WORLD METEOROLOGICAL ORGANIZATION

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

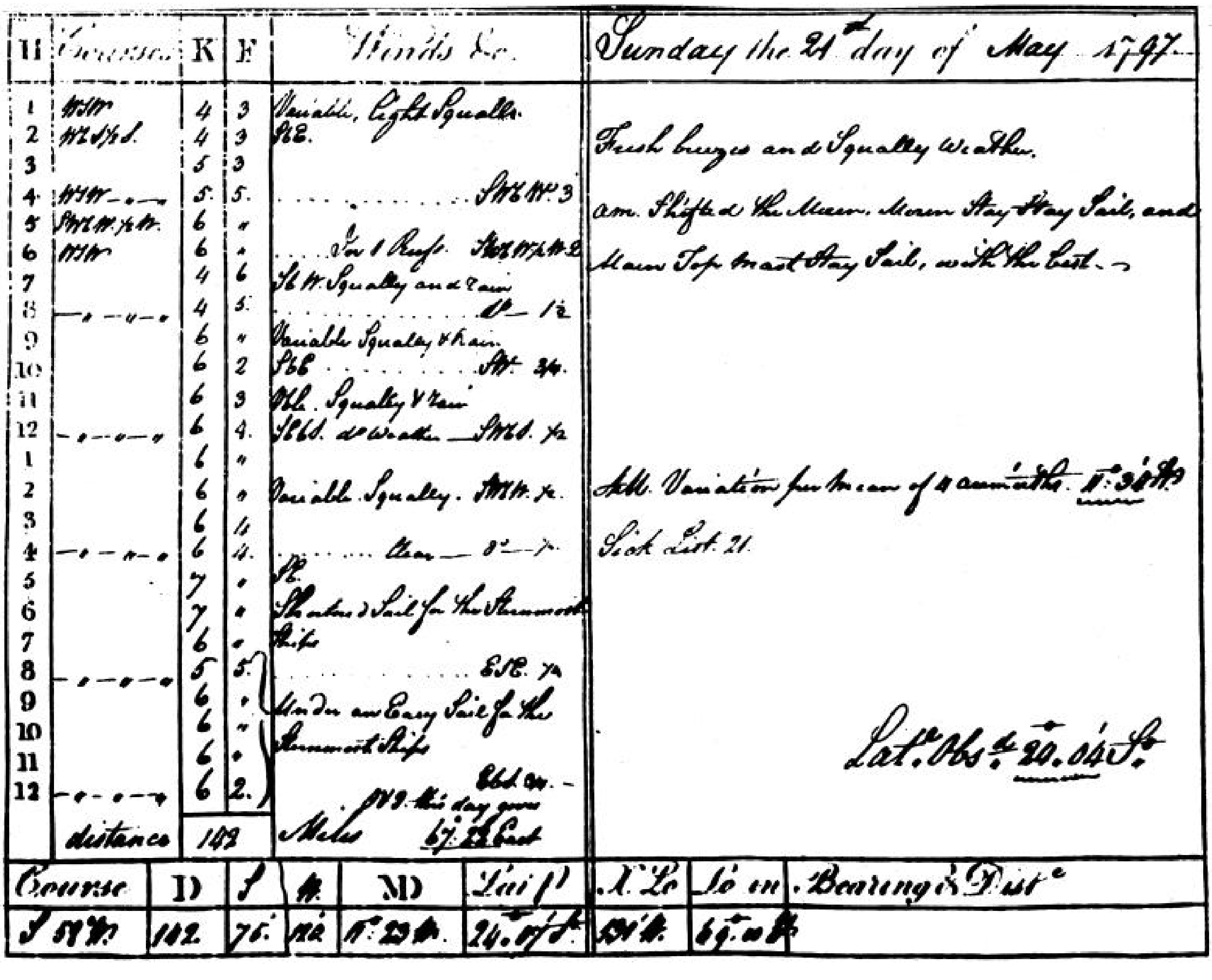
MARINE METEOROLOGY AND RELATED OCEANOGRAPHIC ACTIVITIES

REPORT N° 43

PROCEEDINGS OF THE INTERNATIONAL WORKSHOP ON DIGITIZATION AND PREPARATION OF HISTORICAL MARINE DATA AND METADATA

(Toledo, Spain, 15-17 September 1997)

Editors: Henry F. Diaz and Scott D. Woodruff (NOAA ERL/CDC)


Sunday the 21 day of May 1797
Fresh breezes and Squally Weather.
am. Shifted the Main. Mizen Stay Stay Sail, and Main Top mast Stay Sail, with the best
Sick List 21
Lat. Obs. 20.04 S.

WMO/TD-N° 957

1999

cover: sample page from a log book of a ship from the Maury Collection for Sunday 21 May 1797: column "H" for the hour, and "K" and "F" are for hourly progress in knots and fathoms (a sub-unit of knots in measuring the amount of line run out with the log to determine the ship's speed

Preface

An International Workshop on Digitization and Preparation of Historical Surface Marine Data and Metadata was held 15-17 September 1997 in Toledo, Spain. The workshop was principally organized by the US National Oceanic and Atmospheric Administration (NOAA; Henry Diaz) and the UK Meteorological Office Hadley Centre (UKMO; David Parker), with local arrangements coordinated though Dr. Ricardo García at the University of Madrid. The workshop was also coordinated with the World Meteorological Organization (WMO) and the WMO/United Nations Environment Programme (UNEP) Intergovernmental Panel on Climate Change (IPCC). The workshop was supported in part by NOAA's Office of Global Programs.

Numerous historical ship records for the ocean's surface reside in national archives; the overall purpose of the workshop was to identify these data and foster and develop projects to make them widely available to help support climate and global change research, including the IPCC assessment planned for the year 2000 or 2001.

Specific goals of the workshop were: a) to provide international encouragement for ongoing digitization projects focused on historical marine data sets; b) to identify and prioritize undigitized historical surface marine data collections residing in the world's national archives and other locations; c) to outline near-term plans and cooperative arrangements for manual or automated digitization of the higher-priority collections, plus longer-range plans for the lower-priority collections; d) to discuss quality controls and other preparatory processing of the basic data records and associated metadata (e.g., instrumental corrections); and e) plan for international availability including as part of existing large compilations of surface marine data.

There was a total of 26 participants, from Argentina, Canada, China, France, Germany, Japan, Netherlands, Russia, Spain, Ukraine, United Kingdom, United States, and WMO. In addition to climate researchers and WMO representation, there was participation from representatives of some national archive facilities.

The workshop divided into two Working Groups (WGs):

1) Data Archeology and Rescue.
2) Overcoming Data Biases.

An important thread running through the workshop was the need for governments to support the (unglamorous) work of digitizing old ships' logbook data, because these records are essential to the assessment of natural and anthropogenic climate changes on decadal to century time scales.

Examples of governmentally funded efforts include cooperative China-U.S. projects to key 19th century merchant logbooks, Norway's effort to digitize late 19th century Norwegian logbooks, and a UK project to key 1935-1939 UK merchant logbooks. In other cases, non-governmental funding may be crucial. Japan's ongoing digitization of its Kobe Collection, for example, funded largely by the Nippon Foundation, has provided already more than one million additional digital observations around the hitherto sparsely-documented 1911-1920 period.

Acknowledgements

The workshop organizers gratefully acknowledge partial funding support for the workshop from the U.S. National Oceanic and Atmospheric Administration (NOAA) Office of Global Programs and from the Spanish Instituto Nacional de Meteorología. The organizers are also grateful to David Parker, at the UK Meteorological Office, who co-organized the workshop, to the World Meteorological Organization (WMO) for help and encouragement, and to all the individual participants for making the workshop a success. For extensive help with implementation of travel arrangements for many of the workshop participants, the organizers wish to thank Shannon Andrews at the University Center for Atmospheric Research (UCAR), Beverly O'Donnell at the NOAA-Cooperative Institute for Research in Environmental Sciences (CIRES) Climate Diagnostics Center (CDC), and Dr. Ricardo Garcia and his staff at the Universidad Complutense de Madrid for local arrangements in both Toledo and Madrid. We also wish to thank Craig Anderson at CDC for extensive preparatory word processing, and colleagues at WMO Secretariat for finalization and publication of these proceedings.

Contents

---oooOooo---

Recommendations and findings of Working Groups (WG)

Recommendation (as part of WG1 Report)

The International Workshop on Digitization and Preparation of Historical Surface Marine Data and Metadata, 15-17 September 1997, Toledo, Spain:

NOTING that very large amounts of extremely valuable meteorological, oceanographic, and related data and metadata are at risk of being lost due to, inter alia:

- media degradation (such as fading ink)
- national environmental catastrophes (floods, fires, earthquakes, etc.)
- simple neglect
- retirement of individuals with institutional memory

RECOGNIZING the efforts already being made by WMO in rescuing data through the ARCHISS and Data Rescue Projects

REQUESTS WMO to make these dangers known to WMO Members, NMS, associated institutions with data holdings, and to the scientific community in general

REQUESTS FURTHER that consideration be given by the appropriate bodies to enlarge the ARCHISS and Data Rescue Projects to cover also meteorological and oceanographic data and metadata available from the world oceans.

WG1 (Data Archeology and Rescue) Report (Chair: J.D. Elms)

Initial justification: To better understand the impact of low-frequency variation (e.g., ENSO, NAO, decadal-scale phenomena, and their teleconnections) on climate.

Requirements: Marine databases with optimal global and temporal coverage sufficient to depict the low-frequency events over the period of instrumental records.

1. Goals:

1.1 Complete current digitizing projects.
1.2 Place data in usable form (merge).
1.3 Locate and prioritize additional data to fill in spatial and temporal gaps in the 20th Century using the following criteria:
   a) high quality
   b) easy access and minimal cost
   c) metadata availability
1.4 Extend series back in time (first 19th Century, then earlier data) with the following priority:
   a) all instrumental data (temperatures and pressure)
   b) non-instrumental data

2. How to achieve goals:

2.1 Extension of current bilateral and multilateral cooperative efforts.
2.2 Recognition by WMO-CMM-WCRP of the importance of completing (improving) the existing marine digital archives.
2.3 Recognition of the need by national and international funding agencies.

WG2 (Overcoming Data Biases) Report (Chair: D.E. Parker)

Motivation: Biases must be removed before assessing climate changes.

Recommendations:

1. Metadata-bases are essential so:

1.1 Digitize all of WMO Publication 47.
1.2 Create an extension of, or publication similar to, WMO Publication 47 for buoys.
1.3 Gather and link metadata with data already in existing data banks.
1.4 Recent metadata should include observing instructions.
1.5 Availability of metadata should guide choice of digitization projects.
1.6 Gather and link metadata with newly digitized data. These metadata should include ship name and type, ship libraries, observing instructions.
1.7 Encourage national workshops for Port Meteorological Officers. WMO has already started an international series (CMM XII).

2. There is a need for continued research in the following areas:

2.1 Investigation of biases versus day/night, COADS source deck, country, platform type, measured versus observed winds, etc.
2.2 Disturbed airflow over ships.
2.3 Sea surface temperature biases.
2.4 Capitalizing on advanced statistical techniques for both quality-control and bias-removal.
2.5 Independent validation of satellite data using in situ marine data.

3. Miscellaneous:

3.1 The IMMT format is unsuitable for digitizing old data, because it holds insufficient scope for metadata.
3.2 Historical data should be inventoried to facilitate choice of which data to digitize and to aid the monitoring of progress.
3.3 Feedback is needed from operational centers and reanalysis groups regarding platforms with persistently unreliable data.
3.4 Web site should contain papers on biases, on statistical techniques, and links to relevant experts.

4. Delivery:

4.1 International partnerships.
4.2 Workshops.

Part 1

Digitization of historical ship data and metadata

The Quality Control of the Maury Collection

Guo Fengyi
National Marine Data & Information Service
State Oceanic Administration of China

1. Introduction

According to the Joint Implementation Plan agreed to between the U.S. National Climatic Data Center (NCDC) and the China National Oceanographic Data Center (CNODC), Nov. 17, 1993, CNODC was responsible for digitizing 1.4 million observations from the Maury Collection. The project was finished by CNODC's Data Processing Division, which has 50 staff members including 3 professors, 10 senior engineers and 25 engineers. Most of the staff members have graduated from universities. The computer hardware available for supporting the project is shown in Table 1.

Table 1
Computer equipment available at CNODC for the Maury project.

Equipment type	Quantity
PC Computer: Compaq & Legend 586	80
Compaq 486	30
Server: Sun SPARCServerl000	4
Workstation: Sun SPARCStation20	10
Workstation: Sun SPARCStation5	8
Workstation: SGI Indigo2	1
Calcomp A0 Plotter	6
Calcomp A0 Digitizer	4
Scanner (A0 & A4)	7
HP Printer	20

One of the primary goals of CNODC was to provide high quality digitization of the Maury Collection. Following summarizes the quality control we applied to the Maury Collection.

2. Digitized data processing procedure

Most of the Maury Collection was observed over a century ago, with hand-written information frequently difficult to read. Data were keyed from third generation copies (originals, to microfilm, to enlarged paper copies). In order to ensure the quality of digitized data we used the following procedure:

a) Designed record formats and codes to include all the needed parameters and information.

b) Wrote the digitized data on paper according to the designed record format. This was done by oceanographers because they had sufficient knowledge to correctly convert the described information into digitized data or codes.

c) Input the digitized data into a computer.

d) Printed out listings of the digitized data.

e) Checked the digitized data against the original paper copies. This was done by different oceanographers each time.

f) Corrected errors which were discovered.

Steps a)–f) would be repeated many times until no errors were discovered. Through this procedure we were able to make sure that the digitized data corresponded to original data.

3. Automated quality control procedures

After finishing the above procedure we used a quality control program to perform automated quality control of the digitized Maury Collection data. In the quality control program we used the following procedures:

a) data field verification
b) data relationship verification
c) climatological tests

3.1 Data field verification

The data field verification procedure performed a variety of automatic checks on individual data fields to assure that the characteristics of any given field matched characteristics defined as acceptable for that field. The tests were performed to assure that:

a) Each field contained the expected data type.
b) Data values were written within acceptable ranges.
c) Fields containing code values contained acceptable codes.
d) Mandatory fields were present.

Tables 2 and 3 provide the details of the header record, and data record, field tests. Tables 4 through 10 define the acceptable codes for selected fields in Table 3.

Table 2
Header record field test.

Position	Description	Range
1-2	Microfilm reel number	Number, 01-88 (mandatory field)
3-6	Frame #	Number (mandatory field)
7	Voyage sequence	Number, 1-9
8-37	Type and name of ship	Characters
38-39	Form type	01-04
40-63	Commander	Check spelling
64-87	Departure location	Check spelling
88-111	Destination location	Check spelling

Table 3
Data record field test.

Position	Description	Range
1-7	Serial voyage number	A unique number corresponding to header positions 1-7
8-11	Year	Number
12-13	Month	01-12
14-15	Day	01-31
16-17	Hour	00-23
18-19	Latitude degrees	00-90
20-21	Latitude minutes	00-59
22	Hemisphere	N or S
23-25	Longitude degrees	000-180
26-27	Longitude minutes	00-59
28	Hemisphere	E or W
29-35	Current direction	Characters
36	Current speed indicator	1-8 (Table 4)
37-40	Current speed	0-30 knots; or descriptive terms: 40-58 (Table 5)
41-42	Current minutes lat/lon	Number
43-44	Current time period	01-24
45	Magnetic var. ind.	1 (deg) or 2 (points)
46-50	Magnetic variation	Number
51-52	Barometer indicator	25-27
53-56	Barometric pressure	940-1050
57	Temperature ind.	16 (Table 6)
58-61	Attached thermometer	-20-40
62-63	Barometer ind.	Same as for first barometer indicator (positions 51-52)
64-67	Barometric pressure (second entry)	Same as for first barometric pressure (positions 53-56)
68	Blank	
69-72	Attached thermometer (second entry)	Same as for first attached thermometer (positions 58-61)
73-74	Barometer indicator	Same as for first barometer indicator (positions 51-52)
75-78	Barometric pressure (third entry)	Same as for first barometric pressure (positions 53-56)
79	Blank	
80-83	Attached thermometer (third entry)	Same as for first attached thermometer (positions 58-61)
84-85	Hour of first temperature entry	00-23
86-89	Air temperature (first entry)	-20-40
90-93	Water temp. (surface) (first entry)	-3-37
94-97	Water temp. (depth) (first entry)	-3-40
98-99	Hour of second temperature entry	Same as for first hour group
100-103	Air temperature (second entry)	"
104-107	Water temp. (surface) (second entry)	"
108-109	Hour of third temperature entry	Same as for first hour group
110-113	Air temperature (third entry)	"
114-117	Water temp. (surface) (third entry)	"
118-124	Wind direction (first part or only entry)	Characters
125-126	Wind force (first part or only entry)	0-12 (Beaufort force); or descriptive terms: 20-33, 40-73 (Table 7)
127	Wind force (more details)	1 (breeze), 2 (gale), 3 (trade), 4 (wind)
128-134	Wind direction (middle part)	Same as wind direction; first part (positions 118-124)
135-137	Wind force (middle part)	Same as wind force; first part (positions 125-127)
138-144	Wind direction (latter part)	Same as wind direction; first part (positions 118-124)
145-147	Wind force (latter part)	Same as wind force; first part (positions 125-127)
148-149	Cloud type (first or only entry)	Characters (Table 8)
150-156	Direction of clouds (first or only entry)	Same as wind direction (positions 118-124)
157-158	Cloud type (second entry)	Same as cloud type (positions 148-149)
159-165	Direction of clouds (second entry)	Same as wind direction (positions 118-124)
166-167	Cloud type (third entry)	Same as cloud type (positions 148-149)
168-174	Direction of clouds (third entry)	Same as wind direction (positions 118-124)
175-176	Proportion of sky clear	00-10
177-178	Hour of weather	00-23
179	Weather indicator	1-3 (Table 9)
180-185	Weather	Characters (Table 10)
186	Weather history indicator	1 (keyed as is) or 2 (from remarks)
187-240	Remarks	Characters

Table 4
Codes for current speed indicator.

Code	Description
1	Knots
2	Miles per hour
3	Miles traveled in 24 hours
4	Miles traveled in 21 hours
5	Meters per second
6	Kilometers per hour
7	Descriptive terms
8	Unknown

Table 5
Codes for current speed in descriptive terms.

Code	Description
40	Brisk
41	Declining
42	Faint
43	Fine
44	Fresh

45	Gale
46	Good
47	Hard
48	Heavy
49	Less
50	Light
51	Moderate
52	Nice
53	Pleasant
54	Squall
55	Steady
56	Stiff
57	Strong
58	Unsteady

Table 6
Codes for temperature indicator.

Code	Description
1	Fahrenheit (all temp. fields)
2	Celsius (all temp. fields)
3	Réaumur (all temp. fields)
4	dry bulb Fahrenheit; sea temp. Celsius
5	dry bulb Celsius; sea temp. Fahrenheit
6	dry bulb and attached thermometer Fahrenheit; sea temp. Celsius

Table 7
Codes for wind force in descriptive terms.

Code	Description
20	Calm
21	Light air
22	Light breeze
23	Gentle breeze
24	Moderate breeze
25	Fresh breeze
26	Strong breeze
27	Moderate gale
28	Fresh gale
29	Strong gale
30	Whole gale
31	Storm
32	Hurricane
33	Baffling
40	Brisk
41	Declining
42	Faint
43	Fine
44	Fresh
45	Gale
46	Good
47	Hard
48	Heavy
49	Less
50	Light
51	Moderate
52	Nice
53	Pleasant
54	Squall
55	Steady
56	Stiff
57	Strong
58	Unsteady
59	Violent
60	Gentle
61	Fair
62	Variable
63	Smart
64	Small
65	Increasing
66	Changeable
67	Tremendous
68	Furious
69	Trade
70	Breeze
71	Spanking
72	Prosperous
73	High

Table 8
Codes for cloud type.

Code	Description
CI	Cirrus
CC	Cirrocumulus
CS	Cirrostratus
AC	Altocumulus
AS	Altostratus
SC	Stratocumulus
ST	Stratus
NS	Nimbostratus
CU	Cumulus
CB	Cumulonimbus

Table 9
Codes for weather indicator.

Code	Description
1	Beaufort weather code
2	4-choice code (A = fog; B = rain; C = snow; D = hail)
3	WMO Code 4677 (00-99)

Table 10
Two different codes were used in the logbooks for weather: the lower-case Beaufort weather code, as listed in this table, or a 4-choice code (represented in the digital data by upper-case letters from Table 9). For logbooks containing written remarks, numeric values from WMO Code 4677 (00-99) were selected during digitization. The different code values could be combined in positions 180-185 (e.g., fdgo = fog and drizzle, gloomy with overcast skies).

Code	Description
b	blue sky
c	cloudy sky (detached clouds)
d	drizzle
f	fog
g	gloomy
h	hail
l	lightning

m	mist
o	overcast skies
p	passing showers
q	squall
r	rain (continuous)
s	snow
t	thunder
u	ugly threatening sky
v	exceptional visibility
w	dew
z	haze

3.2 Data relationship verification

The data relationship verification procedure evaluated the appropriateness of data values in relation to other data values. Relationship tests included the following:

a) Time
b) Day and month
c) Latitude and longitude degrees and minutes
d) Current speed
e) Dry and wet bulb temperature
f) Positions and ship speed
g) Time continuity test
h) Total cloud amount and weather
i) Cloud form and weather
j) Wind speed and weather
k) Air temperature and weather
l) Location and landlock check

3.3 Climatological tests

The climatological tests determined the acceptability of data values in comparison to existing values within environmental (climatological) "models." We used the following climatological tests in our quality control program:

a) Wind speed and direction
b) Magnetic variation
c) Air temperature
d) Surface water temperature

Surface water temperature is presented as an example. We used the COADS data (1854-1979) to calculate mean values of surface water temperatures per month per 5-degree square, and a statistical result to check if the observation value was acceptable. The following formula was used: $M–3.5SD < X < M+3.5SD$, where M = Mean value, SD = Standard deviation, and X = Observation value.

---oooOooo---

Digitizing U.S. Historical Marine Collections (1792-1976)

Joe D. Elms
National Climatic Data Center, NOAA, USA

1. Introduction

Presently, there are three major data collections that have been digitized from the U.S. archives awaiting final quality control and conversion to the COADS format (LMR - Long Marine Report, presently version 6: LMR6) before merging the records into COADS (Elms et al., 1993). The earliest data are from the Maury Collection (1792-1910), which was digitized in cooperation with the China National Oceanographic Data Center (CNODC) in Tianjin. Over 1.4 million records were digitized at CNODC from paper copies of the logbooks held by the National Archives and Records Administration (NARA) in Washington, DC. A second collection is the U.S. Merchant Marine Collection (1912-1946) that was digitized, and is currently being quality controlled, at NCDC. Presently two million of the 3.5 million records digitized have been processed through a quality control (QC) system and the software to convert these data to LMR has been developed and is presently being tested. A third, much smaller collection is the Arctic Ice Station collection, containing observations from various research projects between 1922 and 1976. Many of the observational periods in the Arctic Collection are relatively short and broken.

Many lessons were learned while digitizing these data collections. It became clear that much up-front preparation and analysis were required to prepare the logs for digitizing and in establishing a sufficiently flexible format so that all the unique variations in the data could be digitized without being lost or significantly distorted. It was recognized that historical logbooks (handwritten entries and historical observing practices) were difficult to digitize especially in a production mode where speed is a major factor for the data entry personnel. Errors are introduced simply because of the speed and the inability to always recognize what the observer had intended to convey.

The observers also made a number of errors in their original entries. Because of the various approaches taken to digitize these three data sets it was soon discovered that the records should be keyed as closely as possible to the way the information appears in the logbook. Pre-conversions or conversions on the fly introduce errors that can only be corrected by re-keying the data. To the maximum extent possible all conversions should be performed after the observations are keyed; then if a mistake is made it only requires a reprocessing phase.

2. Maury Collection

A number of different form types were found to exist in the Collection including some that were hand drawn. Figure 1 is an example of a typical page from a logbook in the Maury Collection. The difficulty of reading the 19th Century handwriting made the task very tedious and labor intensive. A small percentage of the logs were in languages other than English which required special attention to ensure that the wind directions and other elements were coded correctly. For example a wind direction coded as an “O” (the letter O) in French would indicate a wind from the west while in German or Russian it would indicate east.

For the Maury Collection we did not stress the importance of coding (keying) non-significant weather. Although we initially thought our keying scheme was well covered by introducing three options for keying the “present weather”, the codes for incorporating such entries as “fine weather” or “pleasant weather” were not available as a one-to-one translation resulting in only approximately 26% of the digitized reports containing a “present weather” entry.

This keying practice did not preserve a distinction between a missing report, non-significant report, or a report that was not legible. Lumping all the missing “present weather” reports together and labeling them non-significant would likely introduce a strong bias towards fair weather, an unintended result. However, using the present weather as it was digitized would indicate that there were few, if any, pleasant days observed (biased almost totally towards significant weather).

During the Maury period observing practices and instrumentation were not very standardized especially between countries. The 1853 Brussels Conference did manage to establish important guidance and goals towards global standardization, but it took many years to set these into practice for many of the maritime nations.

These different approaches, practices, and instrumentation affected the logbook entries, thus introducing the probability that errors/biases would appear in the digital database, and making it important to perform a high level of quality control and assurance. For example:

- One logbook reported pressures in Paris inches (Paris inch x 1.066 = 1 English inch) rather than English units. It was noted at the time in preparing the logbooks for keying that the pressures appeared consistently low. After investigation, we discovered the word “Paris” written above the column labeled “bar”.

- In many logs the observer had simply entered a landmark instead of a latitude and longitude, or reversed the numeric characters in the ship’s position. An estimated position was placed into the digital record based on the location of the landmark. When the observer would accidentally reverse the characters in the ship’s position (often reversing the latitude and longitude) the course of the ship would become distorted. When such errors were discovered the position could generally be corrected by reversing the characters or latitude and longitude so that the ship was again on course based on a projected track from its departure port to its destination.

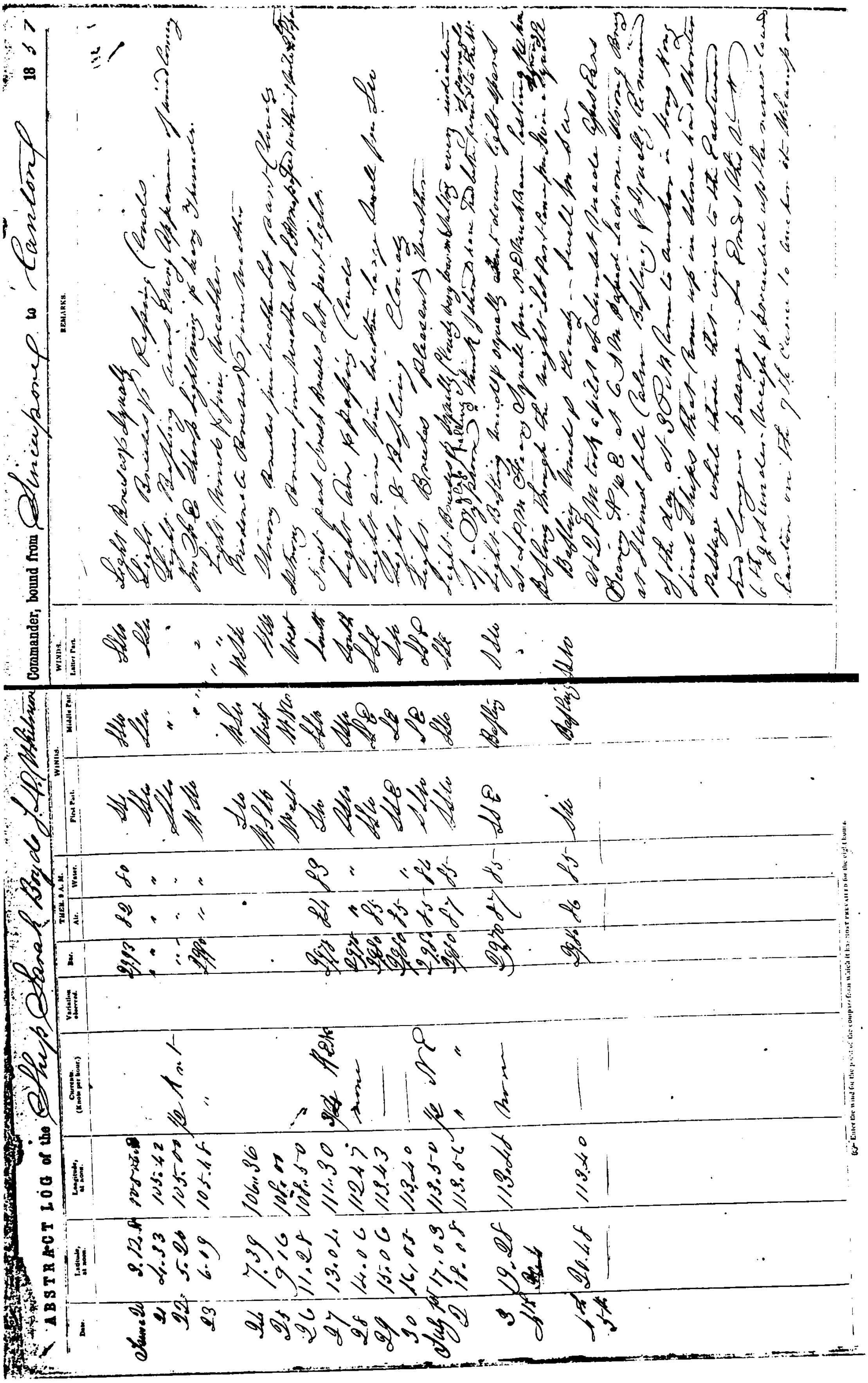


Figure 1: Example of a typical logbook from the Maury Collection. The printed note at the bottom left states: "Enter the wind for the point of the compass from which it has MOST PREVAILED for the eight hours.

- Various methods were used to indicate the wind direction (e.g. Raper's symbols). In one instance a wind force code of 0-8 was used rather than the usual Beaufort (0-12) scale, which we converted to the Beaufort scale as given in Table 1.

Table 1
Conversion of 0-8 code to Beaufort Scale.

Code (0-8)	Beaufort Scale
0 Calm	0
1 Light Airs	1
2 Light Breeze	2
3 Moderate Breeze	4
4 Fresh Breeze	5
5 Moderate Gale	7
6 Fresh Gale	8
7 Tremendous Gale	10
8 Hurricane	12

- There were numerous modes of indicating ditto marks (e.g., do, same). In an effort to determine what to substitute for the ditto mark, the task was often made difficult because (1) at times the handwriting of "do" was difficult to recognize and may have been digitized incorrectly and (2) often it was not clear if the observer meant the value above or to the left of the ditto mark. The reason for this uncertainty as to which value the observer intended was that the forms contained information for the first part, middle part and latter part of the day laid out from left to right across the log sheet and lined up vertically with the previous daily observation (e.g., Figure 1). Generally, the more ditto marks used by the observer to simplify their effort in filling out the abstract log form the more difficult the task of converting the information into the keying/COADS format.

The main reason only one position was established per day and probably the reason for developing a daily observation was that it was most accurate to establish longitude when the sun was directly overhead (local noon) and with a good chronometer giving the time at Greenwich. Knowing the time difference between the position at noon and Greenwich, the observer could compute the noon longitude very accurately because every four minutes difference in time represents one degree of longitude.

Only approximately 65% of the records digitized in the Maury Collection have a complete latitude and longitude entry. Figure 2 demonstrates the global coverage of the Maury digitized records that contained a complete position. It appears that only a few had sufficiently invalid positions to appear over land, a hopeful sign that most records contain a reasonably accurate position.

The total number of records digitized was just over 1.4 million with the vast majority coming from the 1850's as illustrated in Figure 3. It was from this type of graph that we noted that there were no digitized records prior to 1800, yet the archive inventory from the National Archives and Record Administration (NARA) listed several voyages during the 1790's. A year problem associated with 1800 was introduced during the digitizing phase such that all the voyages that took place in the 18th century were displaced one hundred years later in the digital data. Since there were only approximately a dozen voyages that occurred between 1792 and 1799 (e.g. the 1792 voyage appeared as 1892) the dates were easily fixed at NCDC by referencing the microfilm inventory (NARA, 1981).

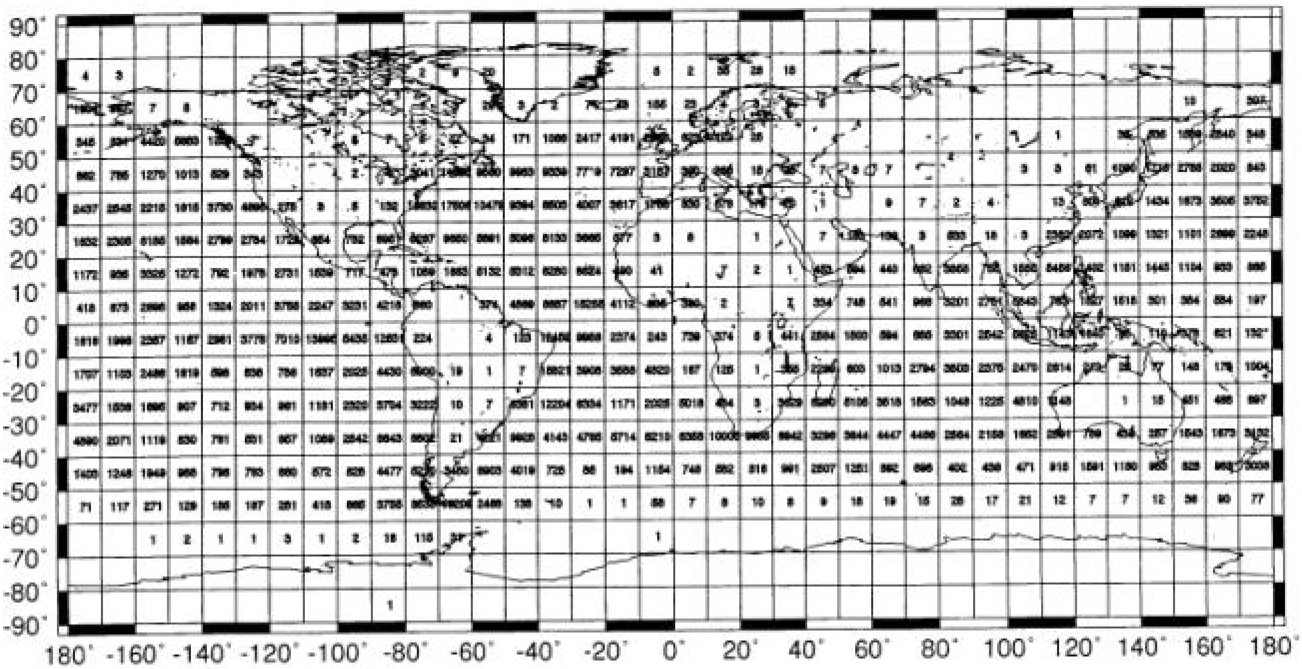


Figure 2: Geographical distribution of the digital records from the Maury Collection that contained a complete latitude and longitude position in the logbook record.

Plans are to place the digitized Maury records as keyed onto CD-ROMs for relatively fast distribution to users. The next step is to convert the records to the COADS LMR format so they can be merged into COADS. To accomplish this a beta version will be produced for distribution and evaluation. A final version will then be produced based on the beta version reviews, and merged into COADS. The tentative plan for converting these records is outlined in sections 2.1-2.5.

2.1 Daily/hourly observations

The bulk of the records in the Maury Collection are effectively daily observations (e.g. Figure 1) with the remainder having specific hours attached to each observation. Where there are daily reports (e.g. year, month and day), we propose that one record be produced for COADS with the time associated with the temperature entry assigned to each record. In contrast, if an observation is reported at a specific time (not a daily report), a corresponding record will be produced for COADS for each hourly report.

The best documentation available for these early records appears to be from the minutes of the 1853 Brussels Conference (Maury, 1854) where the first attempt was made to produce a common universal procedure for taking weather observations at sea. A number of the log forms carried a note stating that whether the day commences at noon or midnight, always call from noon to 8PM "first part". This helped establish the time associated with the information reported in the first, middle and latter parts (a daily observation).

In the daily record format 9AM is generally entered (often pre-printed) above the temperature entries (air and water); these elements plus any additional information (basically prevailing wind information) entered in the "latter part" appears to best represent what occurred during the 8-hour period (4AM to noon), somewhat centered over 9AM.

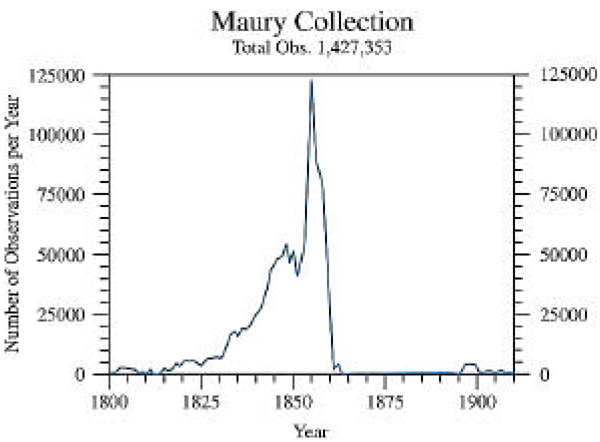


Figure 3: Number of digital observations per year in the Maury Collection.

Therefore, the "latter part" data will be used to complete the 9AM local observation as much as possible in the beta version. If a time other than 9AM appears, procedures will be adjusted accordingly. The remaining keyed information, generally from the first and middle parts (assuming the time of the temperature observation is 9AM) plus from the remarks section, will be placed in the supplemental attachment of the LMR format to ensure no loss of digital information connected with the observation.

2.2 Wind direction and speed

The reported wind direction throughout the Maury Collection is assumed to be magnetic based on the 1853 Brussels Conference minutes (Maury, 1854) where it states; "The direction of the wind is the magnetic direction, with due allowance for appearances caused by the motion of the vessel. It is the direction of the wind which has prevailed for the last 8 hours. It would be expressed to the nearest point of the compass."

The earliest indication that could be found where U.S. recruited ships were requested to report the true wind direction rather than the magnetic wind direction was in the 1906 edition of Instructions to the Marine Meteorological Observer (U.S. Department of Agriculture, 1906).

For the beta version, the following method is proposed to adjust the wind directions from magnetic to true, based on local magnetic declination (based on either the reported or interpolated ship position). We assume that most of the winds were reported according to a 32-point compass, though some may have been reported according to a 16- or even 8-point compass (there is no separate indication of the compass used, and the adjustment will still produce an acceptable approximation in the event of lesser precision in the original compass measurements).

Basically, the adjustment method will (a) assign the magnetic direction to the midpoint in degrees of the 32-point scale, (b) apply the local magnetic declination adjustment, and (c) round the resulting true value in degrees to the nearest 32-point scale midpoint (rounding up when equidistant from two midpoints, etc.). The midpoints we will use for the 32-point scale are defined in Table F2-1 of the COADS Release 1 reference manual (Slutz et al., 1985).

For example if the ship's position was 45°N, 40°W the magnetic declination is approximately 25°W. If the reported wind direction was North the midpoint would be 360°. Adjusted for 25°W the calculated value would be 335°, which rounds to 338° (the nearest midpoint in Table F2-1, corresponding to NNW winds). This adjusted value would appear in the regular section of the LMR format, and would be used to compute the U and V wind components in the COADS statistics.

The original logbook value will be preserved by placing it in the supplemental attachment where it can be recovered (e.g. if the algorithm required adjustment in the future).

Although a few of the observations in the Maury Collection contained wind speeds reported in the Beaufort scale, most were simply in plain-language descriptive terms placed in the Remarks Section. A unique code figure was developed for each descriptive combination, so that the code could be expanded any time a new combination appeared. These coded values will be placed in the supplemental attachment; through cross reference to the keying instructions (documentation), it will be possible for users to determine the exact terms used in the remarks section of the abstract log to describe the wind speed.

Some of these unique codes, (e.g. increasing winds, declining winds, smart winds) likely cannot be converted to an equivalent numeric speed for use in the COADS regular section because of their ambiguity or vagueness.

However those that appear to closely match the Beaufort scale will be converted to the corresponding force and finally to meters per second in the COADS format. This entire process is designed to be reversible.

2.3 Barometric pressure

There is no indication on the Maury abstract logs or in our limited available metadata of whether the barometer entries were "corrected" or "as read". If the values were not corrected it would appear impossible to correct them at this stage because there were no indications as to whether they were "mercurial" or "aneroid", no attached thermometer readings, and no entries as to instrument error. Establishing whether they were corrected values will be difficult, especially if there are mixed entries (i.e. a portion of the entries were corrected, and the others as read). Plans are to produce a monthly mean climatology based on the Maury barometer readings and compare it to a long-term mean climatology (e.g. 1950-1995) based on COADS barometer readings known to be corrected to sea level.

Depending on how well the Maury pressures agree with modern corrected pressures, we hope to be able to draw general conclusions about whether most of the Maury pressure were corrected or not. This type of comparison is most revealing near the equator where the correction for gravity alone is in excess of two millibars.

The pressure entries are assumed to have the same time as the temperature entries. This is based on the minutes of the 1853 Brussels Conference, which state that the dry bulb and wet bulb thermometers should be observed at the same hours as the barometer.

2.4 Clouds

A few of the logs in the Maury Collection contained total cloud information reported as the portion of the sky clear in tenths. This can be converted to total cloud amounts in eighths to conform with contemporary practice and the COADS format. The original reported values will be carried along in the supplemental attachment. There is always a slight bias introduced converting from tenths to eighths. A similar conversion practice was used for a number of the earlier decks appearing in COADS since it was general practice in the early periods to report cloud amounts in tenths rather than oktas.

2.5 Sea surface temperature

The majority of sea surface temperatures recorded in the Maury Collection were most likely observed using a wooden bucket; detailed instructions on how to use a wooden bucket to take sea surface temperature readings were included in Maury (1854). On some of the forms there was also a column labeled water temperature at "depth," for which the following instructions were among those printed at the bottom of each form: "Let the water flow through the ship's cock for a little while; then catch a bucket full, try the temperature, and enter that temperature in the column 'Depth.'"

We have found no clues or anecdotal evidence as to how well the observers followed these instructions; however, they rarely followed the additional instructions at the bottom of the form: "Enter on every page the depth below the surface at which the orifice for the ship's cock enters the sea: in other words fill up the following blank, viz: Depth from which ship's cock lets in water __________ feet _________ inches." The spaces at the bottom of the form were generally blank even when there were entries in the temperature column labeled "Depth."

The scientific value of the Maury Collection has yet to be established. It provides some early observations not previously available in COADS yet the quality and importance of the collection will be determined as the data are examined and used for research.

3. U.S. Merchant Marine Collection

The U.S. Merchant Marine Collection (1912-1946) was selected and keyed in an effort to fill in the critical World War I and II data gaps in COADS. We were rather successful for the World War I period, but largely unsuccessful in locating U.S. merchant marine logs from the World War II period. Unfortunately, the logs from World War II were under the jurisdiction of the Maritime Administration, rather than NOAA, and apparently were destroyed in 1974 under provisions of the Federal Records Act of 1970 (see Elms et al., 1993 for additional information). Figure 4 shows the distribution of the U.S. Merchant Marine Collection's 3.5 million records digitized at NCDC under the World War I and II recovery project. There are a few records in the collection from 1910 and 1911 because these logbooks had been inadvertently sorted in with the later period logs that had been requested for digitizing. Being available they were also digitized as part of the project. Future plans call for us to continue to digitize some of the earlier periods where undigitized logbooks are available.

Most of the reports in this Collection are in a more contemporary format than the Maury Collection and will be easier to convert to the COADS LMR format. However, one significant problem is that a majority of the records, mainly confined to the period prior to the mid 1930's, were observed only daily at Greenwich mean noon. This resulted from the Weather Bureau's desire to produce a daily global synoptic map series, which was considered to be the most important function they could perform in support of a meaningful climatological program. This will introduce diurnal biases depending on geographical location (number of hours from Greenwich), e.g., data at some locations will be systematically weighted towards either the maximum or minimum heating periods of the day.

This is just one more problem that will have to be carefully handled when producing statistics based on COADS because of its large mixture (temporal and spatial) of individual data sources.

4. Arctic Ice Station Collection

The COADS project, with support and assistance from the National Snow and Ice Data Center (NSIDC) and the National Geophysical Data Center (NGDC), produced a digital Ice Station file assembled from all the various sources that could be located (e.g. manuscript, teletype, digital, published, etc.).

Some of the data were previously digitized, but located in various files and formats. In one of the digital files there were no positions included; only an indicator as to which ice island.

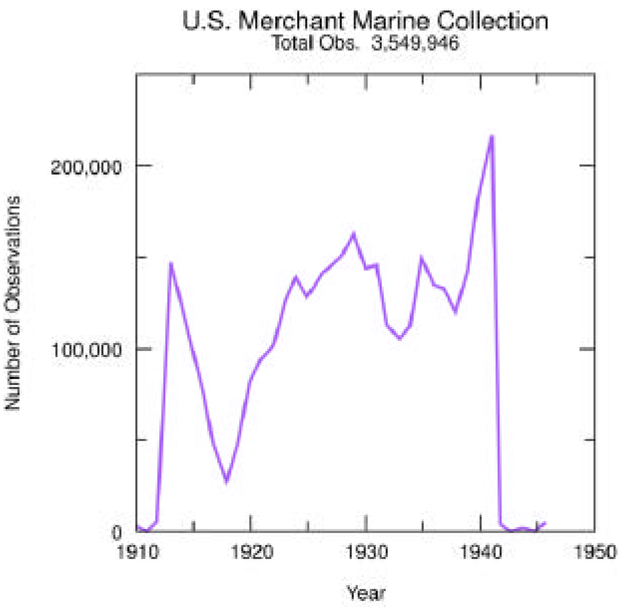


Figure 4: Number of digital observations per year in the U.S. Merchant Marine Collection.

As part of the project positions were inserted into the records, often requiring paging through the original manuscripts to locate the needed positions. In many cases the manuscript records only included a position entry once a day or even less frequently if the observer could not obtain a fix for some reason (e.g. overcast skies).

These data will be converted to the LMR format at which time interpolated positions will be included where feasible. A flag will also be attached documenting the fact that it is an interpolated position. Table 2 lists all the various data sources in the Collection, their periods of record, and a format designator indicating the number of different formats used.

Table 2
Arctic Ice Station Data.
These were manned stations on ice floes, or ships (the MAUD and the FRAM) overwintering in the ice pack.
(Note: none reported sea surface temperature.)

Keying format	Ice station name: sources and time periods
	T-3 Fletcher Ice Island:
803	Synoptic teletype messages July 12, 1963 - May 31, 1966.
804	Plain language messages (coded NCDC) February 19, 1962 - June 12, 1966.
805	WB Forms 610-7; May 28, 1957 - October 25, 1961.
806	WBAN - 10; April 1, 1958 - October 25, 1961.
Deck 117	U.S. Navy hourly ship obs; April 15, 1952 - May 14, 1954; May 1, 1955 - September 16, 1955.
TD3280	WBAN hourly surface obs; June 13, 1966 - April 19, 1971.
	Arctic Research Laboratory Ice Stations (ARLIS):
810	ARLIS I; September 16, 1960 - March 14, 1961.
806A	ARLIS II; January 1, 1963 - December 31, 1963.
809	ARLIS III; June 21, 1961 - September 30, 1962.
	Ice Station “Charlie”:
806B	May 15, 1959 – January 14, 1960.
	Ice Station “A”:
806C	May 22, 1957 – November 6, 1958 (excluding July 1958).
	Arctic Ice Dynamics Joint Experiment (AIDJEX):
808A	(Big Bear, Blue Fox, Caribou, Snowbird); April 1975 - April 1976.
	MAUD - The Norwegian North Polar Expedition with the “MAUD” 1918 - 1925:*
811	September 1, 1922 - August 8, 1924.
	FRAM - The Norwegian North Polar Expedition (1893-1896)**

* During 1918-21 the MAUD recorded full meteorological observations only at 08, 14, and 20 hours. These and other original records (some continuous measurements) for 1918-19 were lost when two members of the party tried to transport the records home by following the coast. Unfortunately, they died en route and the records were scattered by wild animals. For 1922-25 the meteorological records were more comprehensive with six observations per day. Records for a portion of this period (September 1, 1922 - August 8, 1924) were digitized at NCDC. It would be worthwhile to digitize any published records (Sverdrup, 1933) for additional periods, if the necessary resources can be located.

** It is not yet known whether all the observational records collected from the FRAM (July 21, 1893 - April 19, 1896) and published by Nansen (1900-1906) were digitized. Based on spot checks, not all weather elements appear among the data sources presently available in COADS. It would be beneficial to digitize any weather elements, and any full observational records, that prove to be unavailable in digital form, and merge them into COADS.

5. Future plans

These three U.S. datasets along with additional historical foreign datasets will be added to COADS to improve the temporal and spatial coverage as much as possible.

Negotiations are now being conducted between the U.S. (NOAA and NCAR) and China (CNODC) to digitize approximately 1.8 million records from a collection of marine meteorological journals (1879-1893) archived at NARA. These late 19th century data will fill an important gap in the COADS data between the Maury Collection (1820-1860) and the U.S. Merchant Marine Collection (1912-1941).

References

Elms J.D., Woodruff, S.D., Worley, S.J., and Hanson, C.S., 1993: Digitizing Historical Records for the Comprehensive Ocean-Atmosphere Data Set (COADS), *Earth System Monitor*, **4**(2), 4-10.

Maury, M.F., Maritime Conference held at Brussels for devising a uniform system of meteorological observations at sea, August and September, 1853*, Explanations and Sailing Directions to Accompany the Wind and Current Charts*, 6th Ed., Washington, D.C., pp. 54-88, 1854.

Nansen, F., (Ed.), 1900-1906*: The Norwegian North Polar Expedition 1893-1896 Scientific Results*, Volume VI, Greenwood Press, Publishers, New York.

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, Boulder, Colo., 268 pp. (NTIS PB86-105723).

Sverdrup, H. U., (Ed.), 1933: *The Norwegian North Polar Expedition with the "Maud" 1918-1925, Scientific Results*, Vol. III Meteorology, part II tables, Bergen, Norway.

U.S. Department of Agriculture, 1906*: Instructions to the Marine Meteorological Observers of the Weather Bureau*, Washington, D.C.

---oooOooo---

Digitization of the Kobe Collection: The Historical Surface Marine Meteorological Data Collected by the Kobe Marine Observatory, Japan Meteorological Agency

Teruko Manabe
Maritime Meteorological Division, Climate and Marine Department
Japan Meteorological Agency

1. Outline of the Kobe Collection

The Kobe Marine Observatory, which is a field office of the Japan Meteorological Agency (JMA), collected and stored surface marine meteorological data obtained by ships over the period from 1890 to 1961: the Kobe Collection (Uwai and Komura, 1992).

In all, the data obtained by merchant ships, fishing boats and research vessels number about 6.8 million (Figure 1 shows the yearly distribution of data and ships) and those by Japanese navy ships, which cover the period from 1903 to 1944, number about 5 million. Log sheets of the data obtained by merchant ships, fishing boats and research vessels were copied onto microfilm (364 volumes in total) under a JMA-NOAA joint project in 1961.

The same project also digitized all the data after 1933 (about 2.7 million in 185 volumes of microfilm). The digitized data were included in COADS Release 1. However, until recently, the data prior to 1933 had not been digitized.

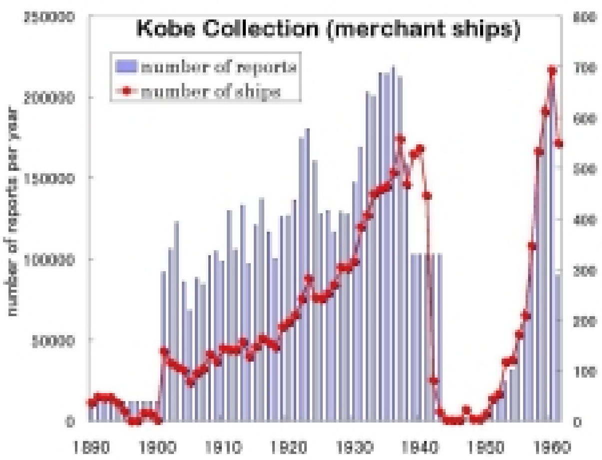


Figure 1: Yearly distribution of reports and ships of the merchant ship data set in the Kobe Collection (from Table 1 in Uwai and Komura, 1992). For the periods 1890-1900 and 1939-1943 only report totals were available; annual bars were estimated by dividing by the number of years per period.

2. JMA project for digitization since FY1995

Because there is great interest in the scientific community in the earlier observations, JMA conducted a two-year project subsidized by the Nippon Foundation with the cooperation of the Japan Weather Association (JWA) to construct a digital data base of the pre-1933 merchant ship data in the Kobe Collection in FY1995 and FY1996.

Considering the data shortage during World War I which can be seen in the presently available COADS, JMA devoted its effort to digitize reports around 1916. Under this project, 1,045,682 reports were digitized.

Unfortunately during the project it was found that a lot of data were not suitable for digitization because of lack of location information, difficulty in interpreting handwriting, etc. In particular, most of the data prior to 1901 were found to be impossible to digitize.

In the digitization procedure, at first, microfilms to be digitized were selected and printed out to make hard-copies of the original log sheets. After coding the information on the log sheets, the data on the coded log sheets were keyed into an “interim” file format, which was designed to include almost all the information on the sheets.

The digitized data in the interim format were converted into the International Maritime Meteorological Tape format, version 1 (IMMT-1), which can be easily handled and widely distributed.

The digitized data were quality controlled in accordance with the minimum quality control standards of WMO’s Marine Climatological Summaries Scheme. The date and location were checked by examining each ship track, and air temperature and SST were compared with JMA climatologies.

About 5% of all the checked data were manually corrected because of errors, which seemed to be mainly based on misinterpretation of the handwritten logs. The digitized and quality checked data will soon be available on CD-ROM.

3. Digitized data

Figure 2 shows the geographical distribution of the reports digitized in FY1995 and FY1996. Data are mainly in the North Pacific, especially along the main ship routes: Japan-North America, Japan-Hawaii-California, etc. For all digitized reports, 82.8%, 11.5% and 5.7% were in the Pacific, in the Indian, and in the Atlantic Oceans, respectively (each basin is in accordance with that shown in Figure 3-2 of Slutz et al., 1985).

Figure 3 and Figure 4 show the yearly distribution of the reports of the Kobe Collection and COADS in the global ocean, and in the Pacific Ocean where a large part of the Kobe Collection exists. In the Pacific ocean, the presently

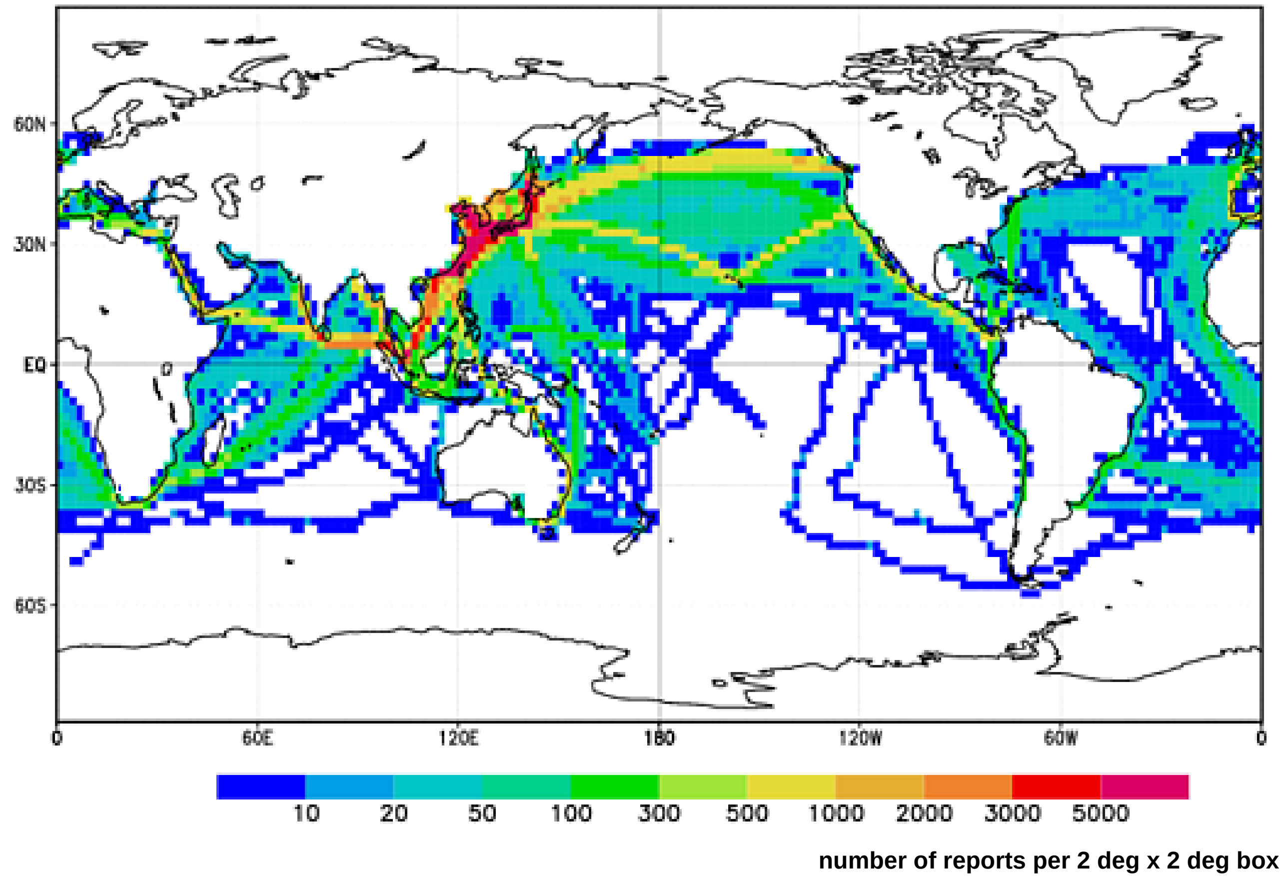


Figure 2: Geographical distribution of the Kobe Collection reports which were digitized in FY1995 and FY1996. Each 2° latitude x 2° longitude box is colored according to the number of reports. (Red indicates that a 2° box has more than 5000 reports.)

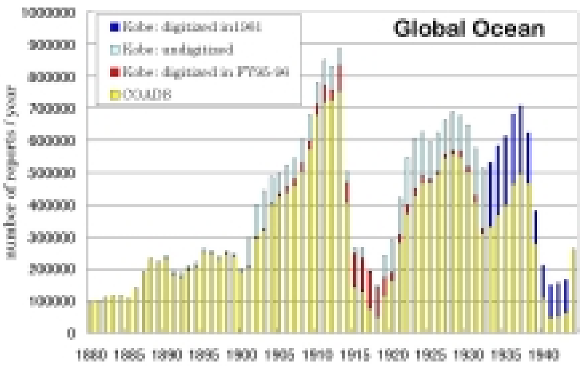


Figure 3: Yearly distribution of the reports of the Kobe Collection and COADS in the global ocean.
Red indicates the Kobe Collection data digitized in FY1995 and FY1996. Light blue indicates the Kobe Collection data which remain undigitized. Dark blue shows the Kobe collection data digitized in 1961 (included in COADS Release 1 decks 118 and 119), which cover the period from 1933 to 1961.

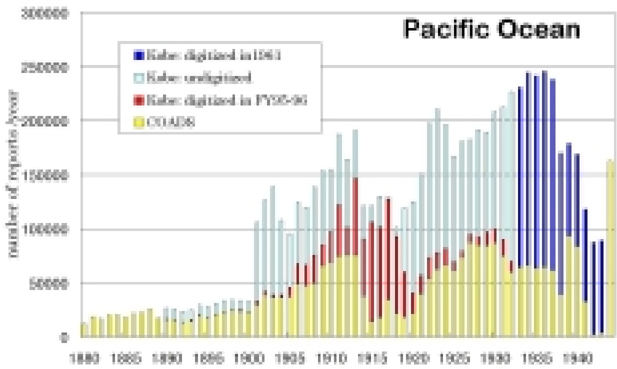


Figure 4: Same as Figure 3 except in the Pacific Ocean.

available COADS, shown by yellow and dark blue, has a significant jump in the number of data between 1932 and 1933, because it includes the Kobe Collection data (1933-1961) digitized in 1961.

Newly digitized Kobe Collection data (1890-1932) will significantly increase the available data, especially in the Pacific Ocean. The number of data during World War I will be greatly increased by the addition of the newly digitized Kobe Collection data to COADS.

4. Future

In FY1997, JMA began a 2-year project to digitize more historical marine records. This is being done with the cooperation of the JWA and is also subsidized by the Nippon Foundation.

At the end of FY1998, it is expected that nearly 1.5 million records will still remain to be digitized. Thus JMA is making an effort to continue to digitize as many data as possible in the period following 1998. Furthermore, JMA is also making a preliminary investigation about the possible digitization of the Japanese Navy data which are also archived on microfilm (257 volumes).

According to this investigation, it seems that only about 10% of all the data include location information. At the moment, there is no specific plan for digitization of these data.

References

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 PB86-105723).

Uwai, T. and Komura, K. 1992: The Collection of Historical Ships' Data in Kobe Marine Observatory. *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*. Diaz, H.F., Wolter, K., and Woodruff, S.D., (Eds.), NOAA Environmental Research Laboratories, Boulder, Colorado, 47-59.

---oooOooo---

Historical Data and its Processing in the Deutscher Wetterdienst

Volker Wagner
Deutscher Wetterdienst, Hamburg

1. Introduction

The basic historical ship logbook data available at the Deutscher Wetterdienst (DWD; German Weather Service) consist of approximately 37,500 journals dating mainly from 1860 to 1945, with a few books back to 1830 (these are estimates based on a partial inventory). The journals are stored in a big cellar with inadequate environmental conditions for books, and the first traces of decay are already visible. The journals contain meteorological measurements and observations, plus information on navigation and shipping as well as on all kinds of natural phenomena. Various observing schemes were used; Figures 1 through 4 provide a selection.

A lot of information was handwritten into the column for "Remarks." Difficult legibility of the handwritten information makes evaluation laborious and therefore expensive.

2. Data distribution

The bulk of the information is from the early 20th century and the 1920s decade. Figure 5 gives the frequency distribution of journals with time for about 20,000 books. Approximately 13,000 books not yet included in these statistics may date from around the end of the 19th century, in which case the distribution will be enhanced for that period near the beginning of the 20th century. The gap between 1915 and 1920 is real and explained by World War I.

The spatial distribution of the journals is shown in Figures 6a and 6b. The world map (Figure 6a) shows the main coastal areas approached and the principal ship routes, with geographical areas denoted by numbers. The frequency distribution of books per each area number (Figure 6b) clearly reflects historical trade routes and big emigrant streams from Europe to America.

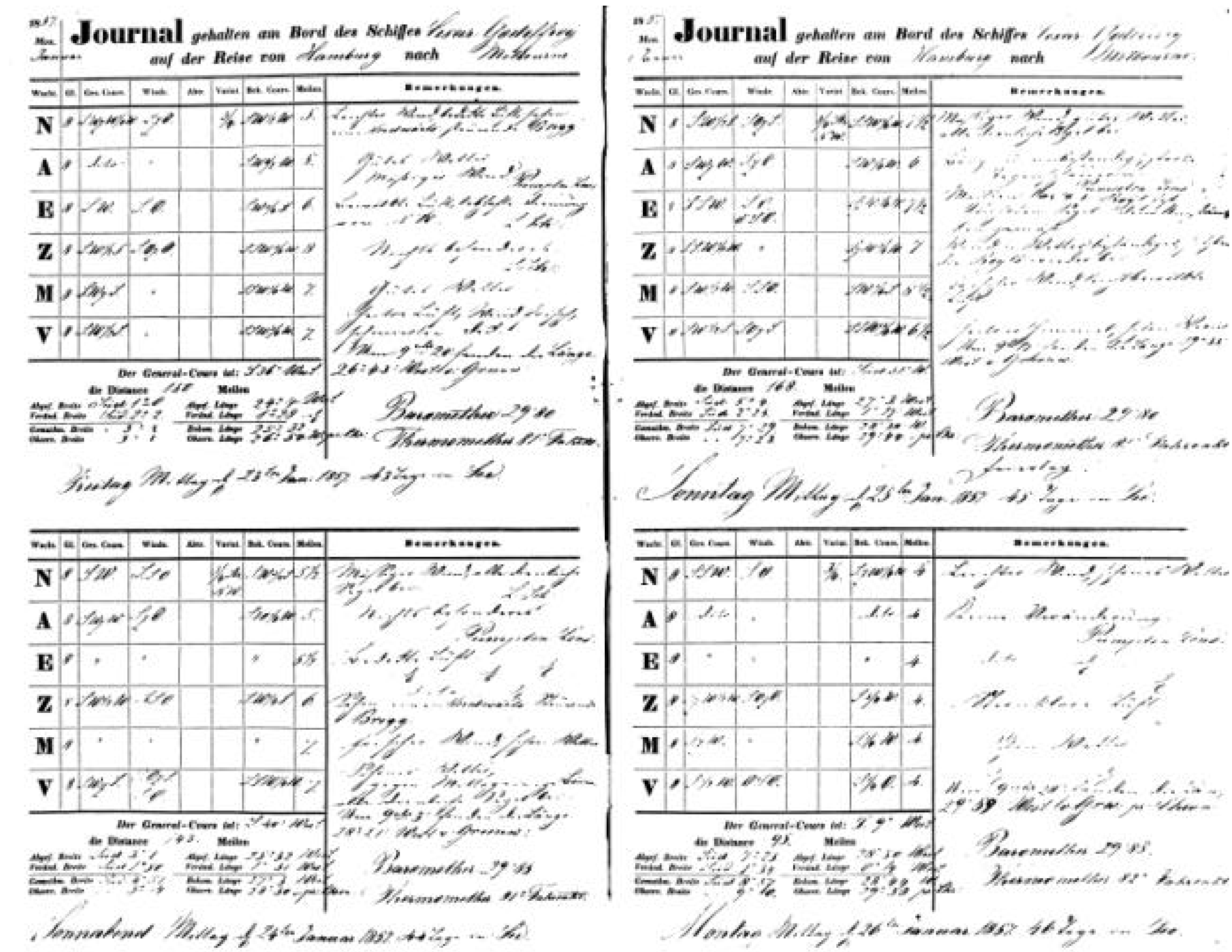


Figure 1: Page from a nautical logbook from 1857 with one daily temperature and pressure reading.

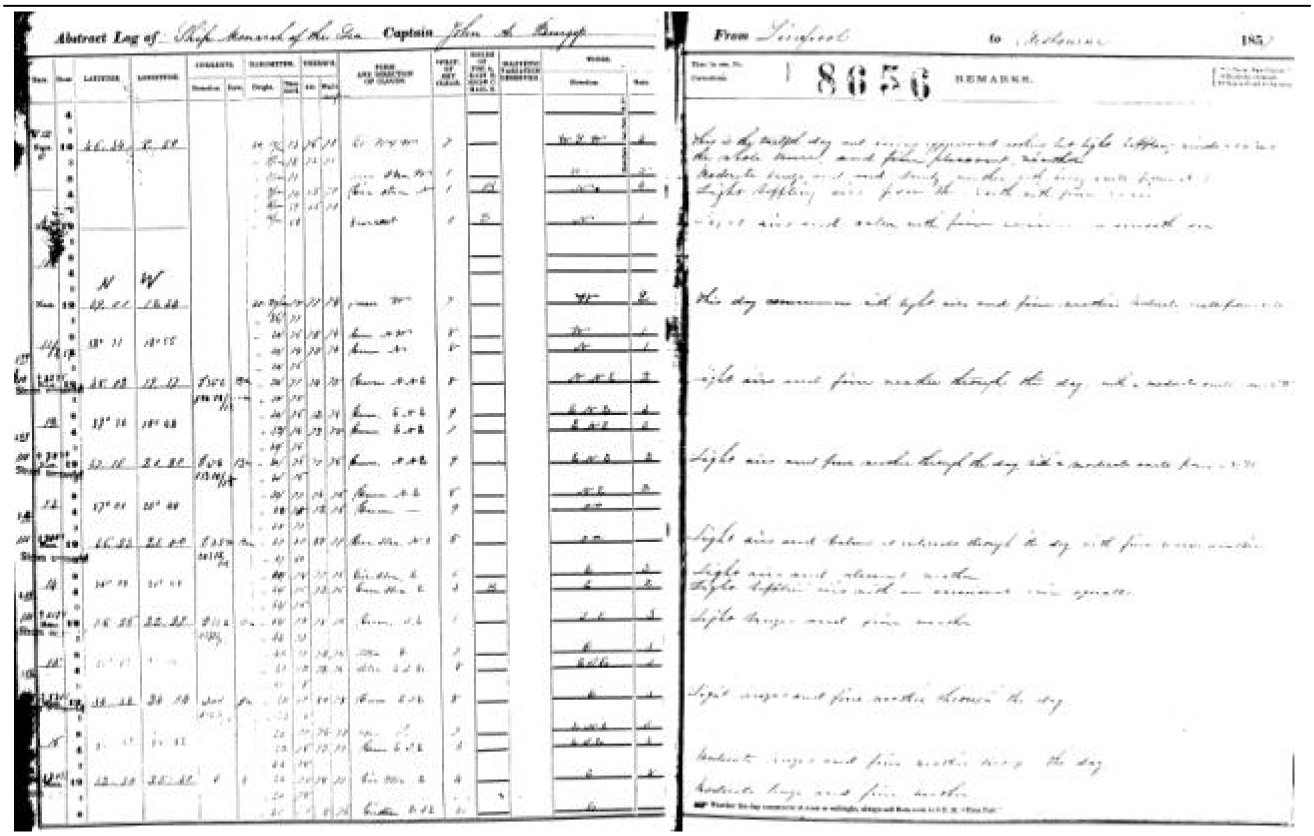


Figure 2: Page from a meteorological journal from 1857 (Maury scheme).

Breaking this down with respect to time (Figure 7) it can be seen that the largest numbers exist in the early 20th century for nearly all voyage areas. But on the Eastern North America route, for example, there is also a considerable number back to the 1870s decade.

So far, we have not been able to accurately assess the amount of observational data related to the numbers of journals, because there is no simple relationship between the number of books and the amount of observations. Unfortunately this information also was not clearly documented during the past history of data typing and extraction, so that this now has to be assessed after the fact.

3. Data processing system

Fundamentally, the system proceeds on a journal-by-journal basis. One journal may contain one voyage, a part of a voyage, or several voyages. Sections 3.1 through 3.4 describe each major step in the data processing system.

3.1 Digitization

Several data "capturing" (digitization) techniques have been tested. But none replaces the laborious work of manually transcribing the desired information, either directly into a computer or onto another medium, e.g., for high-speed typing. The DWD still uses an optical "mark-sheet" reading technique, which is very reliable.

Tests for digital scanning have not been successful to date, because the books do not provide any uniform typefont or appearance that could be used to train interpretation or decoding software.

3.2 Initial processing

The first step is to create a checked copy of the original journal, containing pressure data reduced to sea level. The program-aided steps are:

- check of sequence number and date: wrong observational order, missing lines, and erroneous dates are identified.
- check of completeness of parameters: missing parameters are identified and added from the journal if available.
- check of instrumental corrections and completion of information about physical dimensions.
- transfer of the instrumental corrections to a file and re-calibration of parameters with these correction values (Figure 8 shows instrumental corrections and other information from the front page of a logbook).
- adjustment of parameters including all possible physical corrections.

Processing thus corrects problems in the observational order and dates, adds any data from the journal that may have been omitted, and adjusts parameters using the calibration information.

The last step during this processing is the reduction of pressure to sea level using the barometer height, the temperature at the barometer, and instrumental corrections (calibration data associated with the barometer number).

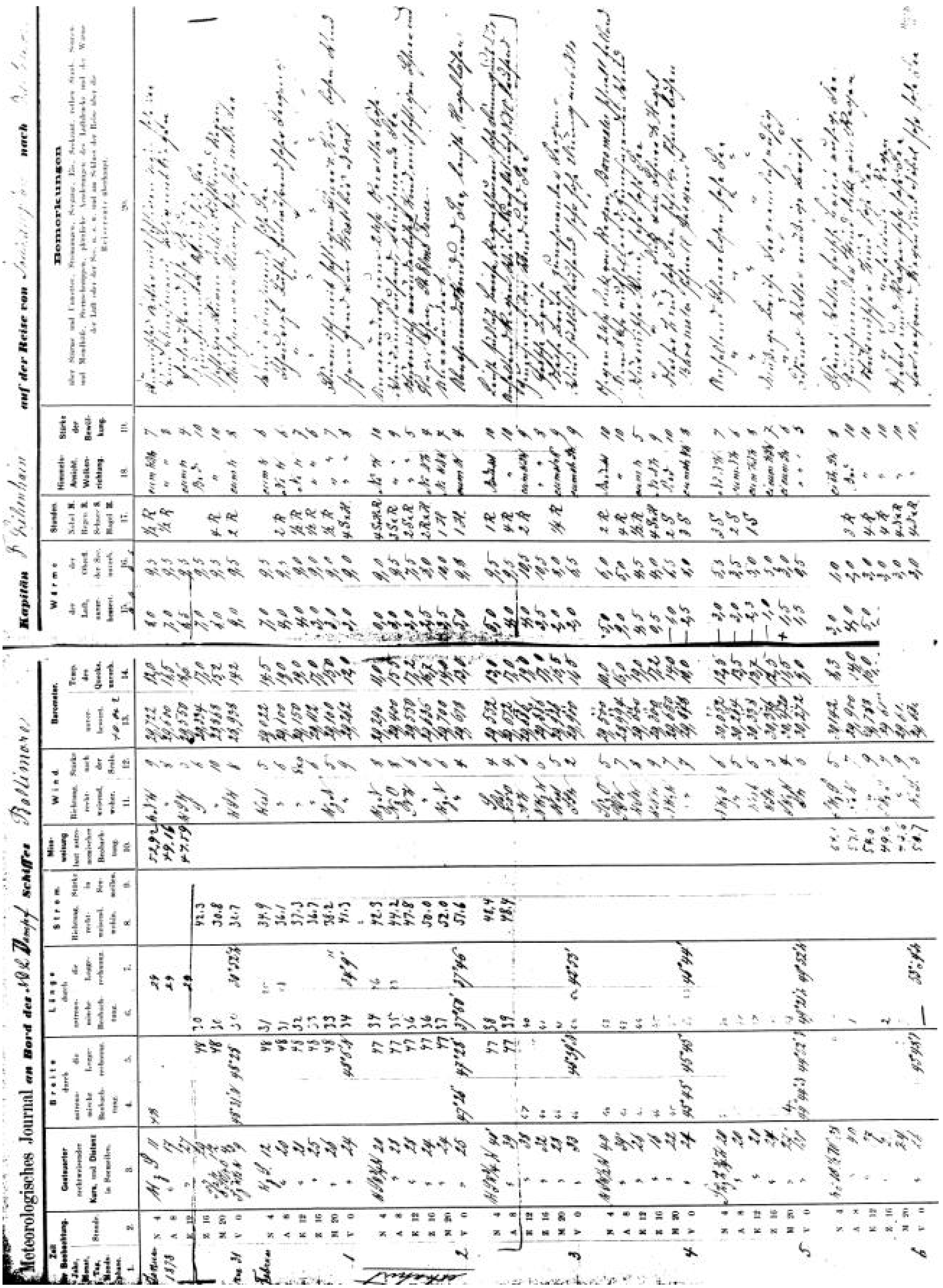


Figure 3: Page from a meteorological journal from 1873.

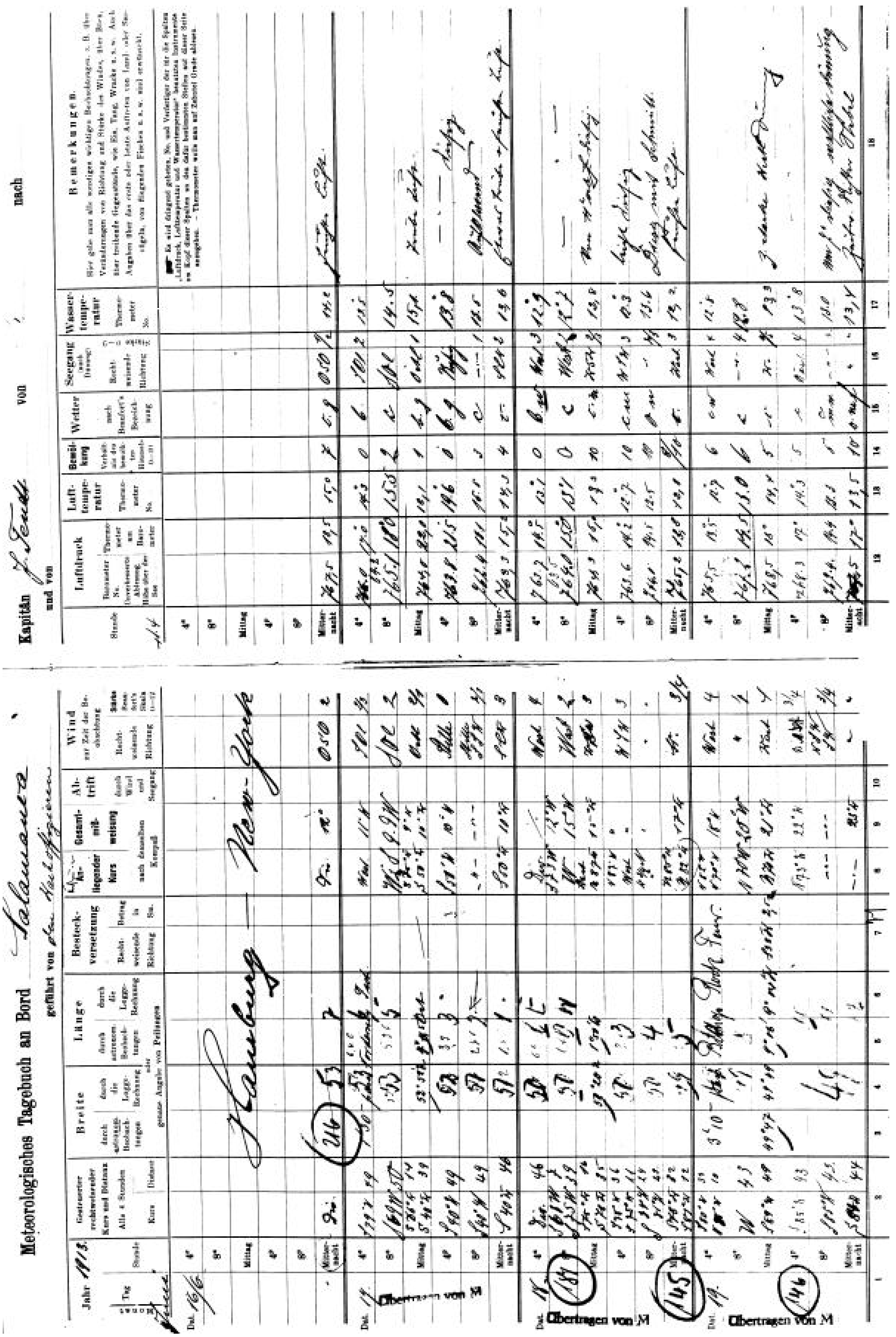


Figure 4: Page from a meteorological journal from 1913; this type of journal was used until the 1940s decade.

For ships under the service of the "Deutsche Seewarte" calibration was done more or less regularly, mostly before leaving port.

Generally, the old calibration data are still available, but sometimes information (e.g. barometer height) is missing. This can mostly be retrieved from the metadata (see section 4).

3.3 Position checking

Starting from a reliable position (obtained by astro-calculation), the dead reckoning positions are calculated by compounding the course and distance for each intervening 4-hour interval until the next astro-calculated (noon) position, as discussed in detail below. If the difference between two positions, the "set," exceeds a critical value, an error is assumed.

The "set" may be variable due to wind, current, or magnetic deviation, and is usually accepted if between approximately 48 and 100 nautical miles per nautical day, depending on the type of ship and on nautical, hydrographic, and meteorological boundary conditions.

The procedure for producing positions for each observation is as follows:

- verification of noon positions.
- calculation of the intermediate positions.
- correction, if necessary, of the intermediate positions calculated by dead reckoning. This is done by adding a weighted part of the total "set" between the adjoining noon positions (weighted, e.g., by wind speed). If the voyage starts without a noon position, e.g., after the ship's departure, backward positions are reconstructed, using the compound course.
- positions over land are checked and corrected.

This procedure is supported by PC-based software that was developed in-house, with the option of manual intervention in addition to the use of commercial navigational graphical-interactive software (especially used for steam ships). The accuracy is 1/10 degree.

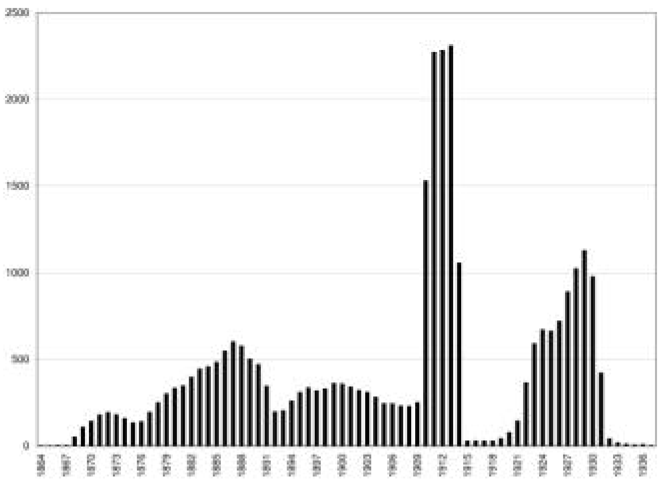


Figure 5: Number of available journals per year.

The most common errors are:

- erroneous quadrant.
- erroneous distances in the data used for dead reckoning.
- computation with declining instead of true course.
- erroneous astro-positioning.
- position on land.

In most cases these errors can be corrected by using the original journal. They are mostly introduced from incorrect journal entries, the interchanging of numbers or letters, or erroneous transcription during the typing process.

3.4 Meteorological quality control

There has been a long history of quality control applied to the archive. Initially, there was a manual quality control, which is sometimes still visible from handwritten notes. These corrected data were typed onto punch cards. A second, independent typing provided a database that reproduced the "corrected" journal information onto a medium for data processing.

Beginning in the 1960s, a computer program was developed for an objective quality checking. This provided primarily internal quality checks for each observation. The core of this program is used still today.

On this basis the whole historical archive was reviewed and once more corrected. It comprises the period 1850 (plus small amounts of information back to 1829) until about 1945. In carrying out the quality control the original journal is very helpful because there are cases when seemingly irregular values are confirmed by the text. This procedure is very laborious.

4. Metadata

During data processing (section 3), two types of metadata are retrieved from the logbooks:

- metadata associated with an individual ship.
- metadata accompanying each individual voyage.

Sections 4.1 and 4.2 discuss these two types of metadata and meta-databases being designed to hold them, although the specific design and contents are still tentative and derived from work with logbooks from the period from about 1890 to 1910 (exclusively steam ships). Section 4.3 discusses a third form of metadata: historical marine observing instructions.

4.1 Individual ship metadata

An example of individual ship metadata is given in Table 1. We try to retrieve all the information that may be helpful for identification of an individual ship, and details concerning its design. Some items are also stored that are not necessarily pertinent to the climatological data, but interesting with respect to shipping history and biographical research. Sources of the stored information are also retained as part of the metadata.

Information about each ship's history and "whereabouts" (Table 2) also is important. Sometimes, for example, a

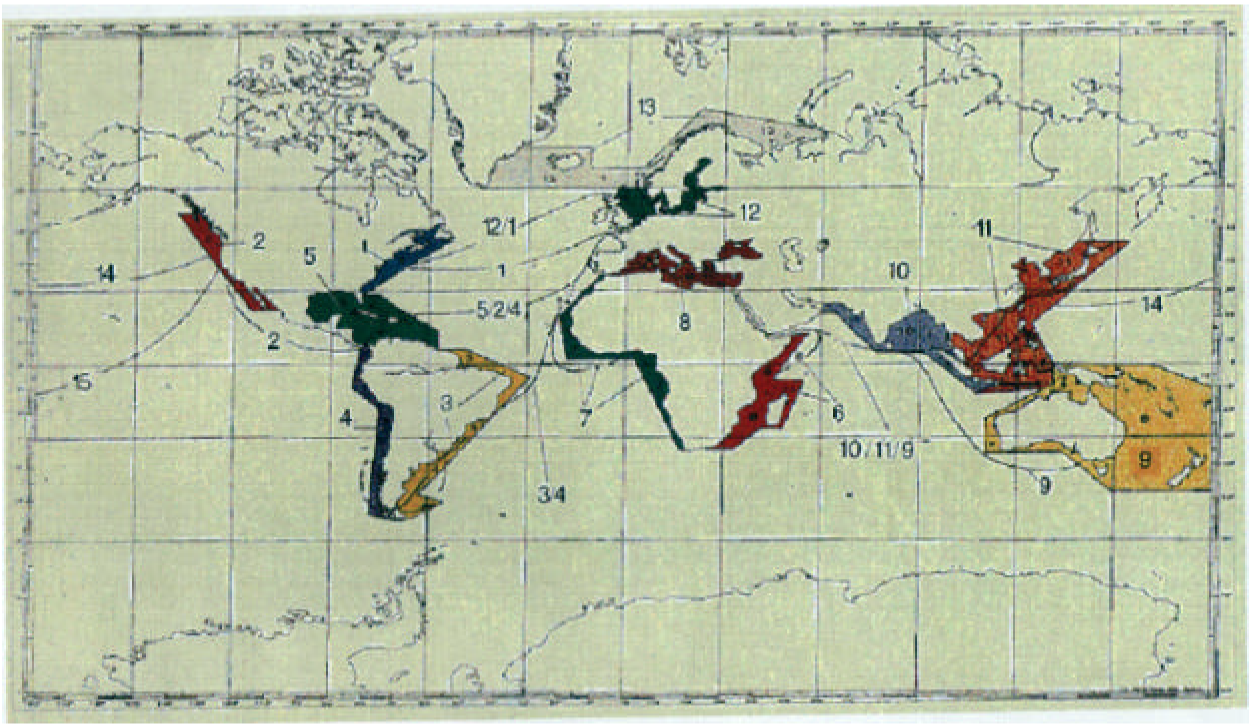


Figure 6a: Numbering of voyage areas.

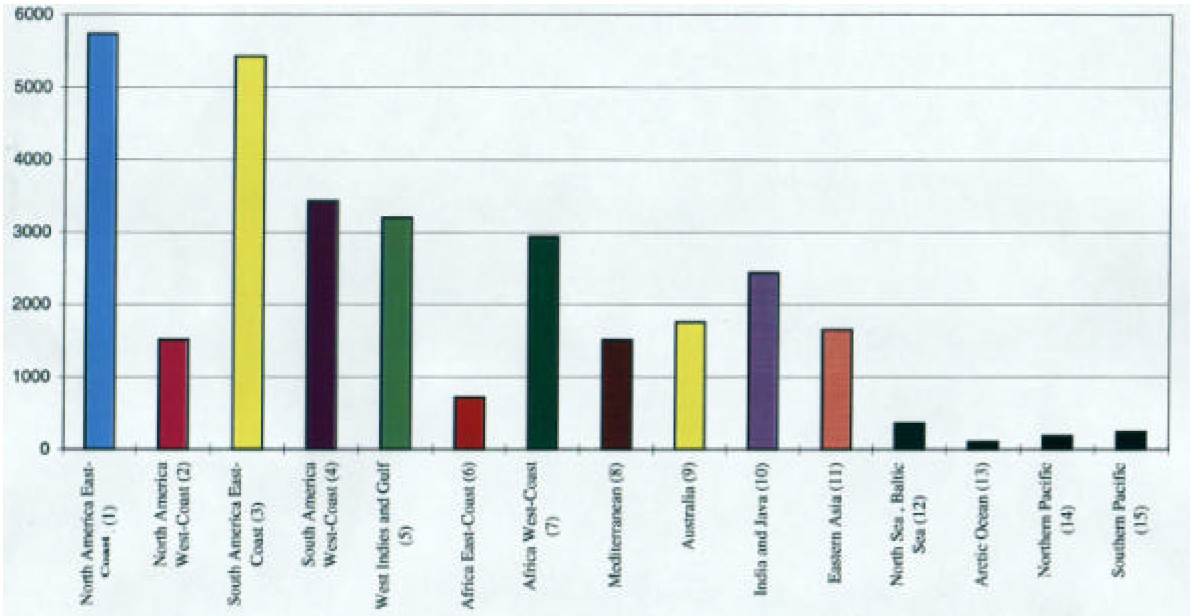


Figure 6b: Number of journals within each voyage area.

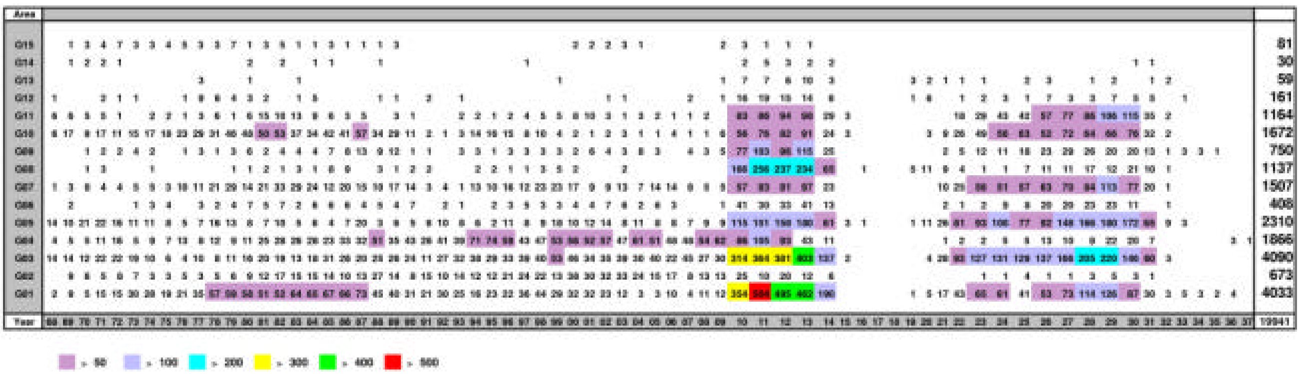
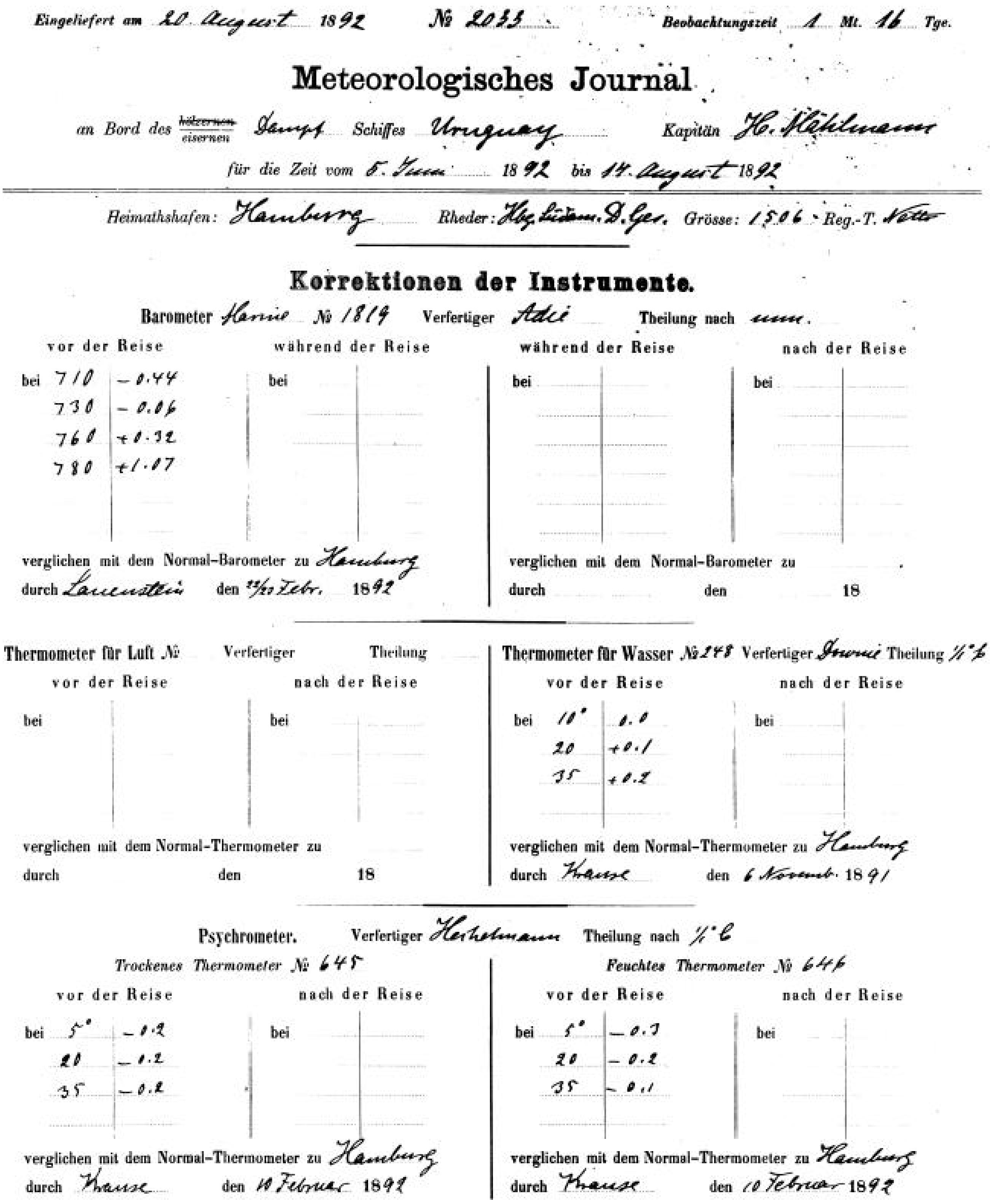


Figure 7: Number of journals per year and voyage area.



Eingeliefert am 20 August 1892 № 2033 Beobachtungszeit 1 Mt. 16 Tge.

Meteorologisches Journal

an Bord des ~~hölzernen~~ eisernen Dampf Schiffes Uruguay Kapitän H. Kühlmann

für die Zeit vom 5. Juni 1892 bis 14. August 1892

Heimathshafen: Hamburg Rheder: Hbg. Süd-am. D. Ges. Grösse: 1506 Reg.-T. Netto

Korrektionen der Instrumente.

Barometer Hanie № 1819 Verfertiger Adie Theilung nach mm.

vor der Reise		während der Reise		während der Reise		nach der Reise	
bei 710	−0.44	bei		bei		bei	
730	−0.06						
760	+0.32						
780	+1.07						

verglichen mit dem Normal-Barometer zu Hamburg durch Lauenstein den 19/20 Febr. 1892

verglichen mit dem Normal-Barometer zu . durch den 18

Thermometer für Luft № Verfertiger Theilung

vor der Reise		nach der Reise	
bei		bei	

verglichen mit dem Normal-Thermometer zu durch den 18

Thermometer für Wasser № 248 Verfertiger Greiner Theilung 1/1°C

vor der Reise		nach der Reise	
bei 10°	0.0	bei	
20	+0.1		
35	+0.2		

verglichen mit dem Normal-Thermometer zu Hamburg durch Krause den 6 Novemb. 1891

Psychrometer. Verfertiger Hechelmann Theilung nach 1/1°C

Trockenes Thermometer № 645

vor der Reise		nach der Reise	
bei 5°	−0.2	bei	
20	−0.2		
35	−0.2		

verglichen mit dem Normal-Thermometer zu Hamburg durch Krause den 10 Februar 1892

Feuchtes Thermometer № 646

vor der Reise		nach der Reise	
bei 5°	−0.3	bei	
20	−0.2		
35	−0.1		

verglichen mit dem Normal-Thermometer zu Hamburg durch Krause den 10 Februar 1892

Figure 8: Front page of journal with calibration information.

series of ships carries the same name. In this case the precise vessel can be identified from the combination of the ship's name and date.

The precise physical identity sometimes is needed in connection with the processing of data from the journals, when the accompanying calibration information is missing. Photos and/or construction plans if available (Figure 9) are scanned and stored with the metadata file.

4.2 Voyage metadata

Metadata for ships' voyages (Table 3) may differ from voyage to voyage, and are essential for recalibration of the measured data. Under the "Comments" field in Table 3, special terms are documented as unique keys, so that an automated query can be based on this information to locate all journals containing a requested term. This information is retrieved when going through the journals during the processing stage and may provide a basis for further research on special weather phenomena, or general ship historical aspects.

4.3 Historical marine observing instructions

Detailed observing instructions are available for a collection of Maury data (dating back to 1850 in the archive), and for German ships starting in 1878 and covering the period up to today. Sometimes individual journals contain instructions on special pages (e.g., Figure 10).

Table 1
Metadata linked to an individual ship
(steam ships only).

Ship's key	RDPV1873 (see Table 2)
Ship's name	VALPARAISO
Nationality	German
Type of ship	2[a], C[b], c[c]
Year built	1873
Material[d]	
Dimensions:	Units:
Length	92.00 m
Breadth	91.05 m
Draft[e]	8.16 m
Gross register tonnage	2247 tons
Net register tonnage	1565 tons
Horse power indicated (PSI)	1160 Ps
Speed[f]	10.0 knots
Home port	Hamburg
Company	Hamburg-Süd
Whereabouts	Y/N
Picture	Y/N

[a] Type of ship, part 1
- 1 = paddle steamer
- 2 = screw steamer
- 3 = screw steamer/tanker
- 4 = special service ship

[b] Type of ship, part 2
- A = bark
- B = brig
- C = topsail schooner/brig schooner
- D = schooner (fore and aft)
- E = schooner

[c] Type of ship, part 3
- a = cargo
- b = passenger
- c = cargo/passenger

[d] Material
- E = iron
- H = timber
- E/H = iron/timber

[e] Draft
- 1 = from measurement
- 2 = from journals

[f] Speed
- 1 = calculated
- 2 = shipyard information

CCCCCCCCCC

Table 2
Metadata of a ship's whereabouts.

Ship's Key: RDPV1873 (call sign and construction year)
Name: VALPARAISO
10 Oct 1873-1893: Operated by Hamburg-Süd
19 Nov 1873: Maiden voyage
1983: Traded in for the "Pelotas" at Edward's shipyard, Newcastle
1894: Sold to W.J. Pitt, North Shields
1896: Sold to Gerard, Valparaiso
1902: Sold to Don Matias, Cia. Esplotadora, Valparaiso
23 Jul 1980: Going down in the bight of Aranco (Chile) after collision, upon voyage from Tocopilia (Chile) to Lota (Chile) with copper ore on board

In addition, some historical publications exist (e.g., Neumayer, 1878). A complete list of observing instructions is being worked on.

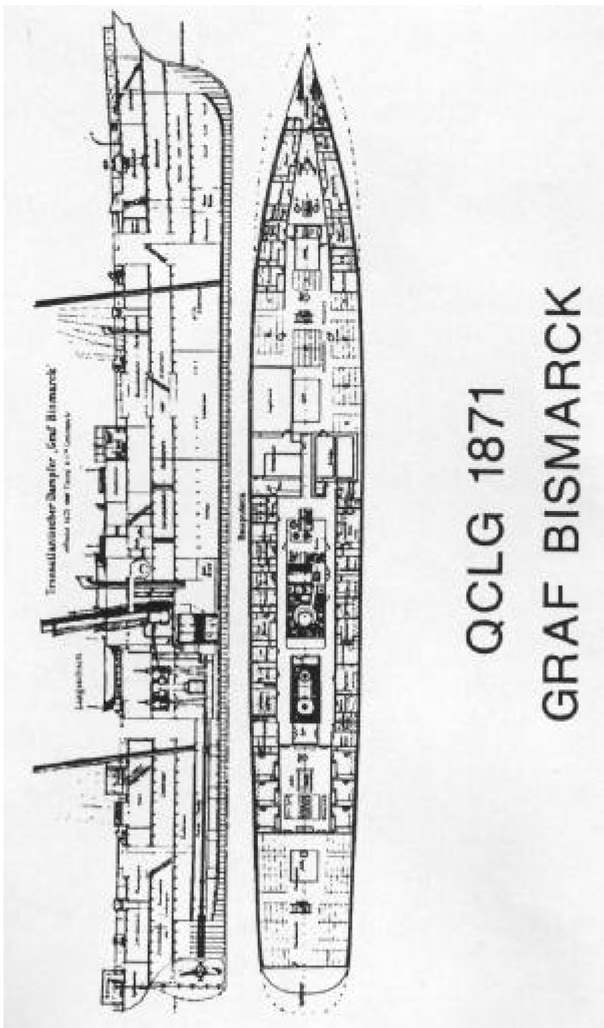


Figure 9: Ship's construction plan added to individual ship metadata. Source used: Koehlers Verlagsgesellschaft mbH, Hamburg, 1991.

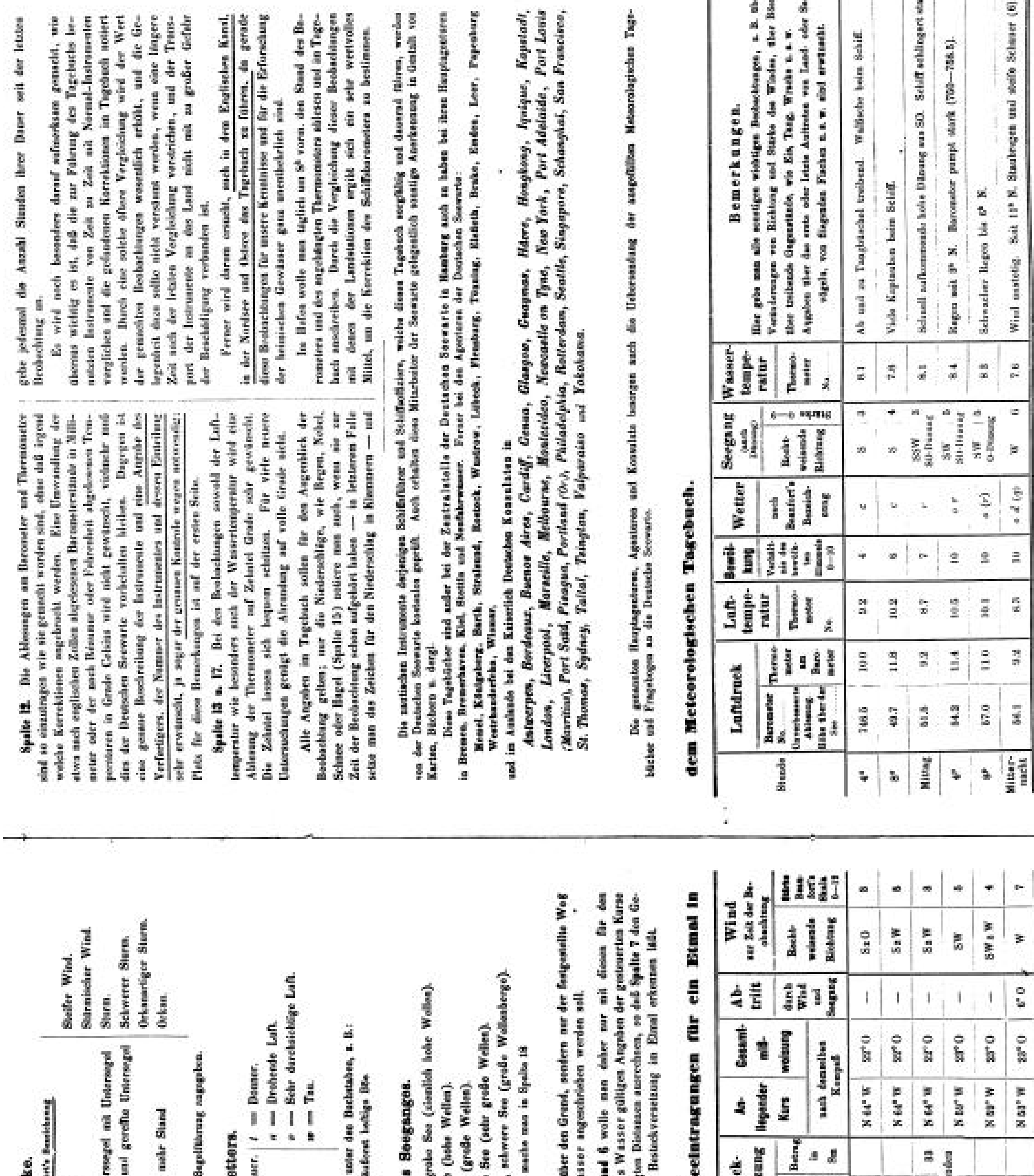


Figure 10: An example from marine observing instructions, extracted from a journal.

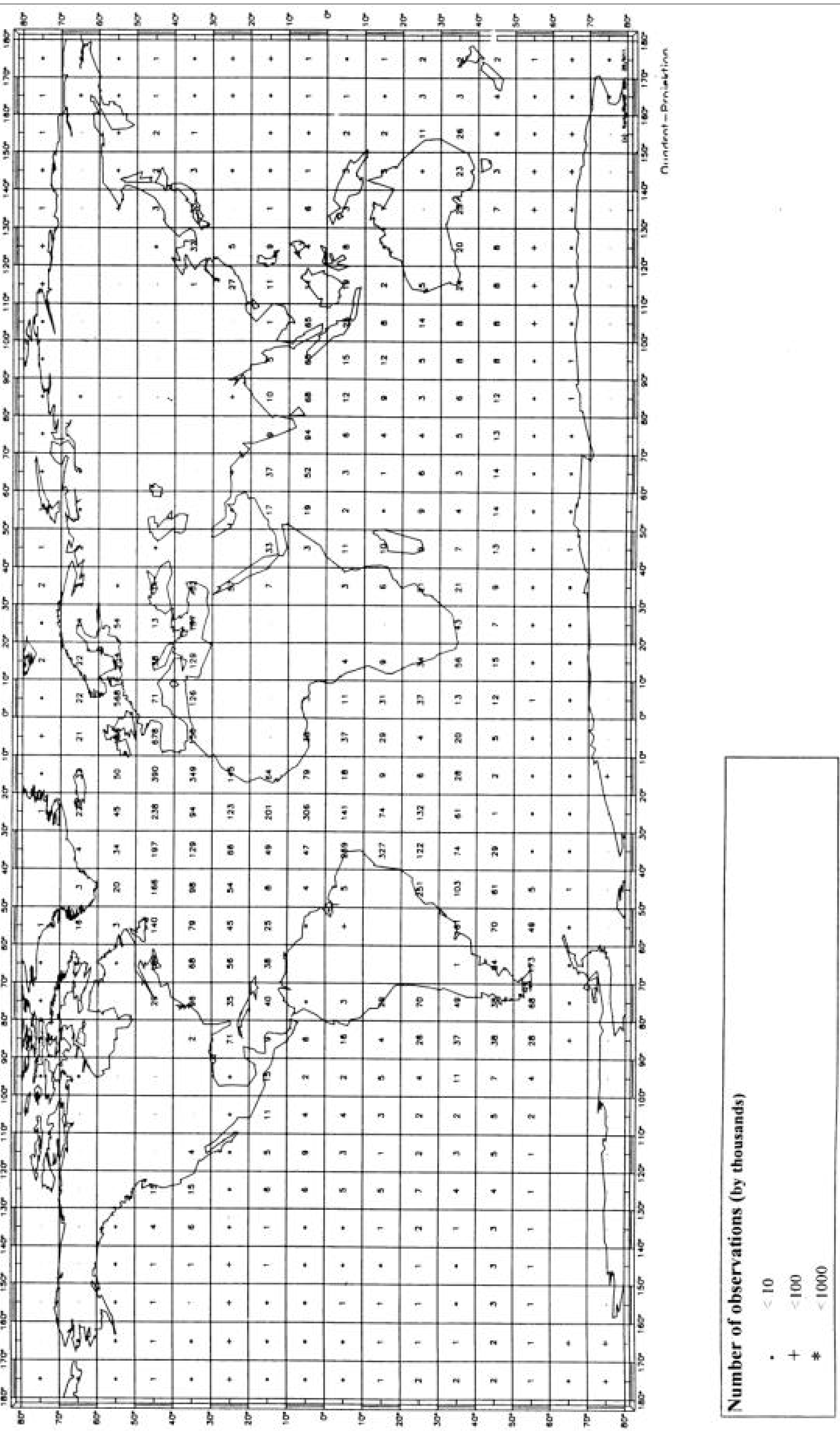


Figure 11: Data distribution of the historical archive, limited to data already available on disk.

5. Data availability

For the journals that have been inventoried, we estimate that they contain about 19 million observations from 37,395 books. About 16 million observational records have been typed. From these, about 12 million are available on electronic media (Figure 11). Another four million records are still under processing. The rest of the approximately three million records still need to be digitized. Besides this there are other journals that still have not been reviewed.

6. Conclusions

DWD holds more than 37,000 marine meteorological journals from the years 1944 back to 1851 and some minor numbers back to 1829. The estimated volume of observational records from these books is about 19 million.

Digitization of these data has a long tradition in the DWD reaching back to the year 1940. Up to now about 16 million observational records have been typed, with varying numbers per year and sometimes dedicated to special projects (e.g. the Historical Sea Surface Temperature Data Project). The worldwide coverage of the observations is generally extensive throughout the period of record.

Part of these data are available in the DWD marine meteorological archive. However, other records have received a lower ranking due to quality control concerns, and therefore are not yet integrated into the archive.

The processing steps applied to the historical data are as follows:

- checking of typed data for completeness.
- computer-aided control and completion of ships positions (use of a navigational electronic chart system, in connection with software developed in-house).
- meteorological quality control.
- storage of voyage-related documentation in a meta-database: the instruments in use and their corrections, plus keywords for weather phenomena and other handwritten remarks (a recent addition to the voyage meta-database).

Another recently installed meta-database is being filled with information on each ship's constructional characteristics, its name, call sign, or other identifiers, history and whereabouts, references to corresponding journal numbers, and a photo/construction plan if available. At present the meta-database covers about 100 ships and is being expanded step-by-step. Some observing instructions, not yet fully documented, are also available.

Table 3

Metadata for a ship's voyage.

Journal No.	D01344
Ship's Name	VALPARIASO
Rack No.	D01250-0135
Journal No. (preceding)	D01239
Journal No. (following)	D01395
Voyage (dates)	22 Oct 1889-21 Dec 1889
Voyage (destinations)	Hamburg/Rio de Janeiro/
Voyage (route)	ENC/GOB/NAO/SAO
Master	Rohlfs, F.
Barometer:	
Barometer Height	6
B-min	5
B-max	7
Dimension	m
Source [a]	
Location [b]	
Scales:	
Pressure	mm
Barometer	°C
Air temperature	°C
Wet bulb temperature	°C
Water temperature	°C
Comments (terms):	hail squall/thunderstorm/ storm/dew/trade winds/St. Elmo's fire/drift bottle/engine stopped, continuing voyage by sailing

[a] Barometer source
1 = journal
2 = calculated mean
3 = calibration information

[b] Barometer location
A = astern
M = midships
V = fore

CCCCCCCCCC

Resources for processing historical data (budgetary and personnel) have been limited during the last few years, although the need for this kind of data is recognized by the DWD. Due to the high degree of manual work needed, the typing and processing of historical data is extremely expensive.

Reference

Neumayer, 1878: *Instructionen fur Führung des meteorologischen Journals der Deutsche Seewarte* (Translation: Instructions for Keeping the Meteorological Journal of the German Marine Bureau), Hamburg.

---oooOooo---

19th Century Navy Shiplogs and 17th-18th Century VOC Shiplogs

Günther P. Können
Royal Netherlands Meteorological Institute

All 19th century Navy shiplogs are preserved in the Dutch State Archive. The logs include readings of pressure, temperature, wind and weather, taken on a regular basis 3 times a day.

The oldest log is from 1812. In total, the logs take 400 m of shelf space, which corresponds to about 4000 volumes.

Even if only 10% of the pages contain useful information, this would correspond to 10 million observations. However, this figure may be easily off by a factor of two. The data were mainly taken along the trade routes from Holland to Indonesia (via West Africa to Cape Town to Mauritius to Indonesia, and possibly Japan) and to America. A digitization project has still to be started.

The 17th and 18th century VOC (Dutch East India Company) ship logs (trade route to Indonesia) are also preserved, containing (visual) wind and weather observations, taken 2-4 times a day. There are about 4000 of them; the oldest is from 1612.

The amount of data in these books is estimated to be of the same order of magnitude as in the Navy books. The accessibility of the VOC books is lower because of the antique handwriting. Summaries of these undigitized holdings are given in Tables 1 and 2.

Table 1
19th-century Navy logs: 1812-1900. Number of logs: approximately 4000. Data elements: pressure, temperature, wind, and weather.

Years covered	1812-1900 (and beyond)
Frequency	3 times a day, no gaps
Quality	depends on the Captain
Total number of useful observations	about 10 million (if 10% is OK)
Accessibility	good
Status	not digitized

Table 2
VOC logs: 1612-1795. Number of logs: approximately 4000. Data elements: weather and wind.

Years Covered	1612-1795
Frequency	2-4 times a day
Quality	probably good
Total number of observations	perhaps 60 million
Accessibility	difficult
Status	not digitized

---oooOooo---

Digitization and Preparation of Historical Surface Marine Data from Russian Research Vessels

Yuri V. Gemish, Vera N. Popova, Elena M. Krakanovskaya, Svetlana M. Somova, Galina M. Nazarova, and Evgeniy D. Vyazilov
Laboratory for Marine Meteorology
All-Russian Research Institute of Hydrometeorological Information-World Data Center-B

1. Introduction

The Laboratory for Marine Meteorology of WDC-B is involved in three main activities:

- creating the marine meteorological data archive on computer-compatible media;
- development of computer technologies for processing of these data;
- research of marine and ocean climate.

We have created the Russian marine data archive named MORMET. This data archive contains about 40,000,000 marine meteorological observations for the period of 1890-1996.

The archive also includes more than 100,000 marine aerological observations and about 30,000 marine actinometric observations. The data archive MORMET was provided to the USA and is now included in the Comprehensive Ocean-Atmosphere Data Set.

The following sources were used in creating the MORMET archive:

- logbooks from ships of opportunity;
- data arriving via communication channels (GTS);
- international exchange data.

In addition to the sources mentioned, there is one more very important data source: marine observations made by research vessels. The WDC-B has a very large number of reports on research vessel cruises. Unfortunately, up to now they haven't been digitized.

This paper is largely devoted to the problem of digitizing data from research vessels; rescue and digitization of some other Russian marine data and metadata sources are also briefly discussed.

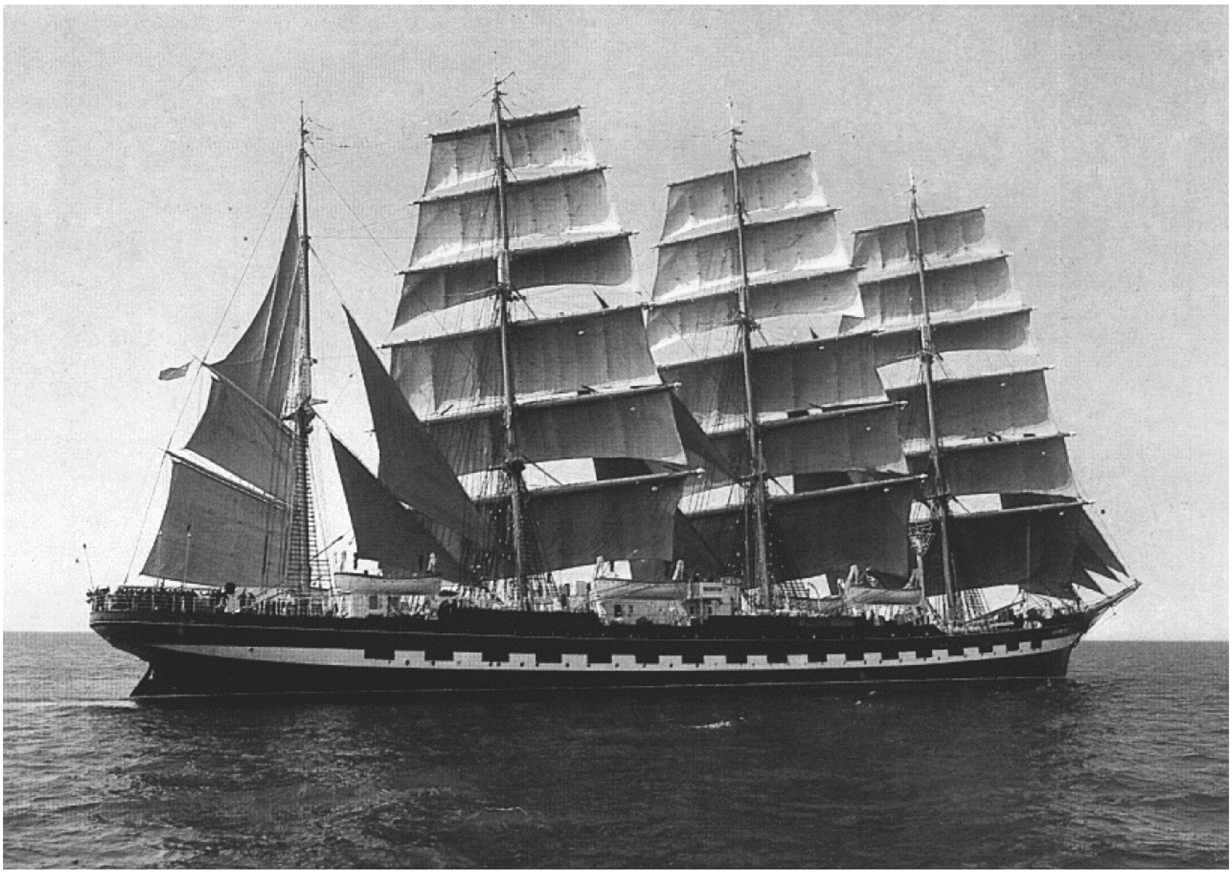


Figure 1a: The research sailing vessel “Kruzenshtern”.

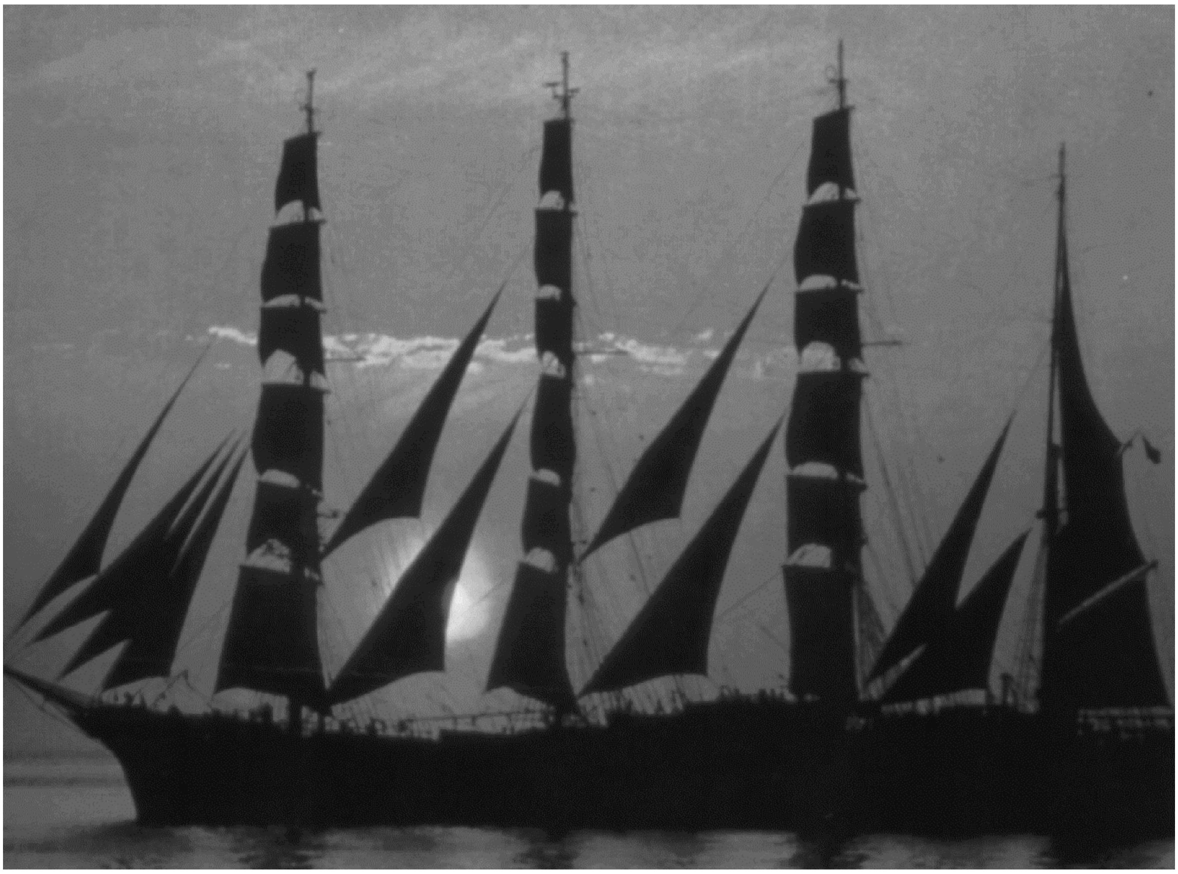


Figure 1b: The largest sailing vessel in the world; bark “Sedov”.

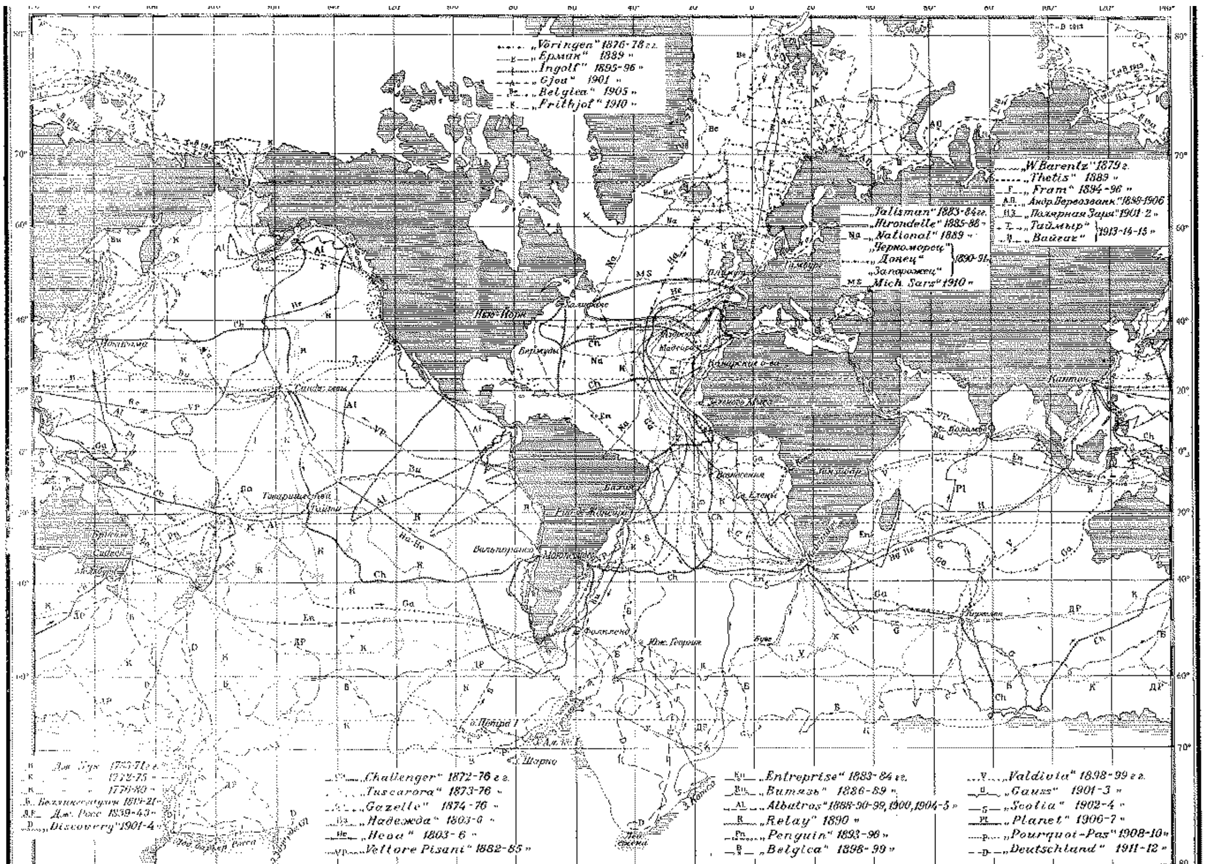


Figure 2: Routes of the most important research expeditions in the 18th and early 19th centuries.

Figure 3a: Research vessel “Academician Petrov”.

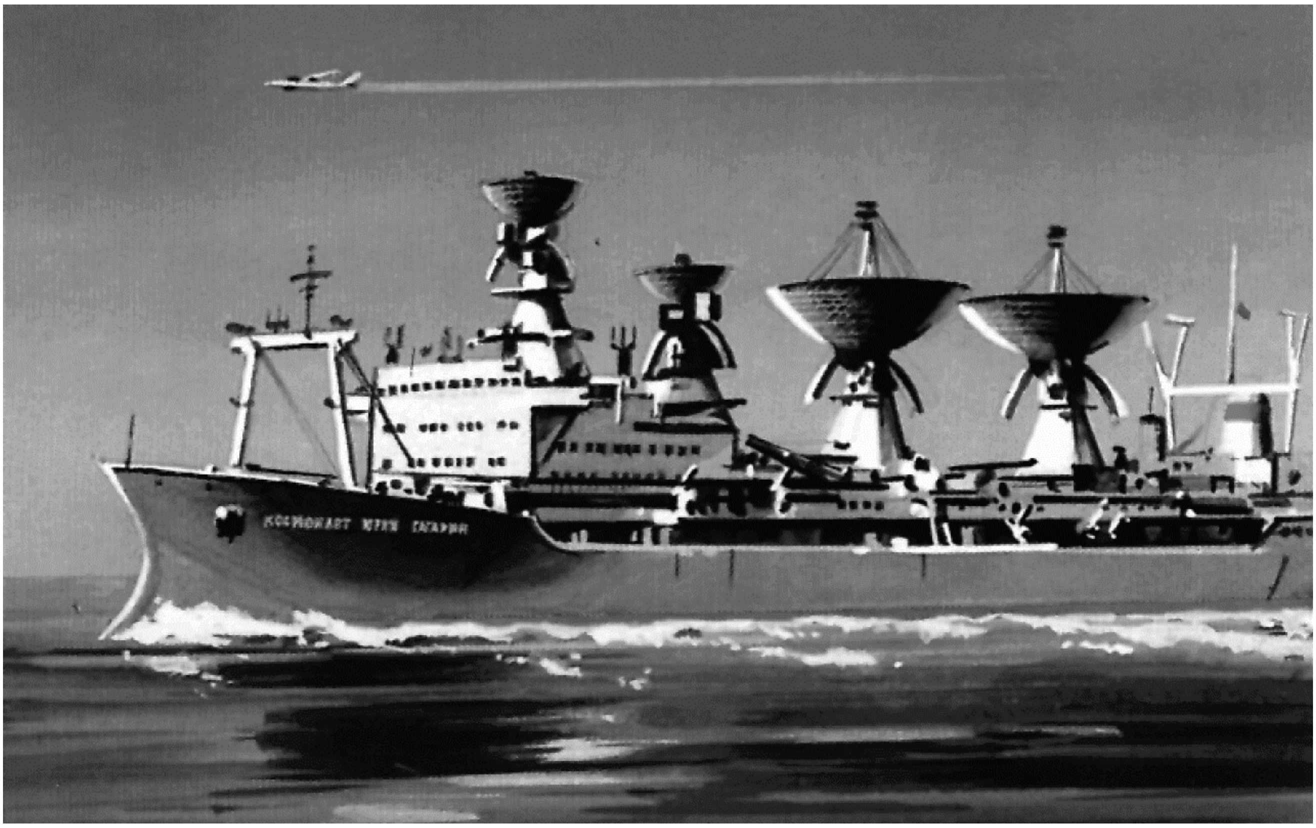


Figure 3b: Research vessel “Cosmonaut Yuri Gagarin”.

Figure 4a: Ice breaker “Arctica”.

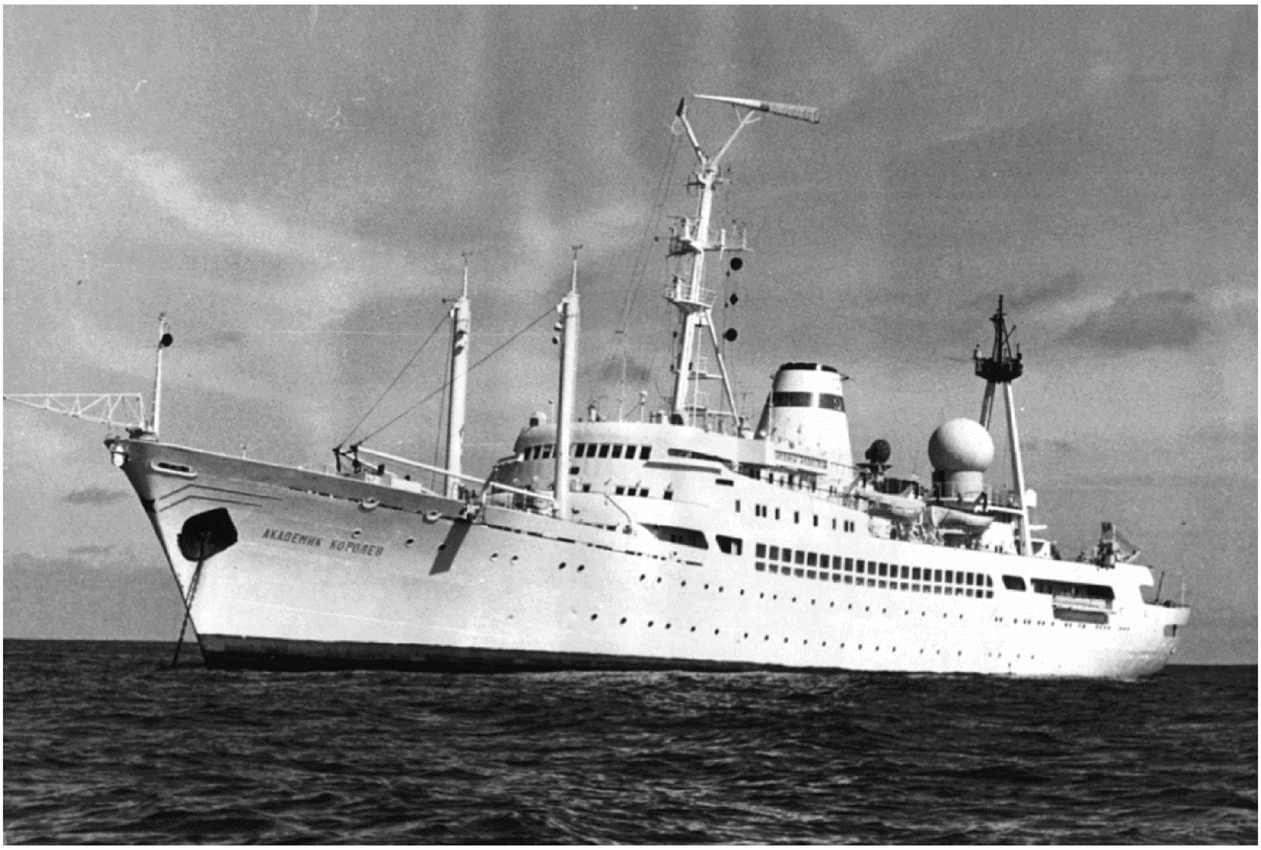


Figure 4b: Research Vessel “Academician Korolyov”.

2. Undigitized data archive

2.1 Russian fleet of research vessels

Russia has the largest fleet of research vessels (R/V) in the world. Russian sailors have been exploring seas and ceans for about two hundred years. The first Russian marine research expedition took place in 1803. Russian vessels "Nadezhda" and "Neva", commanded by captain-lieutenants Kruzenshtern and Lisyansky, had circumnavigated the globe for three years, which made it possible to collect extensive material for marine meteorology.

In Figure 1, two Russian sailing vessels currently in service are shown. They are the "Kruzenshtern" and the bark "Sedov". Vessels like these might have been used by Russian sailors to study the World ocean in the last century. Incidentally "Sedov" is considered to be the largest sailing vessel in the world. Its displacement is 7,500 tons, length of the deck 117 m, width 15 m, maximum speed 12 knots

By the mid-nineteenth century, more than 40 Russian round-the-world expeditions had been performed. In each expedition it was obligatory to make marine meteorological and hydrological observations. Figure 2 shows the most important cruises in the last century. It can be seen that the area of the World Ocean is covered by a rather dense observational network. Considering the time period, this is a substantial achievement.

At present the Russian fleet of research vessels comprises more than two hundred ships. Their sizes vary over a very wide range: from as small as "Academician Petrov" which is 80 m long with a displacement of 2,500 tons to as large as "Cosmonaut Yuri Gagarin" which is 230 m long and whose displacement is 45,000 tons (Figure 3).

The Russian fleet also comprises powerful ice-breakers, such as "Arctica" which is 140 m long and whose displacement is 23,000 tons (Figure 4). However, several tens of medium-class ships, such as R/V "Academician Korolyov" (Figure 4), form the basis of the Russian fleet of research vessels. These are the main characteristics of this vessel: length 125 m, width 17 m, displacement 7,000 tons, speed 18 knots, distance of autonomous sailing 20,000 miles.

2.2 Undigitized data from Russian R/Vs

For two hundred years of its existence the Russian R/V fleet has performed more than twenty-thousand cruises which acquired a large amount of marine meteorological data. These data are mainly contained in R/V reports, as well as in logbooks, books, collections, marine annuals and other materials. At present the WDC-B has information on 32,300 R/V cruises. These are categorized by country of registry in Figure 5 and by ship owner in Figure 6:

Figure 7 shows the distribution of the Russian R/V cruises categorized by the years they arrived at the WDC-B. The total number of cruises performed before World War Two is no more than 1,000. After World War Two the Russian research fleet resumed its activity, and the number of cruises kept increasing from year to year.

The most intensive activity of the fleet occurred in the 70s and 80s. In these years a number of large-scale experiments and programs on ocean-atmosphere interaction were conducted. Several tens of R/Vs at a time participated, and unique scientific results were obtained.

Figure 8 shows the areas of the World Ocean where the most important scientific experiments were conducted. Among these were such programs as SECTIONS (energy active zones), POLEX-SOUTH, POLEX-NORTH, TYPHOON, MONSOON and others.

Unfortunately, during the 90s the activity of the Russian research fleet decreased sharply. This was due to the

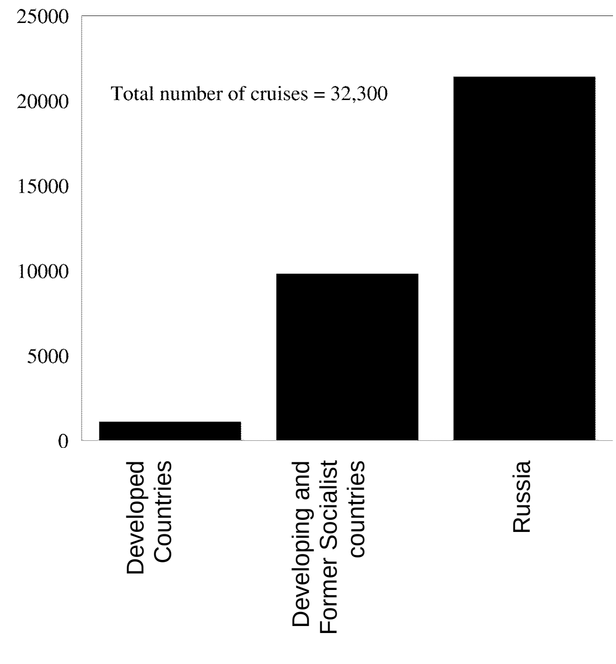


Figure 5: Number of non-digitized research vessel cruises in the WDC-B, categorized by country of registry.

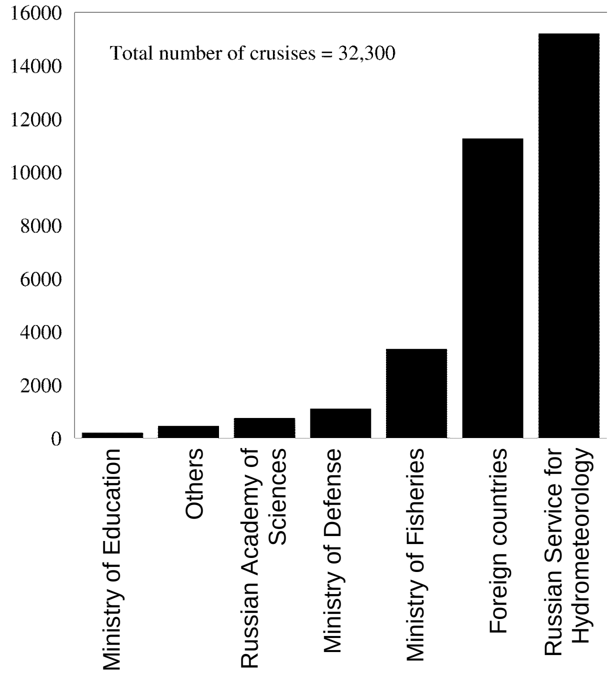


Figure 6: Number of non-digitized research vessel cruises in the WDC-B, categorized by ship owner.

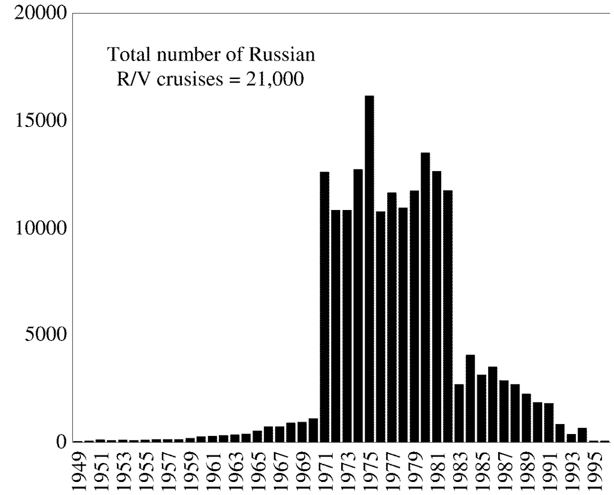


Figure 7: Number of non-digitized research vessel cruises in the WDC-B, categorized by year of cruise.

break-up of the USSR and the current financial difficulties in Russia. Hopefully, these are temporary difficulties and the situation will soon change for the better.

Next we would like to discuss the structure of the archive of Russian R/V undigitized data. On completion of each R/V cruise, a scientific report (SR) is prepared. It comprises all marine meteorological data collected during the cruise. Each SR usually contains from 200 to 2,000 observations. Each observation includes all standard meteorological parameters. The observations are normally made every three hours, but in some cases this time is reduced to every hour (e.g. when actinometric measurements are made). As was mentioned above, the total number of SRs is now 21,000.

The total number of observations in these reports is estimated to be 10 to 15 million undigitized data. About 5,000 reports are now accessible and suitable for digitization. They contain about 6,000,000 observations.

Figure 9 shows the distribution of undigitized R/V data (in percent) as relating to the areas of the World Ocean:

North Atlantic - 30%
Tropical Atlantic - 12%
South Atlantic - 3%
Indian Ocean - 10%
North Pacific - 12%
Tropical Pacific - 22%
South Pacific - 2%
Arctic - 3%
Antarctic - 2%
Seas - 4%

Of the total amount of undigitized data the World Ocean accounts for 96% and seas 4%.

2.3 Plans for data digitization and related costs

So we have 6,000,000 observations from Russian R/Vs which are now accessible and suitable for digitization. It should be noted that as compared to data from ships of opportunity, these observations are of very high quality since they were made by professional meteorologists and oceanographers during research experiments. These observations are unique indeed. If these data are so unique, why haven't they been digitized yet? First, when creating the data archive MORMET, we initially used the largest sources of data, such as GTS and logbooks from ships of opportunity. Second, when few R/V data were available, nobody cared about archiving them. When the amount of these data had increased substantially, it turned out that R/V reports had different formats and structure than logbooks, and therefore it was required that a more sophisticated procedure of digitization be developed.

This procedure was developed in 1988. It was intended to digitize future observations directly on-board; unfortunately, we couldn't bring this into practice for well-known reasons, i.e. break-up of the USSR and almost complete termination of the research fleet's activity.

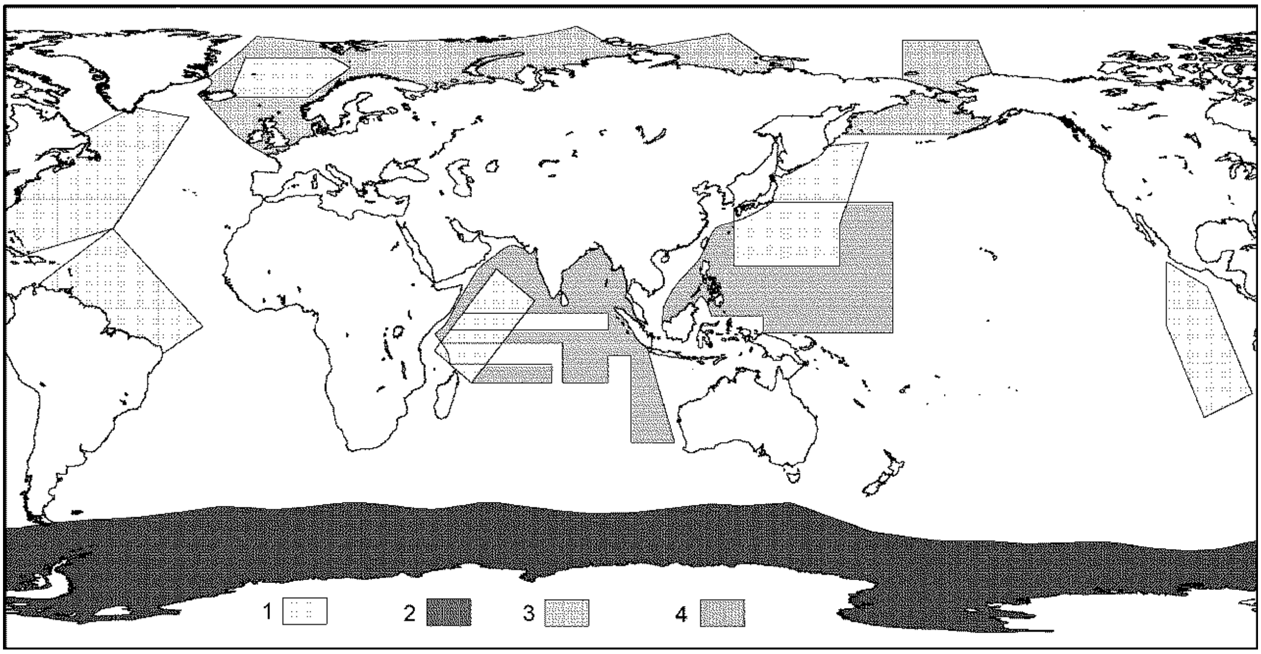


Figure 8: Selected important complex expeditions of Russian research vessels in world oceans 1-SECTIONS; 2-POLEX-SOUTH; 3-POLEX-NORTH; 4-MONSOON & TYPHOON.

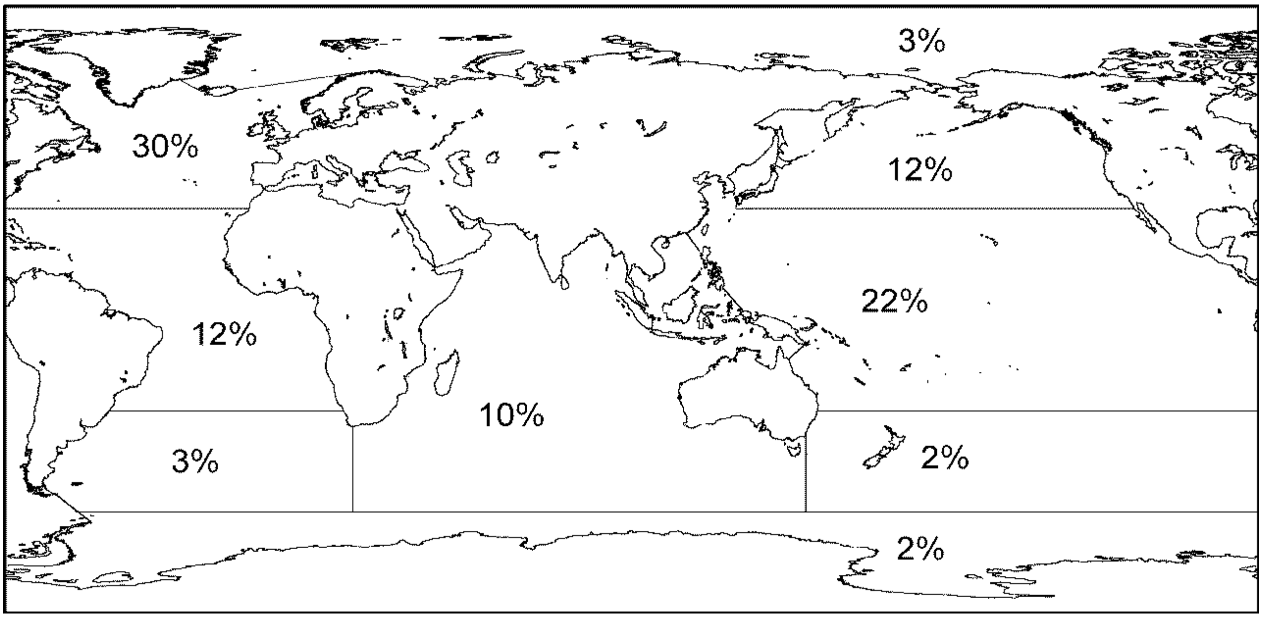


Figure 9: Percentage of non-digitized Russian research vessel observations, categorized by region.

Starting next year we are planning to begin digitizing R/V observations. Two types of digitization, manual and automated, have been evaluated. Figure 10 shows two extracts from the page of an R/V report. The upper extract is of good quality and thus is suitable for scanning. Unfortunately, the number of reports like this does not exceed 5% and some symbols appear that would require manual digitization. The quality of the lower extract is not satisfactory, so it is only suitable for manual digitization. The number of these reports is 95%. Thus we conclude that the manual method for digitizing observations from Russian R/Vs is considered principal.

To evaluate the cost of this work, let us address ourselves to international practice. In the U.S., one estimate of the cost of digitizing one observation (containing 120 symbols) is about 10 cents. Thus, digitization of 6,000,000 observations in the U.S. would require about 600,000 U.S. dollars. It should be noted that this sum includes only net salary and does not include overheads, equipment and taxes. With these costs taken into account, the sum will probably be about twice as large.

Although an imposing sum, this is the reality. In Russia the cost of data digitization is lower than in the U.S. or Western Europe. We are planning to digitize our data within four years. To do this would require 100,000 U.S. dollars annually. This sum includes salary, equipment and overheads. One-half this sum will be provided to us by the Russian service for Hydrometeorology.

As for the other half of this sum, we hope to receive it by involving our foreign partners and concerned international organizations. We are ready for cooperation with any organizations and foundations to accomplish this work in the shortest possible time. We think that 50,000 U.S. dollars annually over four years is not a large investment for the unique Russian R/V data archive containing 6,000,000 observations.

We have developed a plan to implement this project. The plan contains major objectives and tasks of the project, expected results, project period, calculation of costs, staff involved in the project and other information required. Anyone interested in the project can contact Y. Gemish for a detailed project proposal.

3. Rescue of digitized marine meteorological data

This section concerns the rescue of selected digitized marine meteorological data.

Briefly, we have several tens of magnetic tapes with a density of 800 bpi that contain about 3,000,000 observations from ships of opportunity. More than 20 years ago marine meteorological data were transferred from punch cards and punch tapes to magnetic tapes.

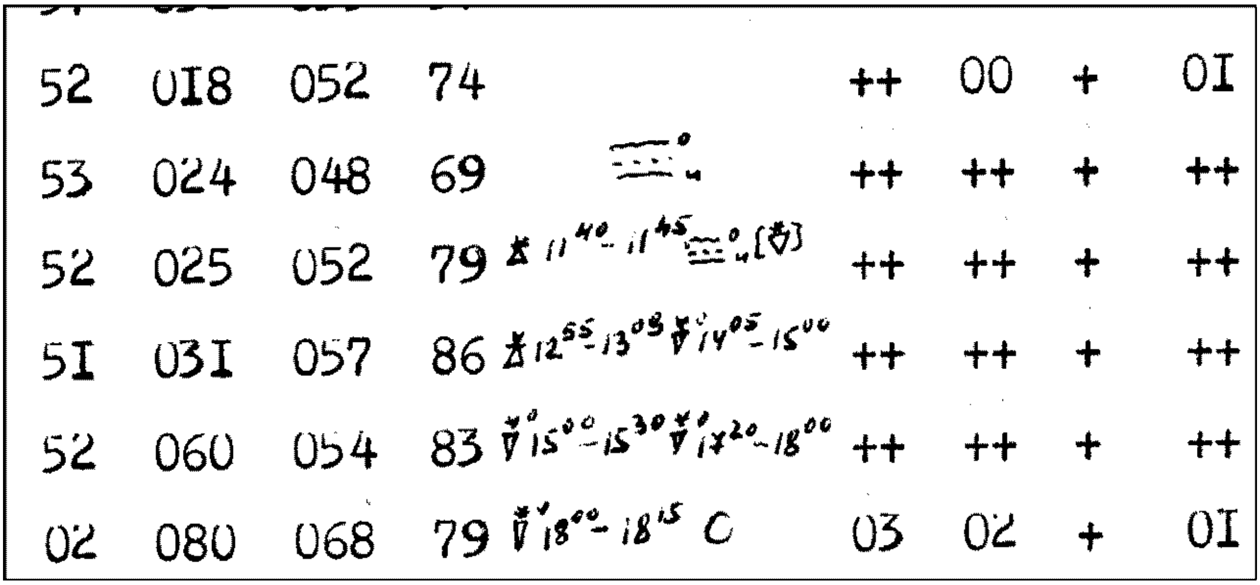
Unfortunately, numerous failures (malfunctions) now arise in reading them. Data on these magnetic tapes are recorded in different formats and are of a complicated structure. Therefore, an effort should be made to decode these data.

Although no easy matter, this work will pay for itself since we will have 3,000,000 more observations in our archive. We have developed a project to rescue this data; contact Y. Gemish for a detailed project proposal.

4. Digitization of marine actinometric observations

The next problem to be solved relates to the digitization of marine actinometric observations. The WDC-B has 700 reports of Russian R/Vs containing about 700,000 marine actinometric observations.

Each observation contains the following parameters: long-wave radiation, short-wave radiation, build-up radiation, direct radiation, back-scattered and scattered radiation, albedo, and six meteorological parameters. The observations were made every hour, sometimes every twenty minutes. These data are unique and of high scientific value.



52	0I8	052	74		++	00	+	0I
53	024	048	69		++	++	+	++
52	025	052	79	11^{40} - 11^{45}	++	++	+	++
5I	03I	057	86	12^{55} - 13^{05} 14^{05} - 15^{00}	++	++	+	++
52	060	054	83	15^{00} - 15^{30} 17^{20} - 18^{00}	++	++	+	++
02	080	068	79	18^{00} - 18^{15} C	03	02	+	0I

Figure 10a: Sample research vessel scientific log, which may produce acceptable results after electronic scanning.

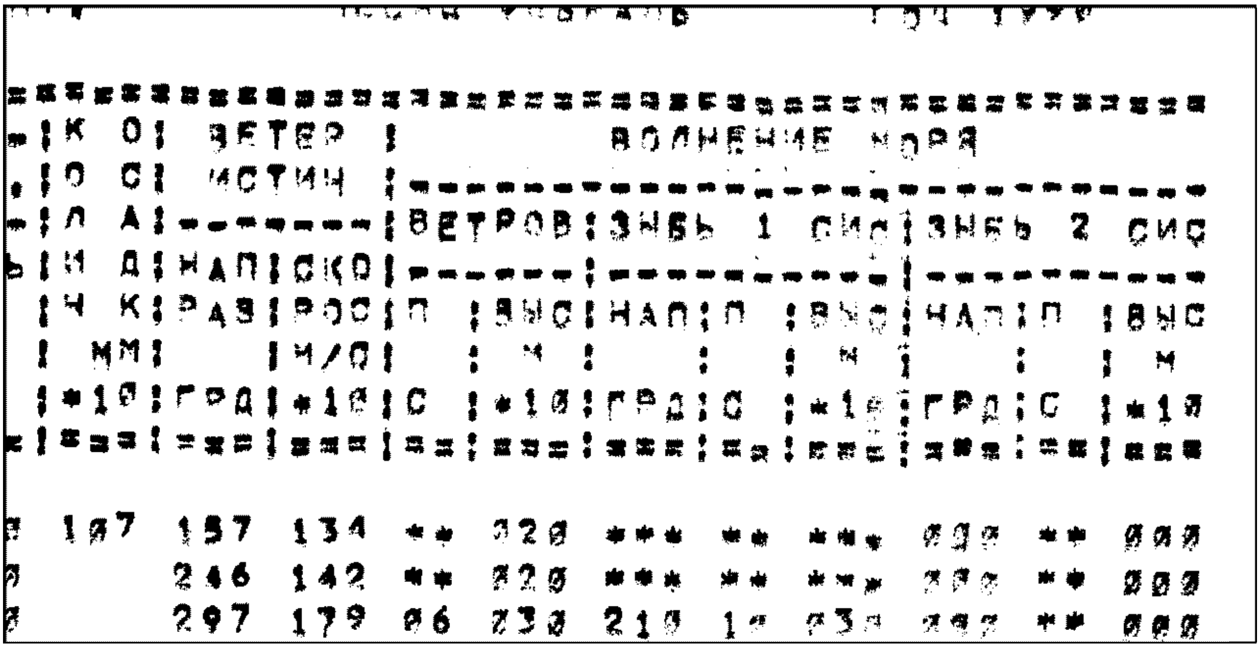


Figure 10b: Sample research vessel scientific log, which is not expected to produce acceptable results after electronic scanning.

5. Metadata

Finally, we would like to briefly discuss our work with metadata. The significance of metadata is difficult to overestimate. Metadata allow us to familiarize ourselves with a large number of sources of marine meteorological observations, and to plan our work for data digitization more efficiently. At present, the WDC-B has a reference database for marine meteorology and oceanography containing the following:

- information on R/V cruises (more than 32,000 documents);
- information on automatic buoy stations (4,500 documents);
- information on marine coastal stations (400 documents);
- information on observation platforms (850 documents);
- information on databases available from different organizations (50 documents);
- glossaries of ships, countries, organizations, geographical regions;
- parameters, instruments, etc. (more than 10,000 parameters).

At present, our database is estimated to contain no more than 20 percent of the total number of the documents. The cost of retrieval, preparation and digitization of one document is 1 - 1.5 U.S. dollars depending on the complexity of the document. We make use of the DBMS (Database Management System) Fox Pro in creating and operating the meta-database. We are now in the process of transforming this meta-database into the DBMS ORACLE for Windows.

6. Conclusions

To reiterate our major conclusions:

a) We have more than 21,000 reports from Russian R/V cruises containing 10 to 15 million undigitized marine observations. These are unique data of high scientific and applied value.

b) At present, about 6,000,000 observations are accessible for digitization. We are ready to do this work within four years. However, we need financial support. We are ready to cooperate with any interested organizations, foundations and private persons.

c) Several tens of magnetic tapes containing about 3,000,000 observations need rescuing. This work also requires technical and financial support.

d) A unique data archive with about 700,000 marine actinometric observations also awaits digitization.

e) We have created a meta-database for marine meteorology and oceanography that contains more than 30,000 documents, which comprises 20 percent of the total number of documents available. This work is being continued.

Thus, it can be seen that we are facing major challenges. We hope that international cooperation in marine data digitization will allow these problems to be solved on short notice.

---oooOooo---

Ukranian Historical Surface Marine Data, Quality Control and Interpolation

A. Polonsky
Marine Hydrophysical Institute (MHI), Sevastopol, Ukraine

1. Introduction

The goals of this paper are to discuss:

- The possible sources and value of surface marine data in the Ukraine (section 2), focusing on Marine Hydrological Institute (MHI) data.
- The quality control procedure (section 3), and the accuracy of a procedure for interpolation (section 4).

2. Sources and value of data

There are four principal organizations in Ukraine that have significant surface marine archives, namely:

- Former Odessa Department of the State Oceanographic Institute (SOI).
- Marine Hydrophysical Institute (MHI) of the Ukrainian Academy of Sciences.
- Former Sevastopol Department of the SOI.
- Southern Research Institute of Fishery and Oceanography (SRIFO, Kerch).

Some Ukrainian data also are archived at the Obninsk Center of Hydrometeorological Information (Russia). The MHI hydrometeorological archive consists of routine observations all over the World from 1960 to 1985 (Table 1).

From 1986 until 1991 the hydrometeorological data were sent to the World Data Center (Obninsk) on magnetic tape. Most historical MHI data are undigitized (except for in the Black Sea region). MHI has the potential ability and qualified personnel for digitization and quality control.

A very complete data set exists in the former Odessa branch of SOI (Table 2). It was obtained mostly in the North Atlantic region. Significant portions of the data are not digitized. The former Sevastopol branch of SOI has the regional Black Sea archive. This contains about 60,000 hydrographic stations. Data from about 8,000 stations have been digitized. SRIFO possesses data from a few thousand hydrographic stations all over the World in fisheries regions. Most of them are not digitized. All of the above marine observations were taken by professional hydrometeorologists.

3. Quality control

Usually quality control consists of the following steps:

a) Double digitization for preliminary quality control.

b) Quality control of coordinates. Repetitive (duplicate) ship data are rejected.

c) Quality control of gross errors. This includes a comparison of each measured parameter with a monthly climatological value. If the difference exceeds some defined magnitude, the measured parameter is marked and carefully checked. This magnitude varies depending on parameter, region, and season. For instance, for SST in the Tropical Atlantic we use a 5°C range.

d) The next step is statistical quality control using criteria of 3 or 4 r.m.s., if the number of data is sufficient for calculation of a monthly r.m.s. After that an interpolation over regular grid points is performed.

Table 1
Inventory of marine hydrometeorological observations (archive of the MHI Sevastopol, preliminary information).

Cruise number	Dates	Number of observations	Region
"Ac. Vernadsky"			
2	16.02-27.04.1970	40 (+46?)	Med. Sea
	04.06-03.09.1970	67 (+87?)	Atl. Oc.
3	23.10-10.02.70-71	54	Atl. Oc.
4	30.06-27.11.1971	89	Pacific, Atl. Oc.
5	27.12-15.05.71-72	45	Pacific, Ind., Atl. Oc.
	11.07-11.08.1972	62	Black Sea
6	17.09-23.01.72-73	89	Atl. Oc.
7	17.03-07.04.1973	42	Black Sea
	26.04-18.08.1973	55	Eq.Atl. Oc.
8	03.09-24.09.1973	25 (+38?)	Black Sea
	08.12-17.04.73-74	68	Atl. Oc.
9	09.10-30.11.1974	53	Med. Sea
10	21.12-20.05.74-75	56	Eq.Ind. Oc.
11	29.05-17.09.1975	32	Eq.Ind. Oc.

12	18.12-16.04.75-76	30	Eq.Pacific Oc.
13	18.06-12.08.1976	70	Med. Sea
14	03.09-20.12.1976	78	Atl. Oc.
	20.03-29.03.1977	26	Black Sea
15	05.04-18.06.1977	65	Trop.Atl. Oc.
16	05.07-05.11.1977	86	Atl. Oc.
17	25.11-25.03.77-78	47	Atl. Oc.
18	26.05-06.08.1978	27	Atl. Oc.
19	21.12-10.04.78-79	5	Ind. Oc.
20	24.04-12.08.1979	266 (+342?)	Eq.Atl. Oc.
21	24.08-10.10.1979	78	Eq.Ind. Oc.
22	01.03-14.07.1980	144	Ind. Oc.
23	05.11-04.12.1980	61	Black Sea
	30.12-29.05.80-81	51	Trop.Atl. Oc.
25	30.12-10.05.81-82	267	Trop.Atl. Oc.
26	06.07-18.11.1982	198 (+275?)	Trop.Atl. Oc.
27	29.12-07.05.82-83	33	Black Sea
	25.05-06.06.1983	236	Trop.Atl. Oc.
28	29.11-14.12.1983	24	Black Sea
	30.12-13.05.83-84	43	Trop.Atl. Oc.
29	15.06-15.07.1984	23	Black Sea
“**M. Lomonosov**”			
7	12.01-15.04.1960	131	Atl. Oc.
8	05.06-22.09.1960	67	Atl. Oc.
17	16.11-28.03.64-65	68	Atl. Oc.
18	30.09-12.01.65-66	63 (+69?)	N.Atl. Oc.
19	29.04-10.08.1966	66	N-W.Ind. Oc.
20	01.11-28.05.66-67	62	Atl., Ind., Pacific Oc.
21	28.04-24.08.1968	92	Atl. Oc.
22	25.12-24.04.68-69	73	N.Atl. Oc.
23	24.06-28.08.1969	27 (+106?)	Med. Sea
24	11.12-10.04.69-70	41	Trop.Atl. Oc.
25	29.07-27.10.1970	40	Med. Sea
26	09.03-24.06.1972	56	Trop.Atl. Oc..
	28.06-27.08.1972	25	N.Atl. Oc.
27	14.12-23.04.72-73	66	Eq.Atl. Oc.
28	22.01-05.05.1974	45	Trop.Atl. Oc.
29	05.06-21.10.1974	36 (+103?)	Trop.Atl. Oc.
30	17.04-28.08.1976	57 (+75?)	Trop.Atl. Oc.
31	16.10-23.02.76-77	35	W.Atl. Oc.
32	04.06-30.09.1977	119	S.Atl. Oc.
33	16.10-13.02.77-78	82	W.Atl. Oc.
34	18.03-18.06.1978	34	Atl. Oc.
36	01.09-05.11.1978	18	N-E.Trop.Atl. Oc.
37	29.12-08.04.78-79	49	Atl. Oc.
38	17.10-28.02.1980	76	W.Atl. Oc.
39	02.04-30.08.1980	104	Ind. Oc.
40	22.10-31.12.1980	27 (+33?)	Eq.Atl. Oc.
41	11.02-12.05.1981	16	Eq.Atl. Oc.
42	04.07-16.09.1981	24	E.Atl. Oc.
44	08.06-10.07.1985	69	Black Sea
“Prof. Kolesnikov”			
3	04.07-07.08.1979	44	Black Sea
	21.03-09.06.1982	92	Black Sea

Table 2
Inventory of surface marine information (former Odessa branch of State Oceanographic Institute).

Region	Number of obs.	Digitization status	Additional information
Odessa-Gibraltar	30,000		
Standard sections in the North Atlantic	49,000		
OWS “C”	130,000		
Bermuda	7,000		
Tropical	3,000		
Newfoundland	11,000		
Others	2,500		
Total	232,500	Only a small amount of information for 1986 to 1991 was digitized.	There are undigitized actinometric data (total short-wave radiation and balance observations).

Table 3
Information about the GATE data set in June-September 1974 (nd = no data).

Observational area number	Latitude (°N)		Longitude (°W)		Number of soundings		
	min	max	min	max	1st phase	2nd phase	3rd phase
0	8.0	9.0	24.0	23.0	137	245	533
1	10.0	11.0	27.5	26.5	138	147	136
2	11.5	12.5	24.0	23.0	152	141	160
3	10.0	11.0	20.5	19.5	152	135	163
4	6.0	7.0	20.5	19.5	117	130	139
5	4.0	5.0	24.0	22.0	58	64	72
6	6.0	7.0	27.5	26.5	158	131	161
7	10.0	10.5	24.0	23.0	150	nd	nd
8	9.0	9.5	25.0	24.0	124	88	81
10	7.0	7.5	24.0	23.0	137	120	73
11	7.5	8.5	22.5	20.5	32	538	159
12	8.5	9.5	23.0	21.5	141	145	296

4. Interpolation

Let us demonstrate the accuracy of an interpolation procedure using data from the Global Atlantic Tropical Experiment (GATE). This is a unique tropical oceanographic data set. The CTD measurements were performed concurrently at 13 observational areas in the Tropical Atlantic typically every 3 hours from June to September 1974 during 3 phases of the experiment (Table 3).

The typical duration of each phase of the GATE was about 20 days. The minimum depth of soundings was 800 m (Deep Sea Res., 1980). Mainly, however, we analyzed data sets for the upper 250 m layer and restricted our consideration to the following standard levels: 0, 10, 20, 30, 40, 50, 75, 100, 125, 150, 200, and 250 m.

We separated the proportion of total temperature variance due to hour-to-hour and day-to-day fluctuations for each observational area. Then we estimated the minimum number of independent measurements necessary for the calculation of an average monthly temperature with a specified accuracy for each area. Finally, we approximated the spatial correlation function of temperature and calculated the optimal weights of measured temperature for use with a $2° \times 4°$ interpolation procedure.

The results of optimal and linear interpolations are compared below. We used a tropical data set from MHI for comparison, containing more than 35,000 CTD soundings performed in the Tropical Atlantic from 1911 until 1990 (Gaysky et al., 1992).

The main GATE long-term CTD survey was situated in the vicinity of the zonal North Equatorial Counter Current (NECC). South/North of NECC, the thick/thin and relatively warm/cold upper mixed-layer (UML) occurs. Within the UML, hour-to-hour variability of temperature dominates in the temporal band ranging from one hour to one or two months; its proportion of the total variance in this band reaches 60%. Below the UML (in the thermocline), day-to-day variability dominates; the proportion of hour-to-hour variability with respect to total temperature variance declines to about 30%. From 10 to 15 independent measurements are necessary to calculate the average monthly temperature with an accuracy less than r.m.s.

Thus, one must perform at least a few daily measurements of the UML temperature, and 10 to 15 measurements of the thermocline temperature per month to obtain a rather accurate average monthly value for each area.

Nobody has enough CTD soundings in the whole Tropical Ocean to satisfy this requirement. Therefore, spatial interpolation is necessary for high-amplitude noise reduction. Optimal interpolation gives the minimum r.m.s. error, but a knowledge of the correlation function is necessary. It is clear that strong anisotropy occurs in the NECC vicinity.

For example, we received the following analytical approximation of the zonal (R_x) and meridional (R_y) correlation function at 75 m depth (which is typical for the thermocline):

$$R_x = \exp(-0.95|x|) * \cos(0.01x) \quad (1)$$

$$R_y = \exp(-0.37|y|) * \cos(0.01y) \quad (2)$$

Table 4

Optimal weight coefficients (each step equals one degree).

Latitude distance	Longitude distance				
	-2	-1	0	1	2
-1	0.00	0.02	0.04	0.02	0.00
0	0.01	0.41	1.00	0.41	0.01
1	0.00	0.02	0.04	0.02	0.00

Table 5

Number of MHI July temperature observations at 75 m depth (nd = no data).

Lat (°N)	Long (°W)									
	28	27	26	25	24	23	22	21	20	19
13	nd	nd	nd	nd	nd	nd	nd	2	1	1
12	nd	2	2	nd	2	2	nd	2	1	nd
11	nd	2	2	nd	14	2	nd	54	19	nd
10	nd	2	2	nd	3	2	nd	14	2	nd
9	nd	2	2	nd	3	2	nd	2	2	nd
8	nd	2	2	nd	70	2	nd	53	2	nd
7	nd	2	2	nd	3	2	nd	12	9	nd
6	nd	2	2	nd	3	9	1	41	2	nd
5	1	2	2	nd	4	10	nd	27	2	nd
4	nd	2	2	nd	4	2	nd	2	2	nd

Table 6

Absolute error of linear interpolation of temperature at 75m depth.

Lat (°N)	Long (°W)									
	28	27	26	25	24	23	22	21	20	19
13	0.00	0.01	0.01	0.00	0.04	0.05	0.08	0.18	0.09	0.74
12	0.06	0.15	0.23	0.26	0.17	0.13	0.36	0.35	0.13	0.45
11	0.19	0.21	0.20	0.45	0.36	0.18	0.37	0.27	0.15	0.15
10	0.62	0.58	0.63	0.37	0.22	0.27	0.30	0.66	0.39	0.11
9	0.06	0.19	0.27	0.46	0.09	0.33	0.23	0.35	0.20	0.31
8	0.49	0.50	0.53	1.23	1.24	0.26	0.30	0.21	0.04	0.08
7	0.04	0.74	0.51	0.25	0.41	0.27	1.25	2.26	1.43	0.39
6	0.46	0.99	1.22	0.43	0.07	0.09	1.72	1.68	1.90	0.08
5	0.64	0.16	1.43	0.42	0.67	1.27	0.96	0.26	0.12	0.62
4	2.87	1.62	0.39	1.00	0.66	0.29	0.88	1.25	0.19	0.39

Table 7

Absolute error of isotropic exponential interpolation of temperature at 75m depth.

Lat (°N)	Long (°W)									
	28	27	26	25	24	23	22	21	20	19
13	0.00	0.05	0.05	0.00	0.10	0.22	0.02	0.07	0.06	0.33
12	0.02	0.05	0.18	0.08	0.14	0.27	0.10	0.09	0.09	0.67
11	0.14	0.10	0.09	0.05	0.20	0.32	0.05	0.21	0.07	0.62
10	0.20	0.37	0.18	0.05	0.32	0.38	0.12	0.21	0.20	0.06
9	0.02	0.07	0.15	0.17	0.61	0.45	0.01	0.21	0.09	0.07
8	0.09	0.28	0.25	0.23	0.18	0.69	0.02	0.06	0.04	0.04
7	0.02	0.46	0.09	0.12	0.52	0.01	2.02	0.05	0.81	0.19
6	0.45	0.48	0.64	0.01	0.08	0.74	0.72	1.21	0.76	0.25
5	0.69	1.46	0.67	0.07	0.57	0.03	1.42	0.17	0.18	0.20
4	2.62	0.84	0.95	0.09	0.17	0.31	0.20	0.11	0.65	0.37

Table 4 gives the optimal relative weights due to formulae 1-2. Let us compare the results of optimal interpolation of the MHI July temperature at 75 m depth, with the results of linear interpolation used by Dzhiganshin and Polonsky (1992). The number of MHI July observations in $1° \times 1°$ degree squares is given in Table 5. We will consider the results of optimal interpolation as the truth, and their differences with the results of linear interpolation as interpolation errors.

It is clear from Table 6 that these errors are usually not large in comparison with the amplitude of the seasonal cycle (typically 2 to 5 degrees; ref., Dzhiganshin and Polonsky, 1992) if the number of observation within the radii of interpolation is more than 10 to 15.

Its typical relative magnitude is 20% of the seasonal cycle amplitude. However, this error may be more than 50% in the vicinity of the NECC core between 4° and 6°N because of strong anisotropy, which is not taken into consideration as part of the linear interpolation procedure. The strong temperature anisotropy in the equatorial Atlantic is clear in Table 7, which gives the absolute error of isotropic exponential interpolation of Sukhovey (1977), who implies that:

$$R_x = R_y = R * \exp(0.661 * \cos(0.011)) \qquad (3)$$

Errors of interpolation may be large also in regions with small numbers of observations as a result of high-amplitude noise. The above estimates are typical for summer in the North Tropical Atlantic because 1974 was not an anomalous year (Deep Sea Res., 1980).

5. Conclusions

The Marine Hydrophysical Institute would be able to organize and to perform the digitization and quality control of the Ukrainian marine observations, with some additional funding. We have qualified personnel and experience in dealing with historical marine data sets. We believe that data taken by professional hydrometeorologists should receive a priority for digitization.

References

Deep Sea Research, 1980: *GATE Supplement 2*, 354 pp.

Dzhiganshin, G.F., and Polonsky, A.B., 1992: *Marine Hydrophysical Journal*, No. 5, 58-63 (in Russian).

Gaysky, V.A., Polonsky, A.B., and Chmutov, M.I., 1992: *Collected Papers: Automated Control of the Hydrophysical Fields*, Sevastopol, 134-147 (in Russian).

Merle, J, and Arnault, S., 1985: *J. Marine Res.*, **43**, 267-288.

Sukhovey,V.F., 1977: *Variability of the Hydrophysical Fields of the Atlantic Ocean*, Hydrometeoizdat, Leningrad, 258 pp. (in Russian).

---oooOooo---

Surface Marine Data Holdings of the Public Record Office, UK

Chris Millard
Public Record Office, UK

1. Introduction

The Public Record Office (PRO) is the Government Archive for England, Wales and the United Kingdom. It runs the public record system of the UK by selecting and preserving the records of the state and making them available to the public and to government. Holdings span 1000 years dating from before the Domesday Book right up until modern government records. These records take up 160 shelf kilometers at the offices at Kew. Access to documents is on site at the Kew Office in most cases. Access restrictions apply to modern records, the standard closure date being 30 years but with some more sensitive documents being closed for 75 or 100 years. In addition some documents have been released early under the Open Government Initiative.

Access restrictions apply to certain important or delicate documents and many of these documents have been put into surrogate form (usually microform.) This is also true of certain popular classes of documents. Copies of documents (subject to approval by conservation) can be provided as can microfilm copies of documents already on microfilm.

The main sets of records providing data relating to surface marine data are of course ships' logs. There are two sets of these records: those of the Merchant Marine (through the Registrar General of Shipping and Seamen), and those of the Admiralty.

2. Merchant ships' logs

The Registrar General of Shipping and Seamen required the masters of merchant vessels to keep ships' logs for all voyages. However with the introduction of seamen's tickets it was generally thought to be unnecessary for these logs to be preserved. The only selection of records that survive in the Public Record Office are those from the period 1902 to 1919. These however do not give the wealth of information that those of the Admiralty do. The information was filled out on a standard booklet supplied by the Registrar General's Office. They are mainly lists of events during the voyage largely to do with the conduct of the crew etc. They give few details of the voyages or of meteorological conditions.

3. Royal Naval ships' logs

Ships' logs compiled by masters and captains of Royal Navy ships must have been filled out from a very early date. Most of those preserved in the Public Record Office run in three partly parallel series: Masters', Captains' and Ships' Logs. An additional series is ADM 55, Exploration Logs, consisting of volumes extracted from these classes. Admiralty orders, regularly re-issued in the Naval Instructions in record class ADM 7, Admiralty Miscellanea, lay down the regulations for the compiling of logs and journals from admiral through to master.

In the 1746 Instructions (ADM 7/201) captains are instructed in Article XL (50) to "...keep a journal, according to the form set down...and to be careful to note therein all occurrences viz place where the ship is at noon, changes in wind and weather, Salutes with reasons thereof; remarks on unknown places and in general every circumstance that concerns the ship…."

In 1808 masters were instructed in Article XXXII "...in form 25 to enter the weather, directions of the wind, the courses steered, distances run, with every occurrence relating to the navigation of the ship; the setting and the velocity of currents and the results of all astronomical observations made…."

The series of records are:

ADM 55 Supplementary Logs Series II: Explorations
Dates: 1757 to 1904
Quantity: 164 Volumes
Copies of these are held on microfilm. These include all of the major voyages of discovery carried out by Royal Naval vessels.

ADM 53 Ships' Logs
Dates: 1799 to 1967
Quantity: 166,777 Volumes, Log Books and Booklets

ADM 51 Captains' Logs
Dates: 1669 to 1852
Quantity: 4,563 Volumes

ADM 52 Masters' Logs
Dates: 1672 to 1840
Quantity: 4,660 Volumes

ADM 54 Supplementary Logs Series I: Masters' Logs
Dates: 1808 to 1870
Quantity: 337

ADM 7 Miscellanea
Contains various early ships' logs from the 17th and 18th centuries

From the introduction of Ship's Logs (ADM 53) the series run consecutively and there is a large amount of replication of information between the classes. The survival of the additional classes does however allow for the filling in of gaps in the information of the main series.

3.1 Content

The format of the information, not surprisingly given the large date range, varies considerably. Additionally the information takes the form of logs and journals for the earlier periods in the Captains' Logs series. The journals repeat some of the information provided in the logs, providing a more discursive account of the events of the particular day, and slightly skewing estimates of the numbers of documents. In addition the period covered by each log varies considerably making estimates about the exact amount of information contained very difficult.

In the logs the most consistent pieces of information are those relating to the position of the ship, usually readings taken on the hour, and those concerning the weather conditions and direction of the prevailing wind at the time. They also provide large amounts of information on life in the ships of the Royal Navy, particularly punishment.

The standard information recorded in the printed ships' log form is divided into two sections, information recorded hourly and information recorded daily. Hourly information is Hour, Knots, Fathoms, Courses, Winds, Remarks (recording significant events and often weather conditions). Daily information is Course, Distance, Latitude, Longitude made, Thermometer, Bearing and Distance at Noon.

Information does obviously vary from captain to captain and master to master within this framework. This is also true over time with the exact form of information changing; however the essential elements do remain throughout the whole period. Journals, which appear in some cases within this class as noted above, have less formalized information and, as you might expect, rather more discursive text.

Included with some of the exploration logs in ADM 55 was the ships' meteorological register. Meteorological Officers (where applicable) were similarly required by the Naval Instructions to keep a register. These contain further information including more detailed meteorological information. It is possible that these may in some cases survive in the main series although their inclusion is not standard; it may however be a peculiarity of the logs of ships on voyages of discovery.

3.2 Geographical coverage

Throughout the 18th, 19th and early 20th centuries, Royal Navy ships were stationed all over the world. In addition many voyages of discovery were undertaken by the Navy. Data therefore is available from these sources at various periods for almost anywhere in the world. The coverage will obviously decrease with the reduction of size of the Navy particularly as the 20th century progressed.

3.3 Format

The format of the documents themselves again varies over time. Some of the earlier documents, particularly, Captains' and Masters' Logs, are in bound journals. Other logs take the form of folded letters; later logs are in the form of booklets.

Earlier bound journals and papers are sometimes hand-drawn to accommodate some or all of the required information, and sometimes printed. The standard form that the information should be entered in is printed in the Naval Instructions. The later booklets are standard, drawn up and printed for the Admiralty.

All entries in the documents to the present day were written in by hand and signed by the relevant ship's officers. This obviously presents certain problems for digitization.

3.4 Organization

The records are organized first into the various classes, and then into periods and within these alphabetically by the ship's name and exact dates covered. Once the information contained in lists has been digitized as part of our AD2001 program to make our lists available online it will be much easier to make estimates about the amount of information held. The lists do not describe the nature and location of voyages undertaken or duty served. In order to establish the geographical area covered by each piece it would be necessary to examine other Admiralty sources, some of which are described below.

4. Other Admiralty sources

4.1 Admirals' Journals

ADM 50 Admirals' Journals
Dates: 1702 to 1916
413 Volumes

These journals sometimes contain information similar to that in the logs (i.e. relating to position and weather conditions). But this is hardly true in all cases; much of the information relates to orders given to the fleet and exchanges with the Admiralty.

These journals do provide information about the ships under the command of, and orders given by, the admiral. This is particularly useful if the admiral was a station commander. In this case the journals can provide an idea, for that area and period, of which types of logs from specific ships are relevant, and may be available in other records classes.

4.2 Station records

"General Station" records (e.g. North American and West Indies Stations) can provide information similar to that listed above in admirals' journals. "Pacific Station" records (ADM 172/2) were compiled on the Pacific Station from hydrographic reports of ships' officers dated between 1845 and 1858.

4.3 Muster lists

These provide the names of ships and their stations. They are station by station accounts of vessels including those which have been sent home. This is very useful for identifying ships of interest for particular areas. These are held in records classes ADM 7 and ADM 8.

4.4 Papers of Robert Fitzroy

Class BJ 7 contains the papers of Robert Fitzroy, who was hydrographer of the Navy from 1829-1836. Most of these papers at the PRO actually relate to his tenure as superintendent of the Meteorological Office at the Board of Trade (i.e. after 1854). His correspondence with Admiral Moorson is contained in piece BJ 7/1. These records may not provide the same kind of raw data as other records discussed above, but still may be of interest. The main body of Fitzroy's papers are of course held at the Hydrographic Office (part of the Ministry of Defense) in Taunton, Somerset.

5. Digitization

5.1 Difficulties of digitizing the records

There are several difficulties with the prospect of digitizing the Public Record Office's holdings as described. The first is that the sheer volume of material makes the task prohibitive. The second is identifying the material that is relevant to a particular study if a smaller regional or periodic study is to be attempted. The nature of the documents may also be a problem, with variations in period covered and of information recorded making a formulaic approach difficult.

Further problems may be presented by the physical nature of the documents; the earlier documents may be fragile and difficult to handle because of their binding and state of preservation. The chief problem would however seem to be that these documents are almost all handwritten. Therefore, while it maybe practical to produce digital images of the document for wider dissemination, the data within them could not readily be made usable by present-day optical character-recognition techniques.

Therefore for any kind of analysis, the information contained in these documents would have to be hand keyed into a machine-readable form.

5.2 Priorities of the Public Record Office

At the present time much of the resources of the PRO are going into the AD2001 Project (mentioned above) which will make all of the Public Records Office's standard catalogue available online by the year 2001. Some resources are being put into digitization at the moment but these are largely restricted to our most popular and widely used classes. These do not include, at the moment, any of the classes discussed here.

6. Conclusions

The PRO has large holdings of material containing surface marine data, although the exact nature does vary over time. Digitization of the material is not imminent. However, subject to conservation concerns and approval, suggestions for digitization projects are welcome.

---oooOooo---

Digitization of UK Merchant Ships' Data for 1935-1939

David Parker and Michael Jackson
Hadley Centre, Meteorological Office, Bracknell, UK

A total of 458,766 records in merchant ships' logbooks (METFORMS), covering the period 1935 to 1939, were found in the UK Met. Office archive, and digitized. The cost was about 10 pence (US $0.16) per record. Data were keyed by Atlantic Data Services Ltd., Blandford, Dorset, UK. Header and data records were created (Tables 1 and 2) and include metadata such as ship's name, voyage, and height of barometer. Basic quality control included removal of illegal characters, and checks that no logbooks had been omitted.

Table 1
Format for header records.

Column number: format #1	Column number: format #2	Type of data	Description of data
1	1	Record type	1
2 - 6	2 - 6	Folio number	33001 to 43000
7	7	Folio number (continued)	Blank, A, B, or C, etc.
8 - 14	8 - 14	Ship type	Steam
15 - 34	15 - 34	Ship name	
35 - 54	35 - 54	Captain's name	
55 - 74	55 - 74	Voyage from	
75 - 94	75 - 94	Voyage to	
95 - 97	95 - 97	Measurement by	Log or revolutions
98 - 102	98 - 102	Loading	Light or deep
103	103	Propeller immersed	Yes or no
104 - 106	104 - 106	Barometer	Mercury or aneroid
107		Not used in Format 1	
108 - 112	107 - 111	Barometer error too high	
113 - 117	112 - 116	Barometer error too low	
118 - 122	117 - 121	at barometer reading	
123 - 127	122 - 126	Height of barometer	
128 - 129	127 - 128	Day	
130 - 131	129 - 130	Month	
132 - 135	131 - 134	Year	

Observers left the whole inches of pressure blank if it was unchanged from the previous observation. These data were keyed in 100ths inch as written, but users can readily restore the full value in 100ths inch using:

IF (VALUE < 100) ADD (PREVIOUS VALUE/ 100)*100,

because of the truncation done by FORTRAN integer arithmetic. Some ships reported in millibars and tenths, or in whole millibars, but simple coding will again restore the full value. Temperature was usually measured in Fahrenheit, often in Celsius and very occasionally in Kelvin. The units were whole numbers or tenths, and simple coding can process these. Directions, route, wind, sea and swell are reported numerically in degrees and also alphabetically in the 32-point compass, requiring additional coding.

Maps of counts of observations from the UK Meteorological Office Marine Data Bank (MOMMDB) and from the newly digitized data were created (e.g. Figures 1 and 2).

The newly digitized data give significantly improved coverage on the routes from Panama via Papeete to Wellington (New Zealand) and from Aden via Colombo to Fremantle (Australia), and in the North Atlantic north of 50°N (Figure 3 and Tables 3a and 3b). Owing to the second world war, the data cease in late 1939.

Table 2
Format for data records.

Column number: format #1	Column number: format #2	Field *	Type of data	Description of data
1	1	N0	Record type	2
2 - 6	2 - 6	N0	Folio number	33001 to 43000
7	7	A0	Folio number (continued)	Blank, A, B, or C, etc.
8 - 9	8 - 9	N0	Year	35 etc.
10 - 11	10 - 11	N0	Month	01 to 12

12 - 13	12 - 13	N0	Day of month	01 to 31
14 - 16	14 - 16	A0	Day of week	MON to SUN
17 - 18	17 - 18	N0	Hour	00 to 24
19 - 22	19 - 22	N	Latitude	000 to 900
23	23	A0	Hemisphere	N or S
24 - 28	24 - 28	N	Longitude	0000 to 1800
29	29	A0	Hemisphere	E or W
30 - 33	30 - 33	N	Course	000 to 360
34 - 35	34 - 35	N0	Speed	00 to 99
36 - 39	36 - 39	NAR	Wind direction	N to N/W (=N by W), 000 to 360, CALM, VAR
40 - 42	40 - 42	NAR	Wind force	000 to 012 or ./. (= missing)
43 - 46	43 - 46	N0	Pressure: as read	0000-0520, 2700-3200, 9500-9999
47 - 48	47 - 49	N	Attached thermometer	30 to 99 °F or -5 to 40 °C
49 - 52	50 - 53	N0	Pressure: true	0000-0520, 2700-3200, 9500-9999
53 - 54	54 - 56	N	Air temperature	0 to 99° F or -20 to 45 °C
55 - 56	57 - 59	N	Sea surface temperature	28 to 99° F or -2 to 40 °C
57 - 59	60 - 62	NA	Present weather	Alpha or 0 to 9
60 - 62	63 - 65	NA	Past weather	Alpha or 0 to 9
63 - 65	66 - 68	NA	Visibility	0 to 9 or ./.
66 - 69	69 - 72	A	Low cloud	Alpha
70 - 71	73 - 74	N	Amount of low cloud	0 to 10
72 - 75	75 - 78	A	Middle cloud	Alpha
76 - 79	79 - 82	A	Upper cloud	Alpha
80 - 81	83 - 84	N	Total cloud amount	0 to 10
82 - 85	85 - 88	A	Sea wave direction	N to N/W
86 - 88	89 - 91	NA	Sea wave amount	0 to 9 or ./.
89 - 92	92 - 95	A	Swell wave direction	N to N/W
93 - 95	96 - 98	NA	Swell wave amount	0 to 9 or ./.
96	99	N0	Remark indicator	

* These descriptions are a guide to how the field is formatted:
N - Numeric characters
A - Alphabetic characters
R - Right adjusted
L - Left adjusted
0 - Value fills the field

Table 3a

Annual concentrations along three shipping routes for the years 1935 to 1938: MOMMDB/METFORMS.

Routes *	1935	1936	1937	1938
1	100/40	80/120	60/130	60/120
2	100/150	150/300	200/250	10/250
3	300/100	400/600	300/650	400/800

Table 3b

Monthly concentrations along three shipping routes for 1939: MOMMDB/METFORMS.

Routes *	Jan	Mar	May	Jul	Sep	Nov
1	5/8	5/10	6/8	8/10	6/0	4/0
2	6/20	8/25	8/15	6/13	4/3	3/0
3	20/50	20/50	20/80	20/80	3/0	4/0

* Key to shipping routes:
1. Panama to Papeete to Wellington
2. Aden to Colombo to Fremantle
3. North Atlantic north of 50°N

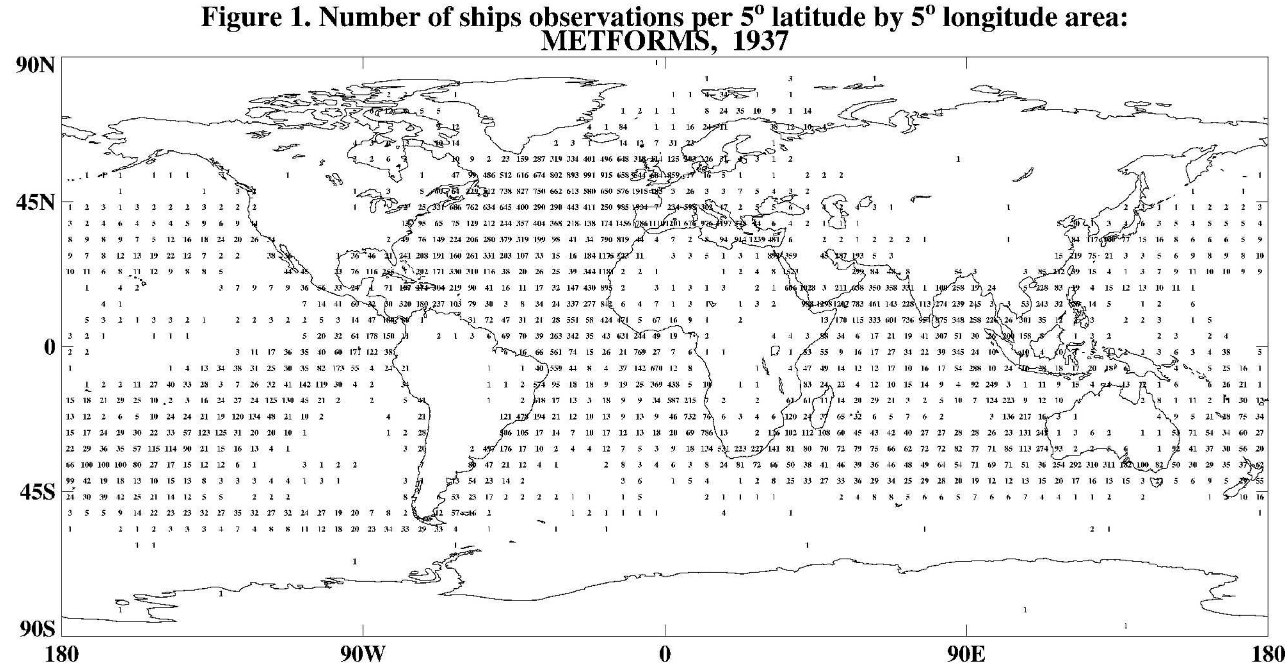


Figure 1: Number of ships' observations per 5° latitude by 5° longitude area: METFORMS, 1937.

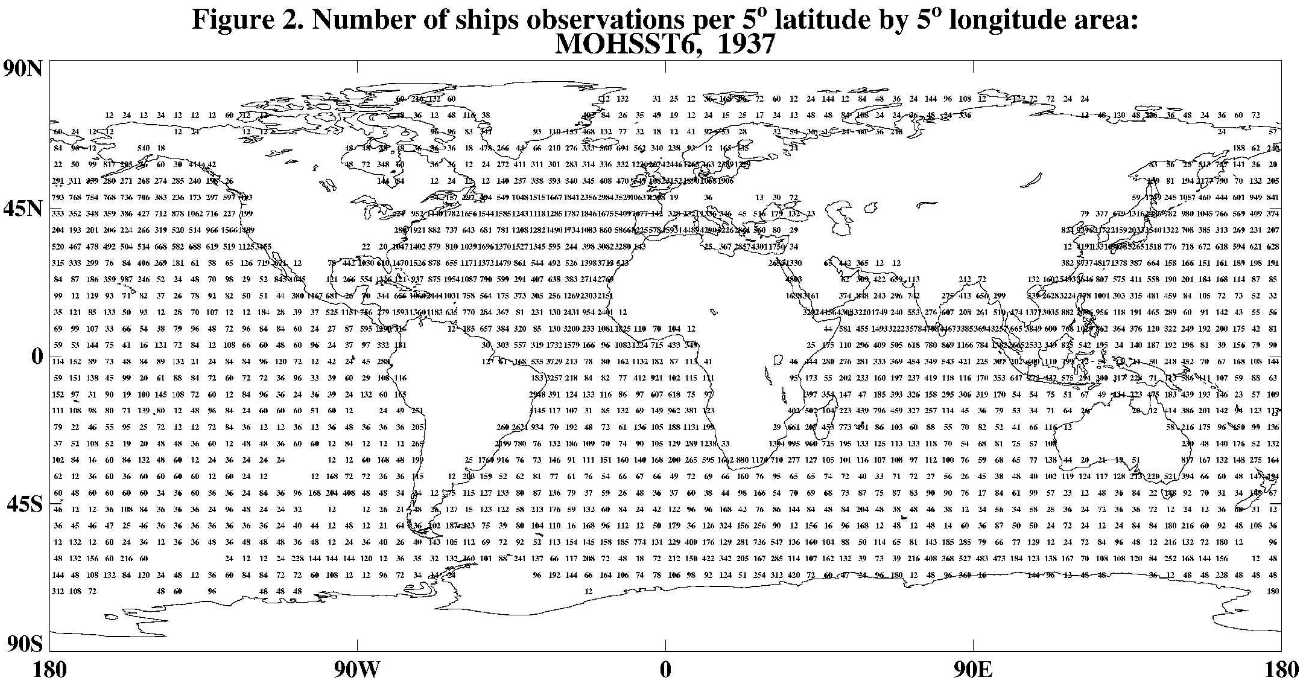


Figure 2: Number of ships' observations per 5° latitude by 5° longitude area: MOMMDB, 1937.

A full set of illustrations corresponding to Figures 1 to 3 is available from the authors.

In the UK Public Record Office there remain about 7 million undigitized observations for 1911-20 and 8 million for 1850-1900. If access is more costly than to the Meteorological Office archives, the price of digitization may exceed $2.5 million.

Nonetheless, these data should yield substantial benefits to climate research and therefore to the world's economy.

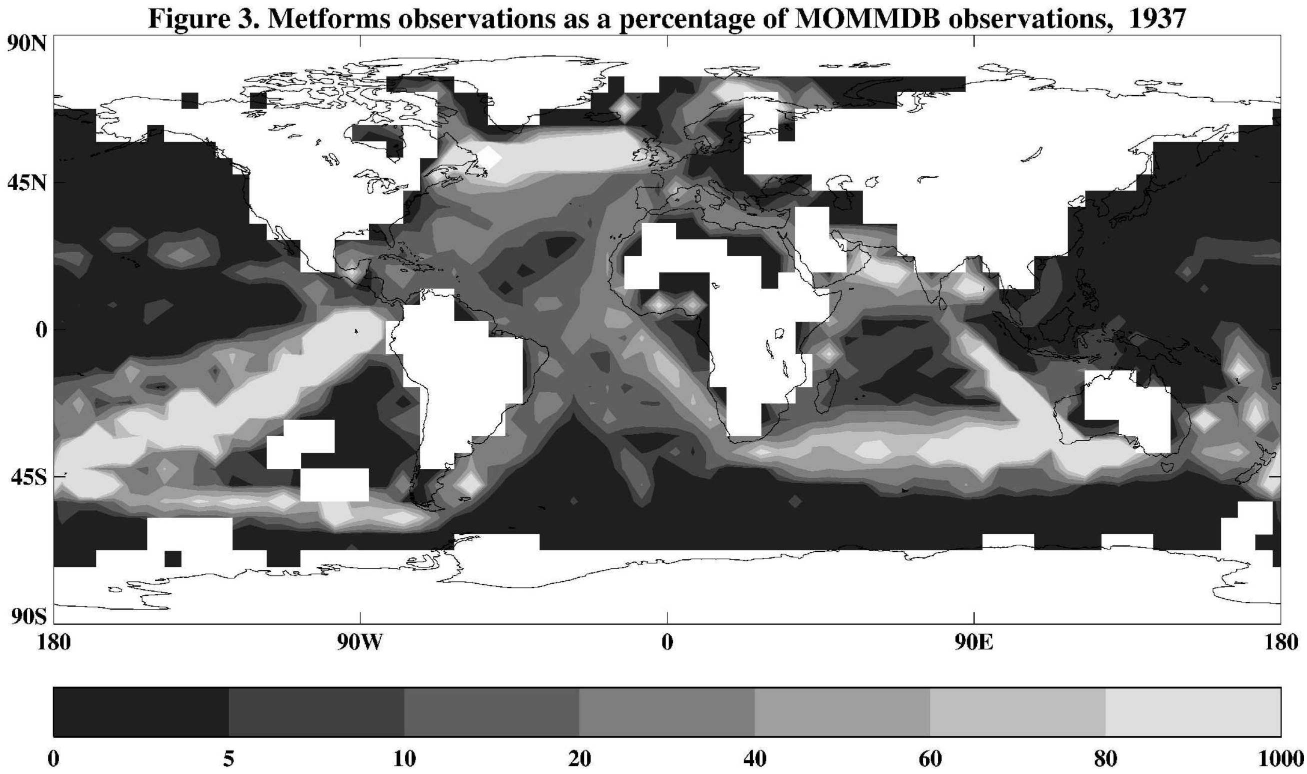


Figure 3: METFORMS observations as a percentage of MOMMDB observations, 1937.

---oooOooo---

Historical Marine Data in American and UK Archives, 1775-1900

Michael Chenoweth
Elkridge, Maryland, USA

1. Logbook characteristics

1.1 Locations of logbook collections

Logbook collections in American and UK archives hold large quantities of instrumental marine data from the late 18th century onward. In the United States, the most important collection of untapped data is that of U.S. Navy logbooks held by the National Archives in Washington, DC. In the United Kingdom, the largest collection is that of the Royal Navy, housed in the Public Record Office in Kew, London. Another important collection is held by the British Library's Oriental and India Office located in London. Other smaller collections include merchant and whaling ship logbooks held, in the greatest numbers, in three collections in New England (Mystic Seaport Museum Library, Mystic, Connecticut; Old Dartmouth Whaling Museum Library, New Bedford, Massachusetts; Peabody Essex Library, Salem, Massachusetts).

1.2 Logbook numbers

The number of logbooks totals in the tens of thousands for the years 1775-1900. The great majority of these logbooks are in the Public Record Office in London. The British Library holds about 1000 to 2000 logbooks in the 1775-1834 period. In the United States, the number of logbooks in regional libraries and archives is on the order of 7000 to 10,000. U.S. Navy logbooks in the National Archives number in the low thousands. Figure 1 gives decade totals of non-naval logbooks available from 1595 to 1900. This figure includes the East India Company collection along with the Old Dartmouth Whaling Museum's collection of more than 4000 logbooks. A few other small American collections are included. The inclusion of naval logbooks will increase these totals by probably an order of magnitude.

1.3 Instrumental data

The amount of instrumental data held in the tens of thousands of available logbooks can only be estimated from small samples of each data set. The British East India Company (hereafter EIC) logbooks held in the British Library probably include 250,000 to 300,000 observations from about 1780 to 1834. This is probably the largest collection of data for such an early time that exists anywhere. Tables 1 and 2 provide a detailed breakout of available instrumental data types in EIC and Royal Navy logbooks for the years 1815-1834.

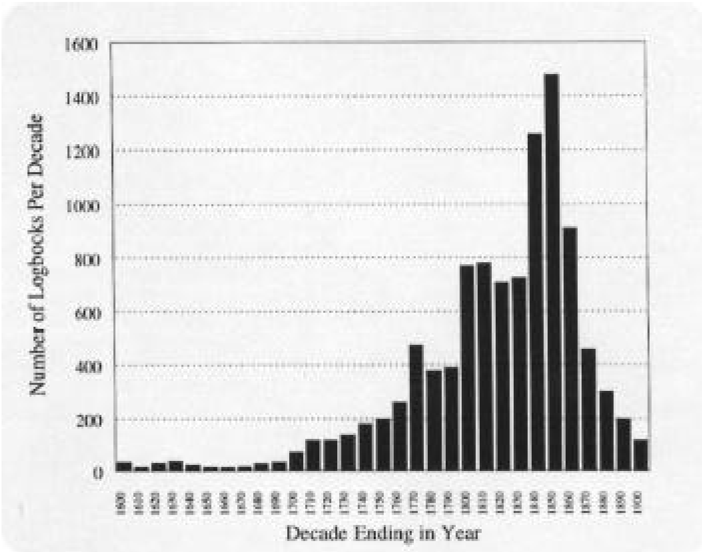


Figure 1: Non-naval logbooks.

Table 1
Number of British East India Company and Royal Navy logbooks with instrumental data, 1815-1816.

	Number of logbooks 1815-1816	Sample size 1815-1816	No Instrumental	Barometric pressure only	Temperature only	Both pressure and temperature
East India Company	76	55	26 47.3%	2 3.6%	6 10.9%	21 38.2%
Royal Navy	500 approx.[a]	120	116 96.7%	0 0.0%	3 2.5%	1 0.8%

[a] This number assumes that 2/3 of the approximate logbooks (750) are extant. Percentages are based on sample sizes of 55 and 120 respectively.

CCCCCCCCCCC

Table 2
Estimated total number of temperature observations available, 1815-1834.

	Estimated number of ships with temperature data based on all available log books	Average number of observations per ship per year	Estimated average total number of observations per year 1815-1834	Estimated total number of observations 1815-1834
East India Company	37	184	6808	136,160
Royal Navy	17[a]	425	7225	144,500
Royal Navy[b]	14	425	5950	119,000

a Based on 500 extant logbooks, 1815 – 1816.

b This row is based on a more realistic estimate of an average 625-ship Navy declining linearly from 750 in 1815 to 500 in 1834. Two-thirds of logbooks are assumed extant. More logbooks are available after 1834.

Table 3
U.S. Navy logbooks, 1804-1827.

Sample size	Instrumental data: with/without	% with data	Number of obs.	Estimated number of observations 1801-1900
30	17/13	57%	~10,000	~2,650,000

After 1834, the number of Royal Navy data is believed to number in the millions of observations, particularly after 1850.

Table 3 provides an estimate of the total number of observations in U.S. Navy logbooks for both the sampled period (1804-1827) and for the entire nineteenth century. Unlike the EIC logbooks, both sea surface temperature (SST) and surface air temperature (SAT) are included in U.S. Navy logbooks. No barometer data have been found in U.S. Navy logbooks for the years 1804-1827. The situation for later years is not known.

1.4 Logbook physical features

Most of the logbooks in the British Library and the U.S. National Archives are in large volumes in a variety of conditions (generally fair to excellent) and with only a limited amount of faded ink or otherwise illegible text. The Royal Navy logbooks are largely in flat formats but some are in large volumes. The larger logbooks (approximately 19-inch spines and 10.5 inches in width) are less amenable to microfilming than flat volumes.

Two important collections of logbooks are already filmed. One is collection ADM 55 in the Public Record Office, London. This collection holds 164 volumes of explorer's logbooks and meteorological diaries, many from Arctic explorations in the 1820s, 1840s and 1850s. The other is the International Marine Archive held by the Old Dartmouth Whaling Museum. This collection of more than 4000 logbooks is global in scale and extends back to 1595. Figure 1 gives a good idea of the temporal coverage of this collection.

1.5 Prioritizing of logbook digitization

Prioritization of logbook digitizing is dependent upon many factors. These factors include geographical coverage, available instrumental data, frequency of observations, coding formats, legibility, and estimated cost. For the early 19th century data, there is no single collection that has a clear advantage over other collections. For example, EIC logbooks are superior to other collections because of the use of pre-printed formats for data recording which provide space for latitude and longitude, barometer and thermometer data. The drawbacks of EIC logbooks are that, on average, there is only one daily observation of SAT and barometric pressure, and no SST data. Reproduction costs are high (on the order of 0.75 U.S. dollar per page for a hardcopy format). U.S. Navy logbooks have an advantage over EIC logbooks in that the typical logbook that includes instrumental data averages 6 to 8 observations per day and has both SST and SAT (but no barometer data). Two days of data per page is typical, which cuts reproduction costs. The greatest drawback for these logbooks is that most do not include the ships' longitude.

Additionally, some data are embedded within the account of daily activity on the ship, which makes locating the data difficult. However, both collections are concentrated in separate geographical areas. This makes both collections unique and valuable.

2. Interpretation of logbook contents

2.1 Observed features

Examination of hundreds of logbooks from the late l7th to mid 19th century has revealed evidence of national and institutional biases in logbook meteorological data. These biases need to be taken into account before these data can be incorporated into the Comprehensive Ocean- Atmosphere Data Set (COADS), or otherwise used by researchers for studying climate fluctuations of the past 400 years. A few of these features include:

- National navies tend to have a more fixed weather vocabulary (for both wind force and observed weather).
- "Present Weather" descriptors are best understood as a combination of low cloud cover amount and prevailing horizontal visibility.

- Assume persistence of weather conditions between observations, especially when the ship is at sea.

High-quality weather data in logbooks usually contain most, or all, of the following features:

- Observations continue while in port (usually seen only in navy logbooks).

- Local weather features (e.g., sea and land breeze) are accurately depicted (i.e., actual wind direction is given as opposed to recording only "land and sea breezes").

- Observations every two to four hours.

- Low frequency use of "ditto" observations.

- More or less fixed wind and weather descriptors

2.2 Interpretation challenges

Even with high-quality logbook data, there are a number of challenges to the interpretation of logbook entries. Probably the most important is the interpolation of ship position between observed positions. The four most important challenges are:

- Observed ship position, particularly longitude.

- Out of date or little known place names.

- Spatial and temporal smearing of observation time and location of a given weather report.

- Archaic or obscure weather descriptors.

Although ship latitudes are probably quite accurate under good weather conditions, longitude becomes increasingly problematic in earlier years. A reliable chronometer only came into general use in 1776. The coordinates of some well-known and frequently visited ports were not precisely known even in the early 19th century. Therefore, ship start point coordinates are likely to differ from modern coordinates. In other instances, longitudes (and sometimes latitudes) are not given when a landmark is observed. A dictionary of place names and their modern coordinates would be useful when digitizing old logbook data.

2.3 Interpretation of wind force descriptors

Perhaps the most common problem in interpreting logbook data is the quantification of wind force descriptors to an anemometer or Beaufort Scale value. Table 4 gives the modern Beaufort Scale (Beaufort number, standard name, and wind speed in knots). Table 5 gives the actual wind force descriptors used by the British Royal Navy in the mid-1810s. Figure 2 is a frequency distribution of Beaufort numbers versus Royal Navy equivalent numbers as given in Table 5 for ships sailing in the tropical North Atlantic from Maderia to the West Indies and from the South Atlantic to the UK in the mid-1810s. This figure indicates that the wind force terminologies are not equivalent. Table 6 gives the estimated equivalent wind speeds for the standard names used by the British Royal Navy in the early 1800s. This was done by forcing the numbers into the modern frequency distribution for the areas corresponding to the early 19th century ship routes in the tropical North Atlantic. Depending upon the country and the ship owner (Navy, EIC, merchant, whaler) there is a different set of most frequently used wind force descriptors.

Table 4
The modern Beaufort Scale.

Beaufort number	Standard name	Equivalent wind speed (knots)
0	Calm	<1
1	Light Air	1-3
2	Light Breeze	4-6
3	Gentle Breeze	7-10
4	Moderate Breeze	11-16
5	Fresh Breeze	17-21
6	Strong Breeze	22-27
7	Near Gale	28-33
8	Gale	34-40
9	Strong Gale	41-47
10	Storm	48-55
11	Violent Storm	56-63
12	Hurricane	>63

Table 5
Wind force descriptors used by British Royal Navy in the 1810s.

0	Calm
1	Light Airs Inclinable to Calm
2	Light Airs
3	Light Breeze (wind)
4	Moderate Breeze (wind)
5	Fresh Breeze
6	Strong Breeze
7	Fresh Gale
8	Strong Gale
9	Hard (heavy) Gale
10	Very Severe Gale[a]
11	Storm[a]
12	Hurricane

[a] Sample sizes are very small for these categories.

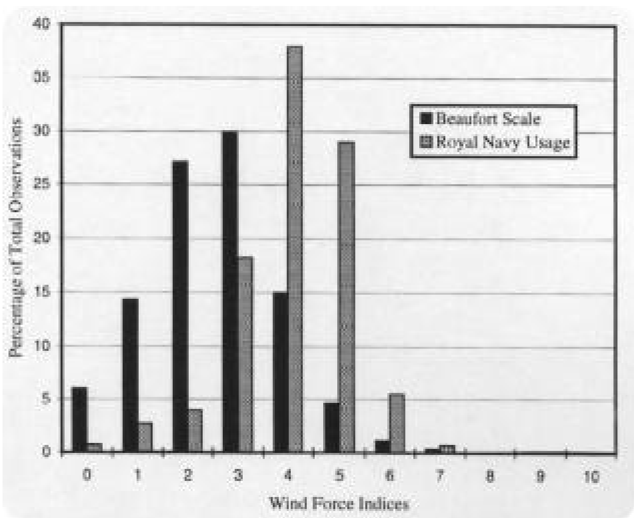


Figure 2: Frequency distribution of wind force indices, 1816 and 20th Century.

Table 6
Estimated equivalent wind speeds for standard names used by the British Royal Navy in the early 1800s.

Beaufort number	Standard name	Wind speeds (knots)
0	Calm	0
1	Light Airs Inclinable to Calm	0-2
2	Light Airs	2-3
3	Light Breeze (wind)	4-7
4	Moderate Breeze (wind)	8-13
5	Fresh Breeze	14-21
6	Strong Breeze	22-30
7	Fresh Gale	Over 30

Most force descriptors can be identified and made equivalent to the approximately equal terms used by another nation or institution. The only significant difference observed is for a bias (3 to 4 times above the others) of strong breezes in EIC logbooks.

Changes in technology through time will introduce false trends in estimated wind speed in time series data. Two ways of addressing this problem are:

1) The equipping of surviving sailing ships with anemometers and comparing these observations with the number of sail deployed and wind force descriptors used by the crew.

2) The drawing of synoptic weather charts from the late 18th and 19th centuries and comparing the observed pressure distribution (and inferred wind speed from the gradient) with the wind force descriptors used by ships within the area of accurate pressure reconstruction

---oooOooo---

The Archivo General de Indias (AGI): A Source of Surface Marine Climatological Information

R. García[1], M.R. Prieto[2], A. López[3], E. Hernández[1], L. Gimeno[4], and R. Herrera[2]
(1) Univ. Complutense de Madrid (Spain), (2) INAGLIA (Argentina),
(3) Archivo General de Indias (Spain), (4) Univ. de Vigo. (Spain)

Abstract

In this paper we evaluate the potential of the Archivo General de Indias (AGI) in Spain to act as a source of climatic marine data. The pertinence of using this historical source to obtain reliable climatic data is examined. The AGI is the main Spanish archive with information about the relationships between the Spanish Kingdom and its colonies in America and Philippines, and its importance for climatological applications is evident. The covered period is 1500-1825, when most of the Spanish Colonies became independent. We present three examples of the possibilities of the AGI as a source of marine data and metadata, related to different projects now being developed:

1) The Manila Galleon voyage 1780-1781.
2) The Coruña-Habana-Coruña mail route.
3) The Coruña-Montevideo-Coruña mail route.

Each of these examples is representative of different areas and climatic phenomena. A brief description of the techniques applied in the reconstruction is presented, and, for every example the extracted information is shown.

1. Introduction

Among the scientific community the most extended view is that the climate in the near future will be warmer than the climate of this century (Houghton et al., 1990, 1992). Most of the studies done to confirm this view are based on General Circulation Models (CGMs). The changes that these models predict do not correspond to the observed data. The differences between observed and modeled data could be due to deficiencies in the model, but also to an inadequate knowledge of the natural variability.

Instrumental records are available only for about a century. So, detecting an anomalous century according to the natural climate variability requires that we have climatic data for at least 1000 years. Obtaining these data involves several techniques like Dendroclimatology, Ice cores, Corals, Lake varves or Historical records. A description of these techniques is outside our work, but articles written by Bradley (1985), Bradley and Jones (1992a) and Bradley and Jones (1995) are good reviews of these techniques.

Among these techniques Historical reconstruction, when available, has the best temporal and spatial resolution. Much of the knowledge about the climate of the last millennium comes from European historical records (Wigley et al., 1981; Bradley and Jones, 1992b). These have permitted us to know many details about the terrestrial climate but very few about the oceanic climate. The need to have more historical climatic data for the oceans is evident if we consider that oceans represent two-thirds of the earth's total surface and consequently two-thirds of the surface covered by a GCM.

The urgency to extract information from historical sources, including the role of the Spanish Archives, was pointed out during the "Workshop on Digitization and Preparation of Historical Surface Marine Data and Metadata," 15-17 September 1997, Toledo, Spain.

Few scientists have used Spanish Archives to reproduce past oceanic climates and when used (Burt, 1990) the search was partial, considering only some types of documents. Among the Spanish Archives the "Archivo General de Indias" represents the most important source of information about oceanic climatic data in the world during three centuries (16th, 17th and 18th). The large number of journeys between Spain and its colonies in America and Asia is described in this Archive with such a high degree of detail as only Spanish bureaucracy could do. The aim of this article is to explore some of the possibilities that the AGI has to reproduce past Oceanic Climate.

2. The AGI

The "Archivo General de Indias" (AGI) is one of the most important archives in Spain. It is located in Sevilla and holds the documents of the Spanish Administration in the American colonies and the Philippines since they were discovered until they become independent (more than 300 years). There are more than 43,000 bundles in its files, covering a wide range of aspects of the life in Spanish America during colonial times. The AGI was founded at the end of the XVIII century by José de Galvez, one of the Government Secretaries which reformed the administration in Spanish America. There were two main reasons that lead to foundation of the AGI. One was organizational problems and a lack of space in the Archivo General de Simancas, which acted as the central archive of the crown since the XVI century. The second reason was of a historical and political nature. King Charles III ordered the writing of a new history of the Spanish colonization, which could balance the books published in England and the Netherlands that originated the "Black legend" of the Spanish colonization.

Thus, Juan Bautista Muñoz was commissioned to classify the documents relative to the American colonization. They were located mainly in the House of Trade (Sevilla and

Madrid), the Consulates (Sevilla and Cádiz), the Indies Council (Simancas and Madrid), and the Secretary of State (Simancas and Madrid) (González, 1996).

In 1785 the documents were moved to their present location. The 43,000 bundles occupy eight linear kilometers of shelves, with around 80,000,000 pages. From these, 12% is now digitized, covering around 60% of the most frequent requests for information.

2.1 Fleets

We are primarily concerned with documents related to the navigation between South and Central America and Spain. During the XVI to XVIII centuries the transoceanic traffic was organized through a convoy system which was related to a monopolistic commerce model (Haring, 1918; Laviana, 1996): the fleets and the galleons. Since 1543 all the ships had to sail compulsorily in a fleet. Two main fleets were organized yearly:

a) "The Flota of Nueva España". It sailed from Sevilla during the period April-August with destination Veracruz via Santo Domingo.

b) The "Galeones de Tierra Firme". It sailed from Sevilla between August and November with destination Portobello via Cartagena de Indias.

Both fleets were connected to a very wide commercial network. In Portobello the merchandise was transported by land to Panama, and from there shipped to Callao (Perú) and Arica (Chile). From there, they reached Oruro and Potosí by land, where the return travel started back to Spain.

Figure 1 shows the main routes, which are of obvious climatological interest. Thus, the journey from Spain to the Caribbean Sea was determined by the intensity of the trade winds and the location of the pressure centers across the Atlantic. During the passage through the Caribbean, they were affected by tropical cyclones.

The duration and location of the return travel represent proxy data of the intensity of the westerlies and the North Atlantic Oscillation. The commerce along the Pacific Coast was very closely connected with the ENSO cycle. The southern navigation was very difficult most of the time.

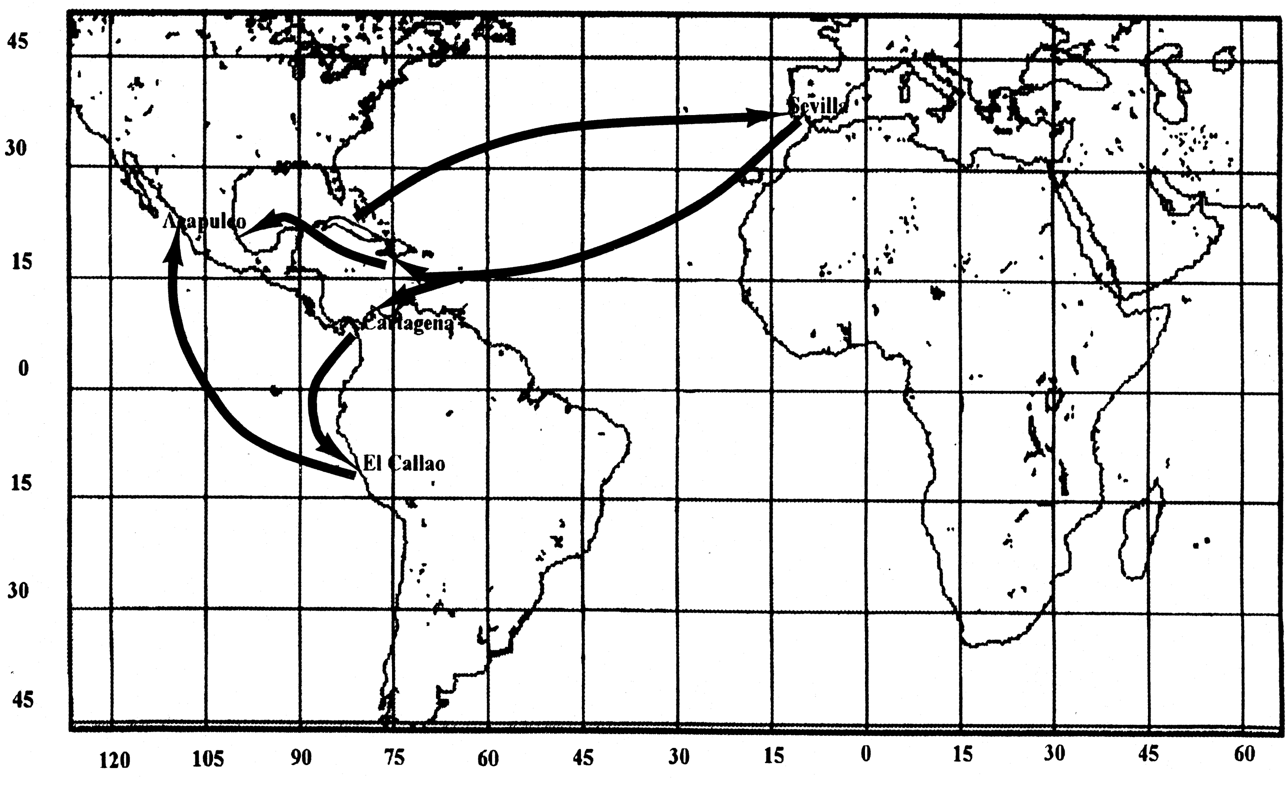


Figure 1: Main routes of fleets to and from the Caribbean and South America from Sevilla.

Since the second half of the XVI century, the annual voyage of the Manila Galleon was established (Shutz, 1939). It was the only link between the Philippines and Mexico and Spain. The Philippines acted as an intermediate market for the merchandise from China.

Chinese merchants transported merchandise to Manila, and from there the annual voyage of the Galleon was organized. It sailed from Manila to Acapulco, the sailing time being determined by the monsoons (they had to wait until the merchandise arrived from China) and the typhoons (they had to sail before the typhoon season started). From Manila they sailed north to reach the Kuroshio Current until they arrived on the California coast (Burt, 1990). The return travel was mainly driven by the intensity of the Pacific trades. The time length of the journey can be considered a good proxy datum of the ENSO cycle.

The Manila Galleon lasted until 1815, and the fleet system finished at the beginning of the XVIII century. Political and economic changes in Spain lead to a change in commerce, which finally became completely free in 1778 (Laviana, 1996). Then a new route was open: the Cape Horn route connecting Argentina and Chile.

2.2 Mail ships

The first postal system between Spain and the Colonies was the Correo Mayor (Main Post) during the XVI and XVII centuries. The mail ships are the main source of systematic information other than the fleets. Since 1628, four mail ships sailed every year from Sevilla to Mexico and Cartagena. During the whole XVII century a ship sailed every three months from Spain to America. However, from 1764 until 1825 the mail ships sailed from La Coruña to Havana, Montevideo-Buenos Aires, Falmouth and New York. From these cities the correspondence was distributed to the rest of the colonies (López Gutiérrez, 1996). From Havana, different ships distributed the correspondence to Veracruz, Florida, Louisiana, Campeche and Yucatan, Puerto Rico and Venezuela (connecting Lima through Cartagena and Portobello) (González, 1996).

A ship sailed every month to Havana, the ships to Buenos Aires sailed every four months, and in 1771 they sailed every two months. At the end of the XVIII century, the Anglo-Spanish war originated the blocking of the ports and the end of this system. The captains had the duty to deliver the logbooks to the Chief Post Officer in La Coruña. In the logbooks (examples shown below) the following information was included:

- Daily position at midday.
- Daily distance covered.
- Meteorological events of the day, with a qualitative scale for the wind.

The Havana route provides information related to the same phenomena as the fleets. The Buenos Aires route can provide a good proxy description of the 40ºN-40ºS region of the Atlantic.

So, we have a highly valuable source of marine climatic data and metadata from the two main areas: the Atlantic and the Pacific-South American coast for a period of around 300 years. However, political, economic and military reasons make the time-space coverage rather inhomogeneous. For instance: wars between Spain and Britain collapsed the fleet route for several years during the second half of the XVII century. The economic strength of the Portobello-Panama-Potosí connection did not allow the use of the Cape Horn route until the mid XVIII century. All the circumstances explain why every geographical area has its own time coverage. The information has to be extracted not only from the logbooks, which sometimes are sparse, but also from:

- Captains' reports.
- Judging procedures made by the House of the Trade.
- Administration reports from the different reports in America.

3. Techniques applied for the climatic reconstruction

One of the main difficulties of handling this kind of information is the lack of precise standards to characterize the climatic phenomena. So, the same phenomenon, with multiple variations, can be qualified by a large number of different words and expressions. In order to validate and homogenize the data, a content analysis technique has been applied. This technique takes into account the linguistics of the considered period and the way in which the climatic phenomena were perceived and expressed.

Table 1

Database of variables considered and their encoding.

Position Source Day/month/year	Degrees site	
Rain	Qualitative measure Quantitative measure	intermittent(1) moderate(2) abundant(3)
Storms	Presence Qualitative measure	light(1) strong(2) very strong(3)
Snow	Qualitative measure Quantitative measure	intermittent(1) moderate(2) abundant(3)
Cloud cover		clear(1) partly cloudy(2) cloudy(3) stormy without precipitation(4) foggy(5)
Temperature	Qualitative measure degrees	very warm(1) warm(2) mild(3)
Sea state		calm(1) short(2) heavy(3)
Currents Icebergs	Presence, direction Presence	
Wind	Qualitative measure type, direction	calm(1) slow(2) regular(3) strong(4)
Observations		

Finally, the correspondence with the present climatic lexicon is considered. This technique has an obvious limitation: it must be applied over a period short enough to guarantee a certain homogeneity in the language, the measurement instruments and the navigation technology. To ensure this, we have considered the second half of the XVIII century in order to illustrate the use of content analysis. The following steps were taken:

a) A technical vocabulary was established with the climatic categories included in Table 1.

b) The whole range of climatic information contained in the logbooks was analyzed. All the expressions and phrases related to weather were extracted.

c) A Dictionary of climatic terms for the considered period was elaborated. The Spanish Marine Dictionary (1831) was the main tool.

d) Each of the climatic expressions was applied to one of the present climatic categories. Tables 2a and 2b show the equivalence between some normalized climatic categories and their equivalent in the language of the XVIII century for two variables: rain and wind strength. Thus we could build up a database. The events were classified according to magnitude and intensity. A numerical range was attributed to every phenomenon, which was coded for entry into the data set. For the wind strength we could not use the Beaufort scale because this variable was described by a different number of criteria whose equivalency to the Beaufort scale is not always evident.

4. Three examples: results and comments

In this section we provide three examples of the information which can be obtained from the documents kept in the AGI. We have used logbooks from the following journeys:

- The Manila Galleon (1780-1781) (AGI, Estado 38B)
- The Carrera de la Havana (1775-) (AGI, Estado 276B)
- The Carrera de Montevideo (1778) (AGI, Estado 191B)

4.1 The Manila Galleon (1780-1781)

Figure 2 shows the route followed by the Manila Galleon on its voyage during 1780-1781. This journey was strange since its departure date was later than usual. As a consequence, the captain could not find the route to reach the Kuroshio current and the Galleon had to sail through the Southern Hemisphere.

Table 2a

Normalized climatic categories and equivalent XVIII Century terms, for rain descriptors.

Intermittent	Moderate	Abundant
• Lloviendo alguna veces • Algunos chubasquillos	• Chubascos • Aguaceros • Chubasqueando • Chubasquillos • Los cielos manifestaba tempestad • Tormenteando con lluvia • Chubascoso • Chubasqueando continuamente • Algunos chubasquillos • Levantaron las lluvias • Muy ofuscado y lloviznosa • Cielo nebuloso con chubasquillos • Tiempo húmedo • Chubascos de poca monta	• Continuos aguaceros • Llovió la mayor parte del día • Muy tormentoso • Gruesas lluvias • Gruesas y ventolinas lluvias • Cielo aturbonado y lluvioso • Pesadas lluvias • Continuos chubascos • Muy tormentoso de aguas • Fuertes chubascos • Tormenta de agua, mar y viento • Lluvias muy fuertes • Turbonada • Fuertes lluvias • Inconstancia del tiempo y continuos temporales • Frecuentes aguaceros • Turbón de aguacero tempestuoso • Buen aguacero • Muchos aguaceros • Aguas muy pesadas • Densas aguas

Table 2b

Normalized climatic categories and equivalent XVIII Century terms, for wind strength descriptors.

Calm	Soft	Regular	Strong
• calma • calma muerta	• vientecito • viento calmoso • viento flojito • bonancible • viento muy bonancible • viento muy calmoso y variable • viento flojo • viento apacible • viento muy flojito • ventolina • viento benigno	• viento con malas semblanzas • brisa fresca • viento fresco • viento fresquito • viento recio	• viento fogoso • viento con malso semblantes • viento turbulento • feurte viento • viento calmoso • tormentoso • viento muy fogoso • viento muy tormentoso • viento muy variable y grueso • viento imperioso • viento furioso • recio huracán • viento con mucha furia • viento con tanta violencia • viento que parecía huracán • frescachón • viento soplaba con desenfreno • recias jugadas de viento

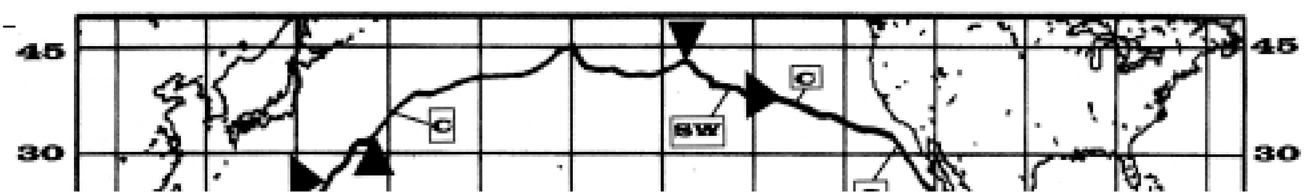


Figure 2: Route of the Manila Galleon voyage, 1780-1781, with abbreviations displayed for wind strength and rain (see Tables 2a and 2b), and arrowheads for wind direction.

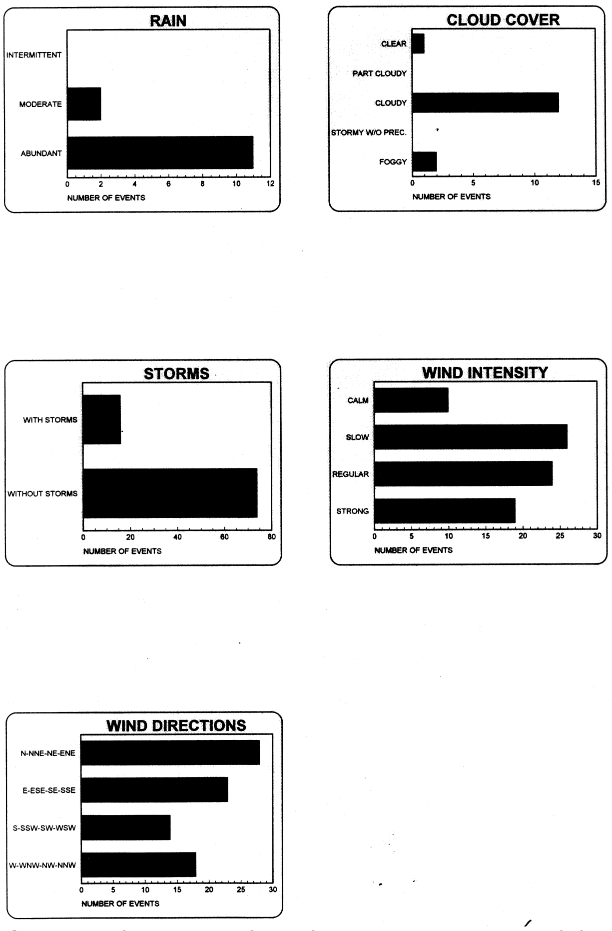


Figure 3: Histograms of weather events encountered during Manila Galleon voyage, 1780-1781 (Figure 2).

On the figure the most representative winds and rain are displayed. During the first half of the journey the galleon suffered from abundant rain and light winds. Therefore, it had to cross the Equator where calm days are abundant and evening convective storms usual. Once the Galleon reached mid-latitudes in the Northern Hemisphere its speed was much higher although calms persisted. The reason was the effect of the Kuroshio current. At the end of the journey and close to the Mexican coasts the captain noted on his logbook the existence of a terrible hurricane with extraordinarily strong winds that could be related to the presence of a tropical storm.

More detailed information about the weather found by the 1780-1781 Manila Galleon is summarized in Figure 3. Weather events are described according to the categories of classification described in section 3. Five meteorological phenomena were considered: rain, cloud cover, storms, wind intensity and wind directions. It is important to note that data from the figure correspond to the days with captain's comments. So, the sum of days with and without storms is not the total number of days of the journey. It is remarkable that the number of days with abundant rain is higher than the days with moderate rain. That number is very similar to the number of days with storms.

The distribution of wind categories shows that slow and regular winds were the most abundant although the number of calms was significant (10 days). These days correspond to the journey through the equatorial regions. By examining the categories of wind directions, we can appreciate that easterly winds are the most frequent. That is why the galleon had to sail through low latitudes during two-thirds of the total journey where trades are the dominant winds.

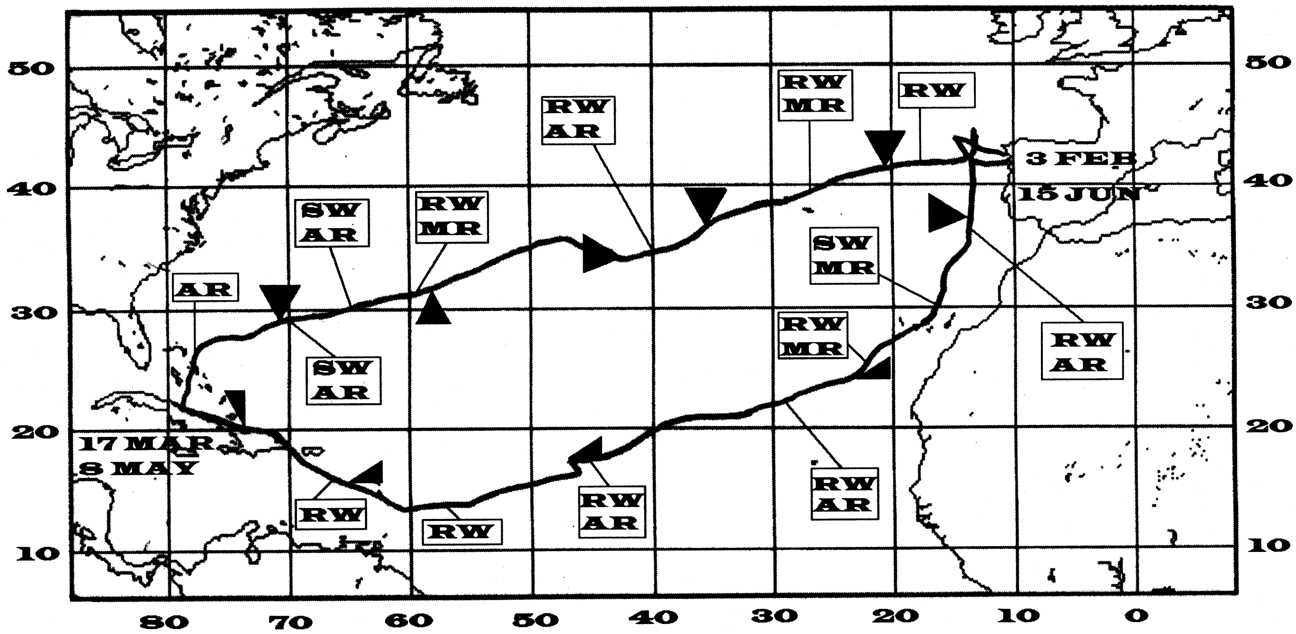


Figure 4: Route of a Carrera de la Havana voyage, 1775, with abbreviations displayed for wind strength and rain (see Tables 2a and 2b), and arrowheads for wind direction.

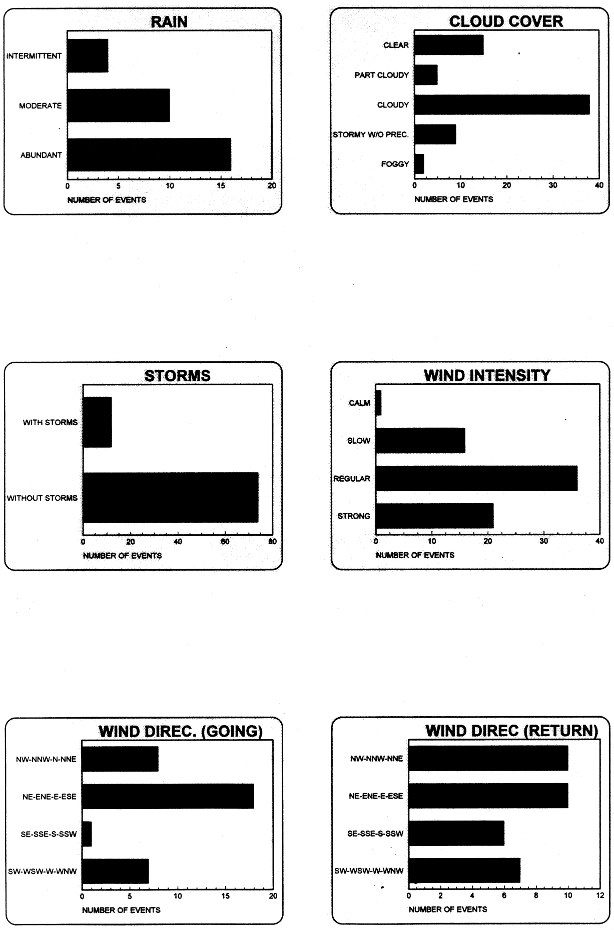


Figure 5: Histograms of weather events encountered during a Carrera de la Havana voyage, 1775 (Figure 4).

4.2 The Carrera de la Havana (1775)

The Carrera de la Havana was a mail ship that sailed from La Coruña to La Havana. The trajectory displayed in Figure 4 corresponds to one of the journeys done in 1775. The outgoing and return voyages lasted a month and a half each. The voyage rounding the anticyclone over the North Atlantic Ocean used the trades to go to La Havana and the westerlies to reach La Coruña. In this example, the outgoing voyage was more typical than the return, since during the outgoing voyage the trades were very constant but during the return there were many times with NE-NW winds. The strongest winds as well as the most abundant rains happened in the return journey close to American coasts. The description of the events happening in this area showed that the captain was afraid of the intense winds. Another remarkable result is the presence of rain close to the Sahara area, which is not very frequent.

Figure 5 summarizes weather events by means of categories. Abundant rains were more frequent than moderate or intermittent rain, the number of days with abundant rain being higher than the number of storms. This result suggests that there were days in which frontal precipitation in mid-latitudes was significant. Only about 20% of the days were clear, with cloudy days the most frequent. Calms were not very abundant, although the ship had to sail for a time along the Equator. Regular winds were the most frequent and strong winds happened more frequently than slow winds. A review of the wind direction categories shows clear differences between the outgoing and return voyages. During the outgoing voyage, NE-ENE-E-ESE categories, due mainly to trades, are the dominant ones. This permitted the ship to travel first to the SW and then to the W. During the return voyage the distribution of wind direction categories is more homogeneous. Winds with a northern component continued to be predominant, as opposed to winds with a southern component, although the sum of these categories was higher than in the outgoing voyage. The number of days with westerlies also rose.

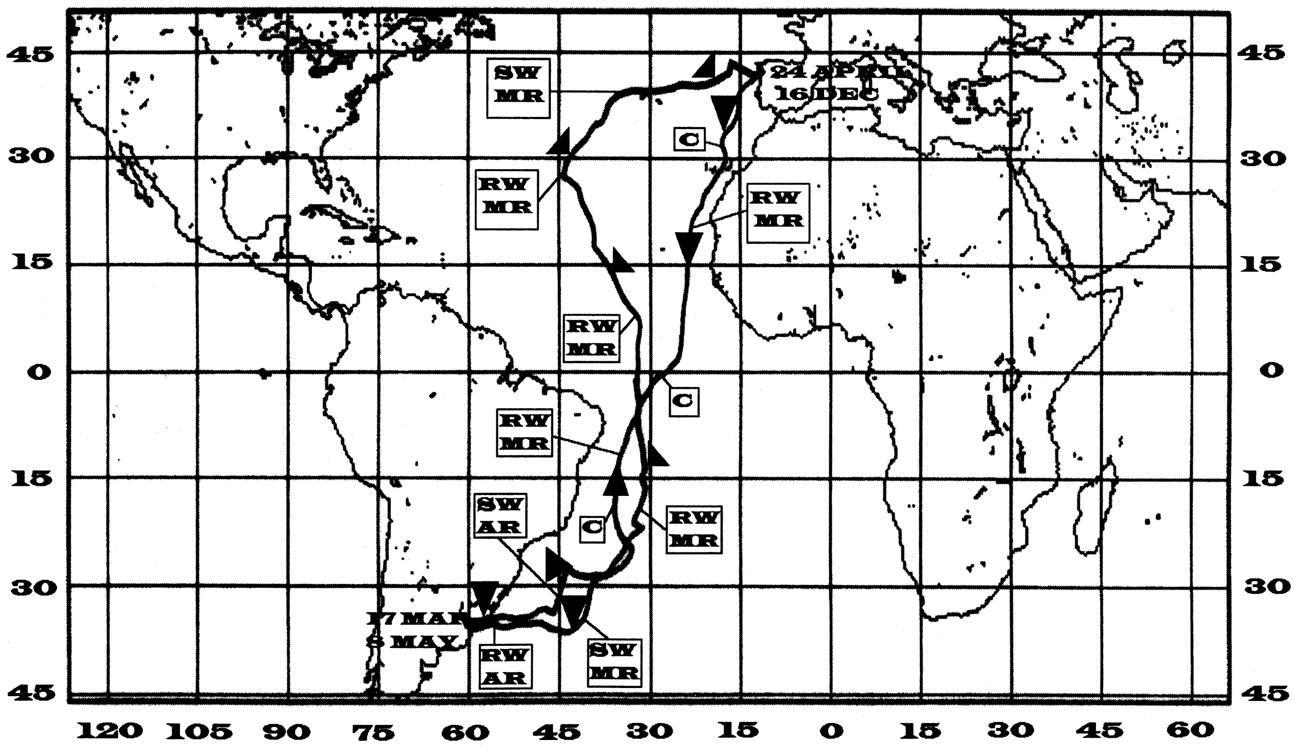


Figure 6: Route of a Carrera de Montevideo voyage, 1778, with abbreviations displayed for wind strength and rain (see Tables 2a and 2b), and arrowheads for wind direction.

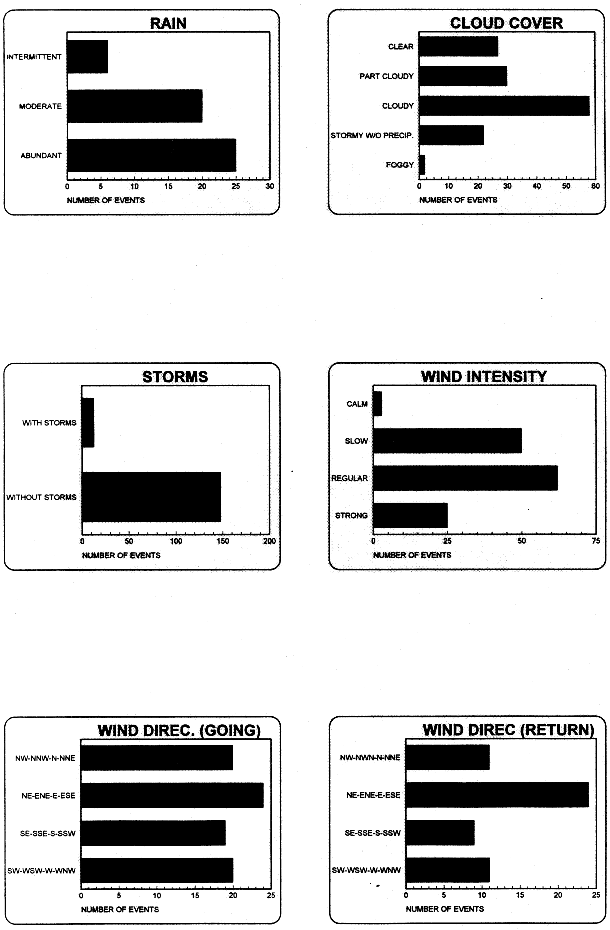


Figure 7: Histograms of weather events encountered during a Carrera de Montevideo voyage, 1778 (Figure 6).

4.3 The Carrera de Montevideo (1778)

The mail ship called The Carrera de Montevideo sailed from La Coruña to Montevideo and then to Buenos Aires, so it had to cross the Equator. The weather events suffered by this ship were very different. Trades, equatorial calms, easterlies in the Southern Hemisphere or westerlies in the Northern, frontal rain in mid-latitudes and convective rain in the tropics could be good examples of the variety of events that could happen during this journey. In our example (Figure 6), many of these events are present, like calms or westerlies, whereas other are not present like easterlies in the Southern Hemisphere or strong rain in the tropics. The most dangerous moment of the journey happened at the beginning of the return voyage from Montevideo to la Coruña because of strong winds and abundant rain. It is also interesting to note the asymmetry between the outgoing and return voyages. This asymmetry is more evident in mid-latitudes because of the different seasons and the route of ships that had to round anticyclones. However, the trajectory had necessarily a crossing point close to the Equator. In the outgoing voyage there was calm, while in the return voyage there were regular winds and moderate rain.

In Figure 7 weather events classified by categories are displayed. The most remarkable results are the differences between the number of days with rain and the number of days with storms. This result is due to the fact that the ship had to sail more than one half of the journey through mid-latitudes both in the Northern and in the Southern Hemisphere. Cloudy days were the most frequent and the number of clear days represents about 15%. Regular and slow winds are the most frequent wind intensity categories. There were 25 days with strong wind and only three days with calms. The wind direction categories distribution during the outgoing voyage shows similar values for the four categories. This is a logical result taking in consideration that the ship had to sail from mid-latitudes in the Northern Hemisphere to mid-latitudes in the Southern crossing low-latitudes. During the journey the ship found many different wind directions. During the return voyage, NE-ENE-E-ESE is the dominant category. The reason for this asymmetry is the existence of strong winds in mid-latitudes both in the Southern and in the Northern Hemisphere. So the number of days crossing regions with trades is higher than the number of days crossing westerlies.

5. Concluding remarks

The previous examples show that the records kept in the AGI, if adequately searched and treated according to techniques such as content analysis, can be a valuable source of marine climatic information.

Some periods and areas have a special interest due to the regular spatial and temporal coverage. The information provided by the mail ships crossing the Atlantic during the second half of the XVIII century is a good example of that. Phenomena such as variability in the trades and westerlies can be adequately characterized for this period.

References

Bradley, R.S., 1985: *Quarternary Paleoclimatology: methods of paleoclimatic reconstruction.* Allen and Unwin, Boston, 472 pp.

Bradley, R.S., and Jones, P.D., 1992a: Introduction in *Climate since A.D. 1500*. R.S. Bradley and P.D. Jones (Eds.), Roukledge, London, 1-16.

Bradley, R.S., and Jones, P.D., 1992b: *Climate since A.D. 1500*. Roukledge, London, 679 pp.

Bradley, R.S., and Jones P.D., 1995: Recent developments in studies of climate since A.D. 1500. In: *Climate since A.D. 1500* (2nd edition). R.S. Bradley and P.D. Jones (Eds.), Roukledge, London 666-679.

Burt, 1990:. The search for the Manila Galleon Log Books. *Bull. Amer. Meteor. Soc*., **71** (11), 1630-1636.

González, P., 1996: El Archivo General de Indias y el correo en América in El correo Español en América. Archivo General de Indias. España

Haring, C.H., 1918: *Trade and Navigation between Spain and the Indies in the time of the Hapsburgs*. Harvard Univ. Press. Cambridge. 459 pp.

Houghton, J.T., Jenkins, G.J., Ephraums, J.J. (Eds.) 1990: *Climate change: The IPCC Scientific Assessment*. Cambridge University Press, Cambridge.

Houghton, J.T., Callander, B.A., Varney, S.K. (Eds.) 1992: *Climate Change; The Supplementary Report to the IPCC Scientific Assessment*. Cambridge University Press, Cambridge.

Laviana, M.L., 1996: La América Española, 1492-1898. Temas de Hoy. Madrid. 146 pp.

López Gutiérrez, A., 1996: Inventario de la sección de correos. Ministerio de Educación y Cultura, Madrid, 794 pp.

Schutz, W.L., 1939: *The Manila Galleon*. E.P. Dutton. New York, 453 pp.

Wigley T.M.L., Ingram M.J., Farmer G., (Eds.) 1981: *Climate and History*. Cambridge University Press, Cambridge.

---oooOooo---

South Atlantic Winds and Weather During and Following the Little Ice Age - A Pilot Study of English East India Company (EEIC) Ship Logs*

Anthony J. Farrington[1], Sandra Lubker[3], Uwe Radok[2], and Scott Woodruff[3]
(1)British Library, London (2)CIRES, University of Colorado, USA
(3)Climate Diagnostics Center, NOAA, USA

Abstract

The times taken by ships of the English East India Company (EEIC) to sail from the Cape of Good Hope to St. Helena Island during the 17th, 18th, and early 19th centuries represent proxy measures of the strength and steadiness of the Southeast Trades which are compared with present-day data from the Comprehensive Ocean-Atmosphere Data Set (COADS). Both wind speed and steadiness appear to have reached maxima in the 1760s and increased again, from lower values in the following decade, to the 1820s.

These changes need to be further substantiated with the available log entries concerning winds, weather, and rates of progress. A similar fleshing out of fragmentary pre-instrumental pre-standardized records with current climatic characteristics is suggested for the routes fanning out east of the Cape towards Arabia, India, China, Indonesia, and Australia.

1. Introduction

The British Library's India Office Collection includes several thousand logbooks and other records of the English East India Company from journeys between England and the East during the 17th, 18th, and early 19th centuries. The sailors of those days recorded their daily positions and, in many cases, their hourly rate of progress as knots (K) and fathoms (F) (Figure 1), measured along a line paid out from the stern. By contrast, they made and recorded only sporadic wind and weather observations. Alexander Dalrymple, Hydrographer of the EEIC (1779) and later of the English Admiralty, unsuccessfully tried to derive winds and currents encountered from the logs of individual voyages.**

The wealth of climatic data available 200 years later, in the form of marine climatologies (e.g. Van Loon, ed., 1984) and numerical archives such as the Comprehensive Ocean-Atmosphere Data Set (COADS, e.g., Woodruff et al. 1993), has made possible the different approach used in the following: instead of single voyages we examine a large number of passages along a single route, selected for its simple and persistent wind and current regimes.

Passage duration statistics then should reflect the broad characteristics and any gradual changes of the wind regime, as well as occasional anomalous conditions. The latter could be created by storms or by non-climatic factors, such as equipment failures or the need to circumvent regions of war and piracy. While individual logs must be searched for evidence of such factors, an initial analysis can use merely the dates of landfalls, recorded by the first author in a comprehensive catalogue of the EEIC logs (Farrington, unpublished).

2. The region and its characteristics

The passages chosen for analysis went from the Cape of Good Hope (34°25'S, 18°29'E) to St. Helena Island (16°S, 5°43'E). Figure 2 shows the average COADS winds for January and July; driven by the South Atlantic subtropical anticyclone, they in their turn drive the Benguela Current.

The present-day average seasonal speeds, in a 44-year sample of 33,000 ship reports from the 2° x 2° degree square halfway along the route in Figure 2, have a spring maximum of 8.2 m/sec and a winter minimum of 7.5 m/sec

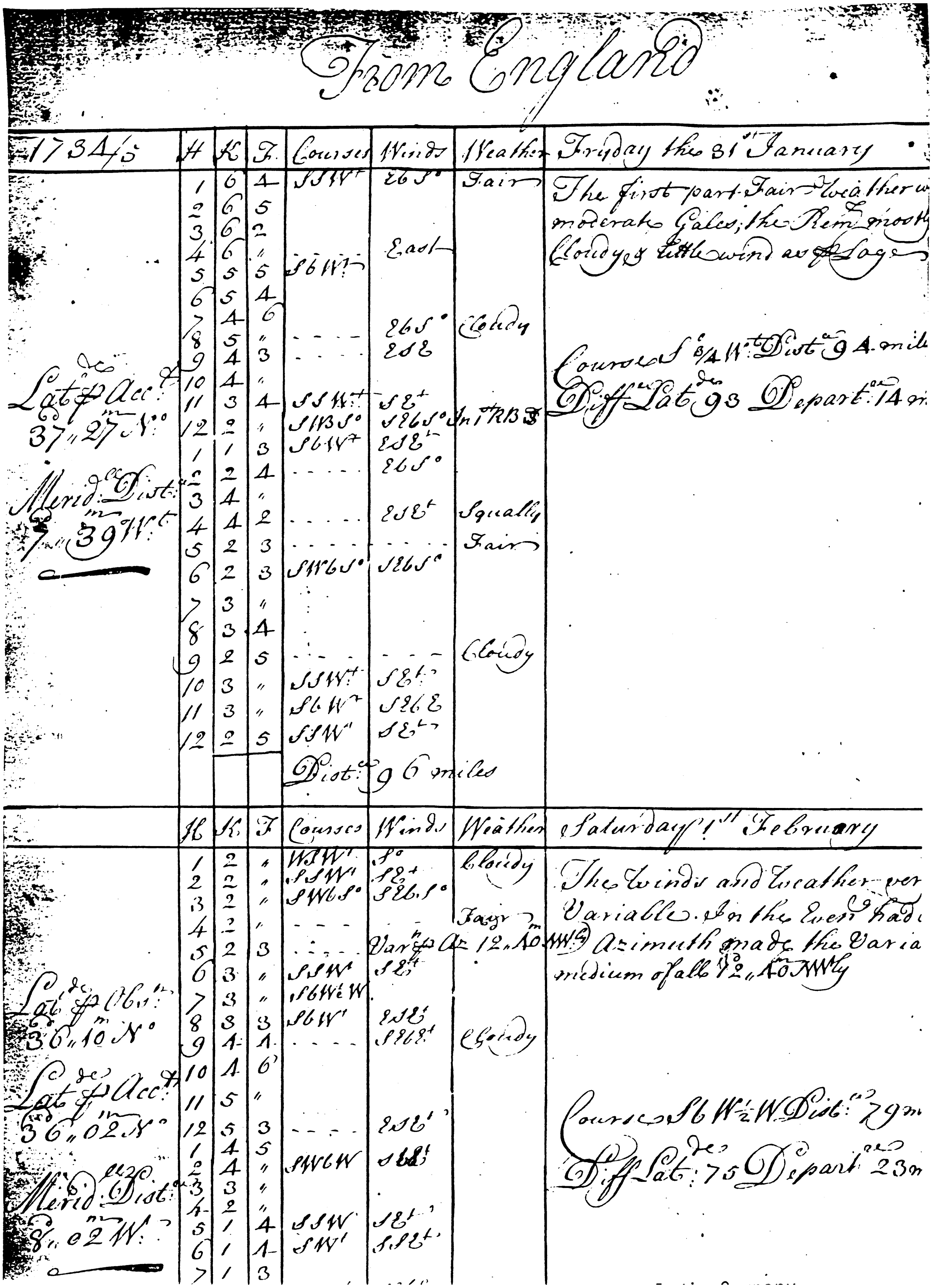
An outstanding characteristic of these Southeast Trades is their high directional steadiness or "constancy" (defined as the ratio of the mean vector and mean scalar magnitudes). Figure 3 shows that it exceeded 0.9 in more than 90% of the 44 summers in the COADS sample; that frequency remained near 80% for the transitional seasons but decreased below 40% for the winters. These seasonal differences may be explained by the varying incidence of upper lows which occasionally invade this area from the southern Westerlies (Lamb, 1957).

3. Passage durations

The EEIC logs include 760 in which landfalls first at the Cape and then at St. Helena Island were recorded, providing a 240-year record of durations of down-wind passages through the South Atlantic. The only set of logs with such dates for upwind passages is especially useful because it covers several consecutive trips in both directions by the same ship. Table 1 shows that the downwind passages took only half the time required by those against the Trades, with one exception.

* This paper is also being published in the H. Riehl Memorial Issue of *Meteorology and Atmospheric Physics.*

** We are indebted to Dr. Andrew Cook for his interpretation of the log entries and for the reference to Dalrymple's work.



From England

1734/5	H	K	F	Courses	Winds	Weather	Fryday the 31st January
	1	6	4	S S W	E b S	Fair	The first part Fair Weather w moderate Gales; the Rem. mostly Cloudy & little wind as ⅌ Log
	2	6	5				
	3	6	2				
	4	6	"		East		
	5	5	5	S b W			
	6	5	4				
	7	4	6				
	8	5	"	- - - -	E b S	Cloudy	
	9	4	3	- - - -	E S E		Course S 3/4 W. Dist. 94 miles
Lat. ⅌ Acc.t	10	4	"				Diff. Lat. 93 Depart. 14 m.
37°27' N.o	11	3	4	S S W	S E		
	12	2	"	S W b S	S E b S	In 1 Reef	
	1	1	3	S b W	E S E		
Merid. Dist.	2	2	4	- - - -	E b S		
7°39' W.t	3	4	"				
	4	4	2	- - - -	E S E	Squally	
	5	2	3	- - - -	- - - -	Fair	
	6	2	3	S W b S	S E b S		
	7	3	"				
	8	3	4				
	9	2	5	- - - -	- - - -	Cloudy	
	10	3	"	S S W	S E		
	11	3	"	S b W	S E b E		
	12	2	5	S S W	S E		
				Dist. 96 miles			

	H	K	F	Courses	Winds	Weather	Saturday 1st February
	1	2	"	W S W	S	Cloudy	The winds and weather very Variable. In the Even. had Az. made the Variation medium of all 12°40' W.ly
	2	2	"	S S W	S E		
	3	2	"	S W b S	S E b S		
	4	2	"	- - -		Fair	
	5	2	3	- - - -	Var. ⅌ Az. 12°40' W.		
	6	3	"	S S W	S E		
Lat. ⅌ Obs.	7	3	"	S b W ½ W			
36°10' N.o	8	3	3	S b W	E S E	Cloudy	
	9	4	4	- - - -	S E b E		
Lat. ⅌ Acc.t	10	4	6				
36°02' N.o	11	5	"				Course S b W ½ W Dist. 79 m
	12	5	3	- - - -	E S E		Diff. Lat. 75 Depart. 23 m
	1	4	5	S W b W	S b E		
Merid. Dist.	2	4	"				
8°02' W.	3	3	"				
	4	2	"				
	5	1	4	S S W	S E		
	6	1	4	S W	S S E		
	7	1	3				

Figure 1: Sample page from a log of a ship of the English East India Company.

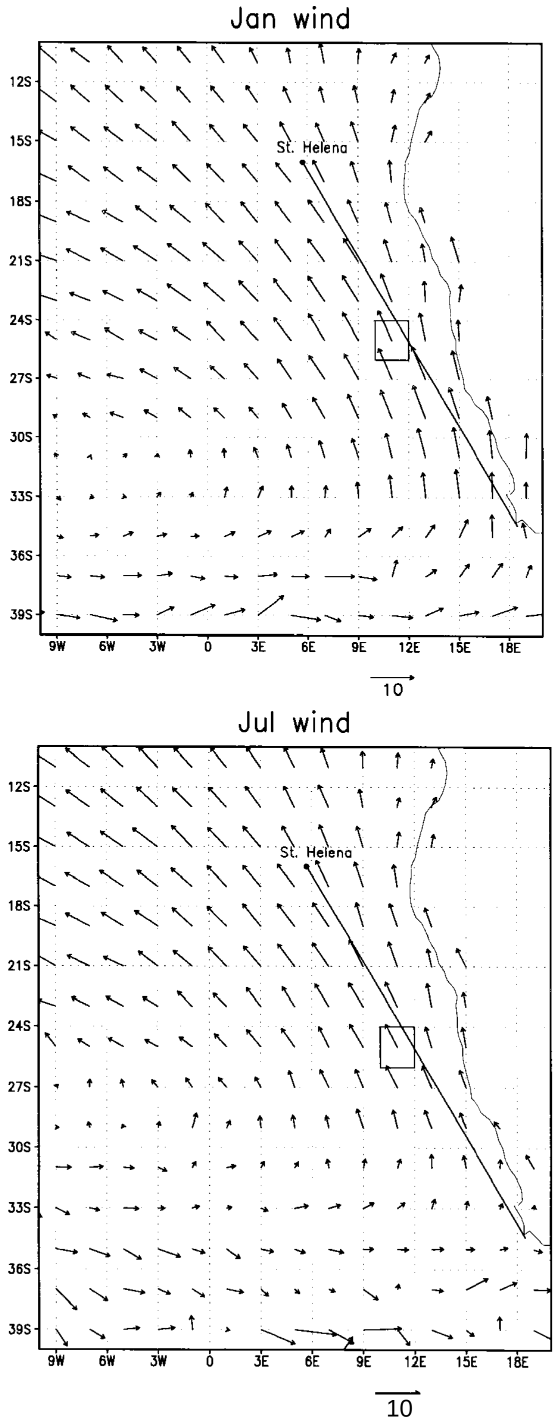


Figure 2: Mean January and July wind vectors (m/sec) for the South Atlantic from COADS (Woodruff et al. 1993).

Entries in the log of that downwind passage (from 22 March to 21 April 1820) are given in Table 2; they indicate that the “steady breeze and clear weather” typical of the Trades was repeatedly replaced by squalls and rain, slowing progress in a “heavy confused sea”.

This is a good example of special conditions, explanations for which must be sought from log entries. No passages are on record between 1650 and 1680, a period of troubles at home and abroad. In England the Puritan revolution had culminated in the execution of Charles I in 1649 and ended with the formation of Cromwell’s Protectorate. The Navigation Act (1651) requiring all commerce to be carried by English ships led to wars with Spain in 1655 and with Holland in 1652-54 and 1665-67. The Dutch occupied St. Helena in 1672 but were driven out a year later. The EEIC in the 1680s resumed operations with a new Royal Charter that left the Company in control of St. Helena until it reverted to the Crown in 1835.

A statistical summary of the Cape-St. Helena passages is provided in Figure 4 with decadal frequency parameters (numbers of passages, medians, quartiles, and minimum durations). The median durations and their interquartile ranges had minima in the 1760s and, after increasing during the 1770s, decreased again until the 1820s. Aside from nautical and technical improvements such trends could reflect a repeated strengthening of the Trades, with the climatic implications discussed e.g. by Fletcher (1994).

However, differences in passage durations could also have been produced by changes in directional constancy. This is suggested by Figure 5 for the only two decades with sufficient passages in each season. The distinctly slower winter passages may be explained as combined effects of weaker and directionally more variable winds.

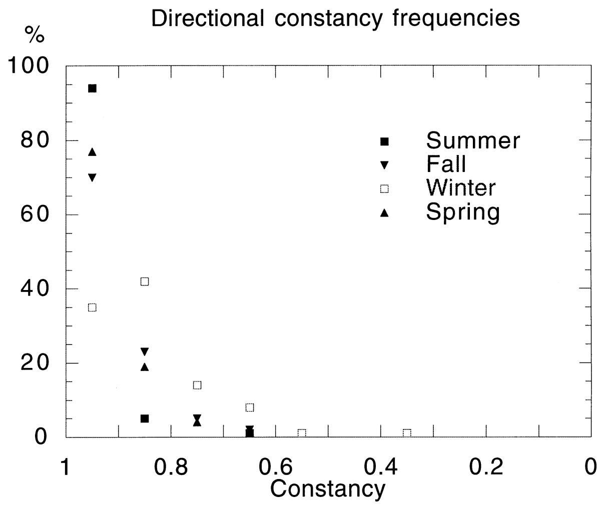


Figure 3: Seasonal frequencies of the directional constancy of the 1950-1993 COADS winds near 25°S 11°E.

4. Further work

To expand this pilot study, entries of weather, winds, and positions must be extracted from individual logs; also line-measured rates of progress which (unlike astronomically determined rates of progress) do not include changes in the speed of the Benguela current. Assuming the fastest passages to occur in the undisturbed Trades, some of the years with systematically slower passages deserve special study.

Other potentially interesting years are suggested by historical reports. Examples are prolonged droughts that prevailed at St. Helena in the 1710s as well as during the 1750s (Gosse 1938).

Table 1
Durations of passages between the Cape of Good Hope and St. Helena Island by the good ship *St.Helena* under Captain Augustus Atkinson.

St. Helena to Cape	upwind days	Cape to St. Helena	downwind days
July 18 to August 11, 1819	24	September 7 to 20	13
October 13 to November 8	26	December 14 to 27	13
January 29, 1920 to February 22	24	March 22 to April 21	30
May 27 to June 18	22	July 17 to 27	10
October 17 to November 13	27	December 5 to 18	13
January 1 to 27, 1821	26		

Table 2
Entries in the *St. Helena* log (L/MAR/B/327A) for 22 March through 21 April 1820.

30 March:	First part squally and rain, latter a steady breeze and clear, with much sea throughout.
31 March:	Squally and rain throughout.
1 April:	First part squally & rain, latter a steady breeze and clear weather.
8 April:	Latter hard squalls & a confused sea.
9 April:	Hard squalls with a heavy confused sea.
10 April:	First part fresh breezes & hard squalls, latter calms & squalls with a heavy confused sea.

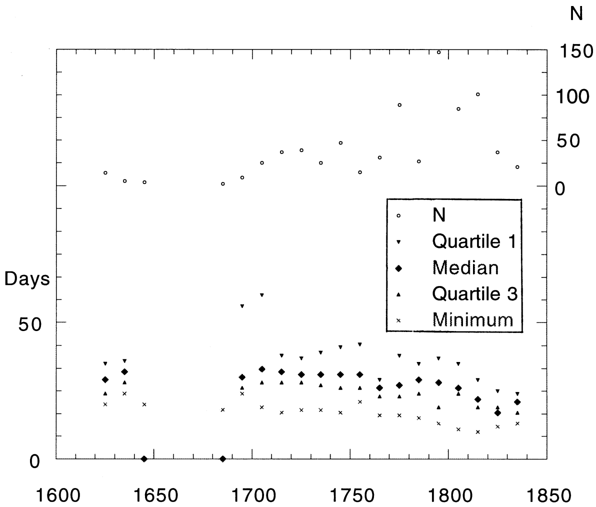


Figure 4: Decadal statistics of passage durations for Cape-St. Helena.

They ended in 1718 and 1756 with flood-producing rains which could have resulted from the southern lows mentioned by Lamb (1957); or from southward displacements of the North Atlantic High and the intertropical convergence, which have been found to be associated with droughts in the Sahel and in Northeast Brazil (Hastenrath 1990, 1993). This could be further tested with log entries for the passages from St. Helena to England.

Finally there are the various routes fanning out from the Cape towards eastern destinations ranging from Arabia to Australia, each representing a different wind system waiting to be studied along the lines here developed.

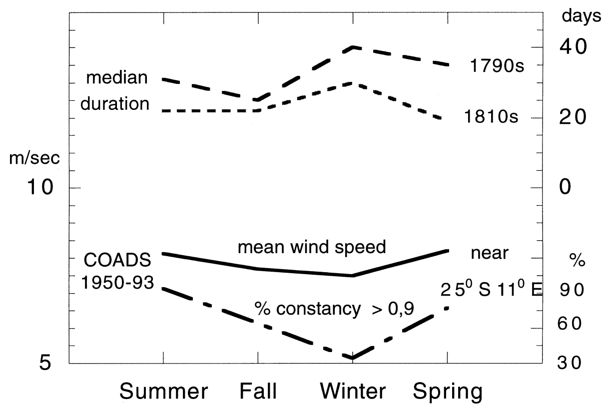


Figure 5: Seasonal mean passage durations during two decades, and mean wind speeds and directional constancies of the 1950-93 COADS winds.

References

Fletcher, J.O., 1994: The importance of COADS winds for understanding climate change. *Proceedings of the International COADS Winds Workshop, Kiel, Germany, 31 May - 2 June 1994*, Diaz, H.F. and Isemer, H.-J. (Eds.), NOAA Environmental Research Laboratories, Boulder, Colorado, and Institut für Meereskunde, Kiel, 4-11.

Gosse, P., 1938: *St. Helena, 1502-1938.* Cassell, London.

Hastenrath, S., 1990: Decadal-scale changes of the circulation in the tropical Atlantic sector associated with Sahel drought. *International Journal of Climatology,* **10**, 459-472.

Hastenrath, S., 1993: Circulation anomaly mechanisms in the tropical Atlantic sector during the Northeast Brazil rainy season. *Journal of Geophysical Research*, **98** (D8), 14917-14023.

Lamb, H.H., 1957: Some special features of the climate of St. Helena and the Trade-Wind zone in the South Atlantic. *Meteorological Magazine,* **86**, 73-76.

Van Loon, H. (Ed.), 1984: *Climates of the Oceans.* World Survey of Climatology, Vol. 15.

Woodruff, S.D., Lubker, S.J., Wolter, K., Worley, S.J., Elms, J.D., 1993: Comprehensive Ocean Atmosphere Data Set (COADS) Release 1a: 1980-92. *Earth System Monitor*, **4** (1), 1-8.

Global Oceanographic Data Archaeology and Rescue (GODAR)

Robert D. Gelfeld
U.S. National Oceanographic Data Center
World Data Center A, Oceanography

1. Introduction

The National Oceanographic Data Center (NODC) is one of the national environmental data centers operated by the U.S. National Oceanic and Atmospheric Administration. The NODC holds global physical, chemical, and biological oceanographic data collected by U.S. federal agencies, state and local government agencies, universities and research institutions, and private industry. NODC acquires foreign data through direct bilateral exchanges with other countries and through the facilities of the WDC-A for Oceanography (WDCA), which is operated by NODC under the auspices of the U.S. National Academy of Sciences.

Each year the NODC provides oceanographic data and information products and services to thousands of users around the world. Among NODC's data products are a growing number of CD-ROMs holding large, frequently used oceanographic data sets.

The NODC is also now using the global communication capabilities of the Internet to make its vast oceanographic data and information holdings more easily accessible to the worldwide oceanographic research community and other users in industry, academia, government, and the general public.

To determine the role of the world ocean as part of the earth's climate system, the international scientific community needs to have access to the most complete data bases of historical oceanographic data possible. Many historical oceanographic data are not available to the international community in digital form. "Data Archaeology" is the term used to describe the process of seeking out, restoring, evaluating, correcting, and interpreting historical data sets. "Rescue" refers to the effort to save data at risk of being lost to the science community.

There are additional reasons for building the most complete oceanographic data bases possible:

a) The international scientific community is often called on to advise national and international bodies on such issues as climate change. Hence, the international oceanographic and climate communities should have access to the most complete digital oceanographic data bases possible.

b) Substantial resources have been or may be allocated for national and international ocean and climate programs such as Climate and Global Change, Tropical Ocean and Global Atmosphere (TOGA), World Ocean Circulation Experiment (WOCE), Global Ocean Ecosystems Dynamics (GLOBEC), and Joint Global Ocean Flux Study (JGOFS), and for the establishment of a Global Ocean Observing System (GOOS). Planners of such programs should have access to all historical oceanographic data in order to optimize measurement strategies. Scientists analyzing data from such programs need historical data in order to place new data in perspective.

Based on personal observations and discussions with scientists and data managers from the international community, it is clear that substantial amounts of historical oceanographic data exist only in manuscript or analog form. In addition there are data in digital form that are not available due to neglect or other reasons, for example, data that were used for one purpose by a scientist or institution but never archived at an international data center.

Effectively these data are not available to researchers and other users. Whether recorded on paper or magnetic tape, these data are at risk of being lost due to:

- media degradation such as fading ink or magnetic fields;
- environmental catastrophes such as fires and floods;
- simple neglect;
- the retirement of individuals who know how to access these data, or know the associated metadata that make them useable to other scientists.

The NODC/WDCA, Oceanography has begun a project to identify and rescue historical oceanographic data in jeopardy of being lost. This project, called the Global Oceanographic Data Archive and Rescue (GODAR), is sponsored by the Intergovernmental Oceanographic Commission's (IOC) International Oceanographic Data Exchange (IODE) Program.

The data recovered via GODAR have been made available without restriction to the international scientific community as a 10-volume CD-ROM set called the World Ocean Atlas 1994.

The use of modern state-of-the-art technology in locating and rescuing historical oceanographic data sets has been very valuable. However, plain old detective work, hard science, and human intervention have also proven to be equally important.

Many historical oceanographic data sets still remain buried in scientists and researcher's offices and the use of modern technologies serve only as an aid to their recovery.

Physical, chemical, and biological oceanographic data, as well as surface marine meteorological observations, are the specific types of data that the GODAR project focuses on. Initially, most data digitized or otherwise rescued have been physical parameters, particularly temperature data from bathythermograph profiles. Data from the open ocean have been given the highest priority. Time series records longer than ten years have also been accorded high priority.

2. History of NODAR and GODAR

The NODC Oceanographic Data Archaeology and Rescue (NODAR) project and IOC/IODE GODAR project had their origins at a meeting held at NODC/WDCA, Washington, D.C., in September 1990. The meeting was supported by the U.S. Climate and Global Change Program. Scientists and data managers from several countries and international centers including the Soviet Union, Korea, Japan, Chile, Australia, United States, and the International Council for the Exploration of the Seas (ICES) met informally to discuss the state of historical oceanographic data, in particular to discuss the loss of data due to media degradation.

The results of the meeting led to the establishment of various national and now an international project that are known generically as “Oceanographic Data Archaeology and Rescue” projects. For example as a result of that workshop, NODC received funding to begin the NODAR project. Similar projects were started at ICES and World Data Centers B and D (WDCB, WDCD).

An international meeting known as the “Workshop on Ocean Climate Data” was arranged by the IOC and NOAA at Greenbelt, Maryland, USA (Churgin, 1992). The meeting was sponsored by the Commission of the European Communities (CEC), International Council of Scientific Unions (ICSU), World Meteorological Organization (WMO), ICES, and IOC.

As a result of the demonstrated progress of various national data archaeology and rescue projects, the workshop recommended the expansion of these projects to band together under the umbrella of an existing international organization. Such support facilitates the exchange of data internationally.

A proposal for a GODAR project was submitted to the Fourteenth Session of the IODE meeting held in Paris, France from 1-9 December 1992.

The IODE recommended to the IOC that this project be adopted as an IOC project. During the March 1993 IOC Assembly meeting, the IOC adopted the proposal for a GODAR project. Specifically the GODAR project emphasizes:

- Digitization of data now known to exist only in manuscript and/or analog form. This effort will have highest priority of all activities.
- Rescue of digital data that are at risk of being lost due to media decay or neglect.
- Ensuring that all oceanographic data available for international exchange are archived at two or more international data centers in digital form.
- Preparing catalogues (inventories) of:
  - data now available only in manuscript form;
  - data now available in analog form;
  - digital data not presently available to the international scientific community.
- Performing quality control on all data and making all data accessible on various media including CD-ROMs as well as standard magnetic tape.

A series of GODAR workshops have been held to bring scientists, administrators, and data managers from nations of a specific geographical region to focus on problems of historical oceanographic data preservation and access.

These workshops helped lay the groundwork for a major upgrading and modernization of the ocean data management system for the entire region and generated faster regional exchange methods. The workshops held to date include:

a) GODAR I Regional Workshop for Member States of Eastern and Northern Europe held in Obninsk, Russia May, 1993

b) GODAR II Regional Workshop for Member States of the Western Pacific held in Tianjin, China March, 1994

c) GODAR III Regional Workshop for Member States of the Indian Ocean held in Goa, India December, 1995

d) GODAR IV Regional Workshop for Member States of the Mediterranean Sea held in Malta April, 1995

e) GODAR V Regional Workshop for Member States of South America held in Cartagena, Colombia October, 1996

f) GODAR VI Regional Workshop for Member States of Africa held in Accra, Ghana March 1997

3. Methods of investigation

From the inception of the national archaeology and rescue projects at various centers, efforts were coordinated to avoid duplication of effort and to maximize the use of scarce resources. Joint activities include the exchange of data, data distribution plots, catalogue information about data holdings, and the exchange of scientists and data managers between centers. Simultaneous “rescue” and exchange of data was emphasized for two reasons:

- some data are at risk of being lost forever if not saved immediately;
- the international research and administrative communities needed a demonstration of how quickly the project could make previously unavailable data accessible in digital format.

Perhaps the most valuable technique to quickly describe data holdings is to produce data distribution plots and tables of the number of profiles on a year-by-year basis for each major measurement type. Levitus and Gelfeld (1992) have done this for each of the major NODC digital archives. This work showed the distributions of NODC holdings for all countries. NODC/WDCA has prepared similar plots on a country-by-country basis and distributed such summaries to data centers, scientists, and institutions in approximately 20 countries as of December 1993.

These summaries have generated much interest and resulted in the exchange of more information and data, e.g., through the generation of such summaries the NODAR/GODAR project was able to conclude that substantial amounts of Canadian MBT profiles (40,000) that already existed in digital form were not in the NODC/WDCA files. These profiles were acquired by NODC/WDCA as part of the GODAR project and were incorporated into the NODC/WDCA archive.

Examination of cruise reports and monographs have also supplied both data and metadata not available in digital form. By examining the report series “Special Scientific Reports...Fisheries” of the Fish and Wildlife Service of the Department of Interior, we found chemical and biological data from U.S. research cruises that had never been digitized even though in some instances the physical parameters of these profiles are part of the NODC digital archive.

Metadata about these biological and chemical parameters, such as measurement techniques, are available in these reports in detail. We also found observations of surface salinity and phosphate that had been measured at the same time that MBT profiles were made. We have begun digitizing all these data.

One example of a monograph providing substantial information about historical oceanographic observations is the work by Muromtsev (1958). He provides a list of “deep-water” observations made in the Pacific Ocean for the 19th and 20th centuries. In addition, Muromtsev provides a table containing the actual temperature measurements from profiles made in the Pacific Ocean during the period 1804-1873.

Although the accuracy of these measurements needs to be examined to determine whether the data are useful, at least the data have been tabulated in a single location with metadata that will allow the determination of the suitability of these data for scientific studies. Another example is the U.S. National Academy of Sciences report by Vaughn (1937).

This document contains contributions by Defant, Sverdrup, Helland-Hansen, and Wust who, in total, describe a great deal of the profiles made globally up to the date of the report. Some of this material, which may also have been published in other reports or atlases (e.g. Wust, 1935), describes data not archived at NODC/WDCA in digital form.

4. Results

When one considers that the global historical data base of temperature data has increased by more than 1.4 million profiles in the past two years, and the results of the GODAR workshops (which have identified on the order of another one million profiles that are in manuscript form), it becomes clear that the international scientific community will have access to a much more complete data base than ever thought possible. This includes many MBT profiles that will help determine interannual variability of the upper ocean thermal structure. In addition, numerous Oceanographic Station Profiles that include temperature, salinity, and other parameters are now available.

5. Future work

The data sets received so far are being processed at NODC/WDCA and will be distributed internationally without restriction. Several media will be used for distribution including CD-ROMs. CD-ROMs have several advantages for the distribution of in situ oceanographic data:

- there exists an international standard for reading and writing information to CD-ROMs;
- CD-ROMs have relatively large storage capacity for the in situ oceanographic data sets we are distributing;
- they are relatively inexpensive to master and duplicate;
- their small size, which makes physical distribution easier as compared with magnetic tapes;
- a relatively small investment is required to obtain a personal computer and CD-ROM drive for accessing these data. Neither mainframes or even minicomputers are required for digital access to large data bases.

While we fully expect the data, and analyses of these data to be made available over an electronic network, we believe that CD-ROM technology best serves the international distribution of these data. Some countries simply do not yet have the required network capability.

During 1998 the following amounts of additional data are expected to become available via CD-ROM:

a) 140,000 Ocean Station Profiles
b) 103,000 High Resolution CTD Profiles
c) 253,000 MBT Profiles
d) 103,000 XBT Profiles
e) 600,000 Plankton Observations
f) 120,000 Chlorophyll Profiles
g) 300,000 XBT Profiles from the Global Temperature Salinity Profile Project (GTSPP)

These data will be known collectively as “World Ocean Database 1998” (WODB98). This new data base will include all GODAR data, plus all data held at NODC/WDCA, Washington, D.C. It will be a joint product of the IOC/IODE system and the ICSU/WDC system.

WODB98 will contain parameters not previously included in the WOA94 database such as meteorological data, nitrite, pH, alkalinity, chlorophyll, and plankton observations as well as metadata such as originator and NODC cruise numbers. Distribution will be made using CD-ROMs and magnetic tape media.

Numerous institutions from the international community are now participating in the GODAR project. The navies of several countries have been declassifying oceanographic data and making these data available internationally without restriction. For example, the Russian Navy is participating in the GODAR project by making manuscript data available for digitization and distribution.

The British Navy released for international distribution in 1994 approximately 175,000 bathythermograph profiles. Some of these data were observed as early as 1945.

More comprehensive data bases will lead to a better description of both the mean state and the variability of the world ocean. This will lead to better understanding of the role of the world ocean in the earth's climate system. Such work is international in nature.

The recent atlas by Olbers et al. (1992) is an excellent example of an analysis of oceanographic data that has benefited by international cooperation resulting in the exchange of data.

It is our hope that other investigators, research groups, and data centers from all countries with marine research and operational programs will participate in and benefit from this project.

Appendix
Minimum Requirements for Quality Assurance of Oceanographic Data

Data Acquisition

a) the data should be in a media-compatible format (preferably ASCII)
b) the data format documentation should contain all available metadata, which include:

    (1) techniques used to sample the data (including number of replicates)
    (2) instrumentation used to collect the data (including manufacturer and detection limits)
    (3) standards and protocols which apply to the data

c) the primary keys of the cruise and station should be supplied including: time, date, position, ship, originator's station number
d) the units for the parameters of the data taken should be supplied
e) the data supplied should also include all physical parameters (i.e. temperature and salinity) and all other parameters measured including data calibration information

Data Formatting

a) the data should include all depth levels and all measurements in all cases
b) the precision at which the data were measured should be supplied
c) any meteorological information should be available
d) the source of the data (i.e. in situ or remote) should be indicated
e) ROSCOP-like information should be supplied (i.e. Principal Investigator, parameters measured, and the estimates of the numbers of parameters)
f) the definition of time should be indicated (start of sampling)
g) there should be a comparison to operational standards such as WOCE, JGOFS, etc.

References

Levitus, S. and Gelfeld, R., 1992: *National Oceanographic Data Center Inventory of Physical Oceanographic Profiles - Global Distribution by Year for All Countries*. Key to Oceanographic Records Documentation No. 18, NODC, Washington, D.C., 242 pp.

Muromtsev, A.M., 1958: *The Principal Hydrological Features of the Pacific Ocean*. Main Administration of the Hydrometeorological Service of the USSR Council of Ministers, State Oceanographic Institute, (Published for the National Science Foundation, Washington, D.C. by the Israel Program for Scientific Translations, Jerusalem, 1963, 417 pp.)

Olbers, D., Gouretzki, V., Seib, G., and Schroeter, J., 1992: *The Hydrographic Atlas of the Southern Ocean*. Alfred-Wegener-Institute for Polar and Marine Research, Bremerhaven.

Vaughn, T.W., 1937: *International Aspects of Oceanography: Oceanographic Data and Provisions for Oceanographic Research*. National Academy of Sciences, Washington, D.C., 225 pp.

Wust, G., 1935: *The Stratosphere of the Atlantic Ocean*, English Translation edited by W.J. Emery, Aprevind.

---oooOooo---

Archival Climate History Survey (ARCHISS)

Mike Baker
ICA, Unesco, WMO Archival Climate History Survey

Abstract

The Archival Climate History Survey was launched in 1988 in order to obtain, on a systematic basis, the help of the National and State Archives to locate and retrieve relevant measured and proxy historical climatological information. Initially the International Council of Archives (ICA), the International Council of Scientific Unions (ICSU), the United Nations Educational, Scientific and Cultural Organization (Unesco) and the World Meteorological Organization (WMO) participated in a pilot project based on Europe. The proposal was in part a response to the conclusion of the Intergovernmental Panel on Climate Change (IPCC) that to increase the current assessments of climate change a more complete spatial record and longer time series of historical climate information are required.

1. European pilot project

In order to examine the validity of the proposed survey it was decided to carry out a pilot project in Europe. The help was sought of archivists in five countries (France, Germany, Italy, Spain and U.K.) and archives were searched in Arras, Galicia, Lancashire, Lübeck, Lucca, Milan and Ulm to try to reconstruct weather conditions in the 50-year period from 1725 to 1775. In order to facilitate comparisons of the information standard forms were produced by archivists and climatologists which made provision for the recording of meteorological and hydrological informations and also of archival information that would facilitate finding the material again. No attempt was made to tag information of potential interest for two main reasons: i) the surveys were made by archivists with little knowledge of climatology; and ii) searches for information of potential interest can be both time-consuming and distracting.

The results were both positive and diverse. For example, the search in the Arras region gave a wide range of instrumental and proxy information for Northern France; that in Lucca produced several long-time series of instrumental observations by individual observers, including one for the period 1777 to 1827 giving daily measurements of temperature, pressure and precipitation; while that in Lübeck provided information about riverine and estuarine conditions.

In the course of the European pilot project studies were made into the use of different methodologies for the reconstitution of various meteorological parameters, but particularly temperature, using incomplete information, both instrumental and proxy, so as to obtain series covering longer periods of time. For example, average/mean temperatures at Lille were reconstituted from 1757 to 1793 using a multilinear regression. Studies into the most effective methods are continuing. The intention is to prepare series as long as possible and to make these available as quality controlled data sets for use by the scientific community so as to help to improve studies of climate change and variability and in the validation of climate model outputs.

1.1 Scientific evaluation

The results of the project were submitted for scientific evaluation to a panel of four climatologists appointed by ICSU, Unesco and WMO. The evaluation underlined the great value of the pilot project and stressed the enormous potential of such searches for developing or improving climatic baseline data sets.

Some of the points brought out in the evaluation include: i) high-density serial data, especially instrumental observations, are the most valuable to the climatologist; ii) the importance of good metadata; iii) the large amount of time required to find potentially useful information and the difficulty of forecasting how much time might be needed; iv) the climatologists need to be prepared to inspect the retrieved information and to put it into digital form so as to be useful in global change studies.

The evaluation also indicated that the project should be expanded and become operational. It suggested that several steps are necessary to ensure that the work of the archivists is useful. These include:

a) searches for data should be undertaken in South America, Central America and Africa;

b) initial searches should focus on measured information;

c) a group of climatologists should be selected to perform the initial evaluation of the information;

d) a climatology Centre should be selected for the digital entry of the data and their quality control;

e) a plan should be devised for the exchange of these digital data throughout the World Data Centre system;

f) a formal coordination committee should be established.

As a result of the pilot project and the preliminary searches in Latin America it has been decided to keep to a minimum the handling of the archives, because many are brittle, and that wherever possible the transfer of the information from the archival material to machine-readable format should be done in situ. For Mexico and Cuba (about which I will discuss later) this has been made possible by contracts with the National Archives for on-the-spot digitization of the information found. Also as a result of these suggestions particular attention has been given in the current phase to the need to ensure that digitized, evaluated data should be readily available to both the national community of climatologists and to the international community by providing a copy of the data to the national authorities and the transmission of the data on disk to the Global Historical Climatology Network (GHCN) data base.

2. Latin America and the Caribbean

After careful consideration of the positive scientific evaluations of the material obtained in the European region and after consultation with climatologists it was decided to concentrate on Latin America and the Caribbean. Several possible countries were selected and it was finally decided to begin the first archival search phase of the project in Cuba and Mexico, to be followed by Brazil, Colombia and Ecuador. Contacts have been made with the National Meteorological Services to ensure maximum cooperation, to avoid unnecessary duplication of effort and where appropriate to provide the benefit of the experience of the national archivists in the archiving of material at National Meteorological Services.

A presentation of the ARCHISS project was made in 1995 by G. Tori, State Archivist Lucca, to a meeting of the *Asociacion Latineamericana de Archivos* (ALA) at which there were representatives of 15 National Archives in Latin America. The project received a warm welcome and immediate offers to participate were made by archivists from several countries.

2.1 MEXICO: Archivo General de la Nacion

In February 1995 a six-month search began of national and state archives in Mexico for meteorological data that are not included in either the global or the national meteorological data bases. This has been completed and a report on this first phase has been submitted to Unesco, WMO and the ICA.

An on-the-spot scientific evaluation was then made in July 1995 by K. Davidson, Chairman of the WMO Working Group on Climate Data. In his evaluation he stressed the success of the search and indicated that very significant amounts of data were brought to light, including a long-term data set that provides daily data from the late 1700s. For example, instrumental data for about 30 stations throughout Mexico have been found from the mid-18th Century in 3000 sources. In particular, studies of the *Diario Oficial* from 1753 to 1894 provided data on the weather (for example, daily measurements of temperature, pressure, winds, precipitation and humidity) during this period, but mainly in the years 1868 to 1894.

A map (not shown) gives the location of the stations, three of which are on the Pacific coast and of potential interest in historical studies of the El Niño - Southern Oscillation events. In addition the archivists searched through a large amount of material for proxy data but it was decided that this was too time-consuming, less effective and provided less information.

During the second phase of the research monthly summaries were made of the data from more than 40 stations during the period 1870 to 1910; this has been followed by quality control and digitization of the data and checking of the original locations of the stations, some of which are no longer in the original locations. Mean temperature data sets have been transmitted to the GHCN for this 40-year period for between 40 and 60 stations (the number tends to vary with the year but with a maximum in 1895-96).

2.2 CUBA: Archivo Nacional

In April 1995 different archive groups were identified some of which contained archives of potential interest for proxy data going back to the 15th Century. Following the development of a strategic plan the first phase of the searches in Cuba began.

An evaluation of the results of the survey in Cuba was carried out by D. Phillips, Rapporteur on Data Rescue for WMO Regional Association IV for North and Central America. In his evaluation he drew attention to four important time series:

a) The century-long time series identified at the *Real Colegio de Belén de la Orden de los Jesuitas* in Old Havana, later moved to the *Observatorio del Real Colegio de Bélen* in Marianao on the outskirts of Havana. The observing programme included temperature, precipitation, relative humidity, atmospheric pressure and wind. Although the time series is broken by a move in location into two sets: 1853 to 1926 and 1926 to 1953, careful statistical analysis by climatologists with the Institute of Meteorology might possibly generate a single homogeneous record.

b) Observations for Havana can be traced from 1791 to 1839 and beyond. To achieve such a clean time series will need some statistical analysis from at least 10 identified data sources and careful data construction by climatologists, but it is potentially a very significant data set.

c) Another significant time series (1903 to present) exists for Casablanca, site of the Institute of Meteorology.

d) Especially significant is the continuous five-century data record of the occurrence of cyclones. The data set lists the day and month of cyclone occurrence from 1494 to 1870.

The current retrieval and digitization phase is almost complete. A set of disks giving the digitized information has been sent to the GHCN for quality control and checking. It is expected that these data will soon be available from the GHCN data-base.

2.3 COLOMBIA: Archivo General de la Nación

Following a positive preliminary survey a contract is in the process of being signed with the National Archives in Colombia and it is expected that the retrieval and digitization phase will begin shortly.

3. Future projects

In addition to the studies projected in Brazil, Colombia and Ecuador the possibilities are being investigated of carrying out a supplementary survey to extend the search to archives in Europe with climatological information concerning Latin America and the Caribbean and to reports of scientific expeditions, such as those of G. Markgraf, A. von Humboldt, A. Codazzi, etc. In addition it has been suggested that particular attention be given to archives of data-sparse areas where there is likely to be a response to climate system forcing, for example archives in the Pacific region that might have historical climate information that would help extend backwards in time our knowledge of El Niño - Southern Oscillation (ENSO) events.

It has also been suggested that:

i) cooperation between the scientific and archival communities in Europe be extended to Eastern Europe and the Southern Mediterranean; and

ii) the positive outcome of the European pilot project should be followed up, possibly by searches in the archives for meteorological data obtained during the former colonial periods. These searches will depend on the interest and cooperation of the concerned communities and appropriate funding.

The ideal result would be to have daily instrumental records of the basic meteorological variables, however in many parts of the world these do not exist. The ARCHISS project in addition to trying to locate data obtained with reliable instruments that are not in the global data sets is also trying to provide information on significant signals in the climatological record covering the longest possible periods, and wherever possible at least the dominant weather pattern month by month, or failing that season by season for the principal climatic regions of the world.

The exciting discoveries made as a result of the searches in Cuba and Mexico show clearly the need to continue and extend the ARCHISS searches in Latin America and the Caribbean. Such searches should be complemented by searches in Europe of archives that have numerical data of climate-sensitive regions that were obtained in the colonial period and of scientific expeditions of more than a few months' duration.

4. Historical surface marine data and metadata

Up to the present, ARCHISS searches have not been concerned with historical surface marine data and metadata although climatological information has been obtained from some coastal stations. Contacts have, however been maintained with representatives of the Global Ocean Data Archeology and Rescue (GODAR) programme of the Intergovernmental Oceanographic Commission (IOC) and the author drafted the following recommendation that was adopted by the GODAR Regional Workshop in Cartagena de Indias in October 1996:

> "The Workshop noted the wide range of programmes to rescue data and to recover climatological data from archives being implemented by different international organizations such as ARCHISS, GODAR (IOC), DARE (WMO), IGBP-PAGES (ICSU), etc. and recommends that efforts be made to establish and/or ameliorate links with the organizers of such programmes in an endeavour to improve cooperation and to avoid unnecessary duplication. A meeting of representatives, of these and other similar programs is recommended for 1997."

Discussions are in progress about the possibilities of holding such a meeting in mid-1998.

5. Publications

A report *Using Archival Resources for Climate History Research* was published by Unesco in 1994 as IHP 4, within the framework of the International Hydrological Programme.

A *Report on the Status of the Archival Climate History Survey (ARCHISS)* has been published by WMO in 1996 as WCDMP-No. 26 (WMO-TD No.776).

Briefer reports about progress have appeared in the *ICA Newsletter, Unesco Features* and the *WMO Bulletin.*

6. Finance

The continuation of ARCHISS activities will be dependent on adequate funding. Up to the present the major financial inputs have come from Unesco and WMO with the Meteorological Services of Canada and the USA providing funds for the archival searches in Latin America.

The other sponsors of ARCHISS have provided contributions in kind and several national archives have provided help in the preparation of the forms and in the initial studies. The continuation of the project is dependent on extra-budgetary funds being made available to ICA, to Unesco or to WMO.

7. Conclusions

The European pilot project and the results obtained so far in the Latin American-Carribean project prove that the archival survey for climate history (ARCHISS) is a particularly useful and effective method for completing the climate record. The involvement of professional archivists, who have proved to be particularly enthusiastic about the ARCHISS project, has brought a new and positive element into such surveys.

Archivists are basically impartial with regard to questions of climate change, are used to searching archival material, generally have experience of reading difficult manuscripts and often have knowledge of complex and ancient language forms. In addition the experience obtained so far in Cuba and Mexico shows that the development of closer cooperation between the national archival and climatological communities brings great benefits and it seems probable that this will also be the case in other Latin American and Caribbean countries and in other regions of the world.

We can expect wide variations in the amount of data, particularly serial data, that will become available through searches in archives in different regions of the world. It also seems probable that the criteria used for the selection of information may have to vary widely, in some regions the information may only give indications of the monthly weather characteristics and in others be so sparse that it gives only seasonal indications, but even these can be useful when carrying out long-term statistical studies of climate fluctuations and change. This extension of our knowledge of climate change and variability, as requested by the Intergovernmental Panel on Climate Change, will be of interest not only to scientists and to decision makers but also to the public at large.

---oooOooo---

Part 2

Processing and usage of digitized historical data and metadata

Detecting Climatic Signals from Ship's Datasets

Pascal Terray
Laboratoire d'Oceanographie Dynamique et de Climatologie, Paris, France

1. Introduction

Marine ship observations over the vast oceanic regions are crucial to studies of climate variability on timescales from the seasonal to multidecadal. However, any climatic analysis of this historical record is hampered by two difficult problems, namely:

- The systematic instrumental errors which contaminate the ship observations. For example, it is well-known that most of the ship-reports before 1940 contain a large majority of uninsulated bucket Sea Surface Temperature (SST) measurements which are biased low, while the data after the 1940s are mostly injection or insulated bucket SST measurements which are biased high (Bottomley et al., 1990).
- The irregular space-time sampling of the ship-reports. For example, COADS summaries provide meteorological variables in the form of monthly means for 2° x 2° latitude-by-longitude cells (Woodruff et al., 1987). In such datasets, the number of observations used to compute a particular monthly mean reflects the number of ships that cross the box that month. Thus, for a particular month, one cell's mean may be computed from hundreds of observations, while others may be based on only a few, and there may be many cells with missing means due to the poor spatial and temporal coverage outside the main shipping lanes.

The former problem is particularly relevant to studies of multidecadal variability, and has led researchers to design instrumental correction procedures for the meteorological and oceanic fields derived from ship-reports and used for assessing climatic changes, e.g. SST and wind.

The latter problem attends almost all climatic studies from seasonal to multidecadal timescales, but is particularly relevant to the interannual to multidecadal. The classical solution to cope with this problem is to use some kind of objective analysis. This technique spatially smooths the oceanic fields by filling the data-void areas with reasonable values which are a linear combination of climatology and anomalies observed in the neighborhood of each grid's cell. The drawbacks of this solution are: First, the need for a very good climatology which has to be constructed before the analysis. Second, the oceanic fields derived from objective analysis are generally over-smoothed with the undesirable consequence of a decrease in the spatial resolution of the data.

The main objective of this work is to present a new multivariate statistical method to deal with this last problem. The method may be termed weighted Empirical Orthogonal Function (EOF) analysis or weighted Singular Value Decomposition (SVD) analysis and is a generalization of the traditional EOF analysis, or more precisely, of truncated SVD analysis. This method accounts for the irregular space-time sampling of the ship-reports by the use of weights (a weight is associated with each cell-month entry of the data matrix) in approximating the data matrix by a lower rank matrix in the least squares sense. In contrast, the traditional SVD analysis assumes that all the cells have equal weights in solving the same optimization problem.

The organization of this paper is as follows: first, the formalism of the weighted SVD analysis is presented and its relationships to traditional SVD analysis are outlined. Second, we illustrate with some examples how weighted SVD analyses are useful for extracting seasonal, interannual and multidecadal climatic signals from ship's datasets such as COADS summaries. Finally, we highlight the utility of the weighted SVD analysis for different common tasks in meteorology and oceanography.

2. Theory

The widespread acceptance of EOFs for data reduction purposes, to aid in determining the variability of oceanic and atmospheric fields, or to identify coherent modes of atmospheric parameters suggests that the adaptation of this method to ship's datasets can provide us an improved tool to extract climatic signals from such noisy data. However, traditional EOF analysis is not well-adapted to ship's datasets since the method gives the same weight to all the data matrix entries without taking account of the irregular space-time sampling of the ship's reports when determining eigenvectors and principal components. Moreover, it is impossible to compute EOFs if some data are missing.

By contrast, the new method of analysis we will develop takes directly into account these uncertainties of the data while estimating the EOF model. In order to introduce this new method, it is first useful to review some of the optimal properties of traditional EOFs (Kutzbach, 1967). This is a necessary step to understand the new method.

Let X denote an p x n data matrix consisting of n time observations (columns of X) for p grid cells or stations (rows of X). In the complete case, where X is a full matrix with no missing values, the "full EOF model" can be expressed as a matrix product, X=E $ C where E is an p x p orthogonal matrix, E^T $ $E=I_p$ whose columns are the eigenvectors of the p x p symmetric matrix,

$$R = \left(\frac{1}{n}\right)\left(X \bullet X^T\right)$$

If the data are centered in rows, this last matrix is simply the covariance matrix between the grid's cells. Furthermore, the elements in the *i*'th row of C represent time variations associated with the *i*'th eigenvector.

One of the most important optimal properties of EOFs, especially for data reduction purposes, is that maximum inertia of the data matrix is explained by choosing in order the eigenvectors associated with the largest eigenvalues of R.

More precisely, it can be shown that the fraction of the total inertia, V_k, explained by the eigenvectors associated with the *k* largest eigenvalues can be obtained from

$$V_k = \sum_{l=1}^{k} l_l \div \sum_{l=1}^{\min(p,n)} l_l$$

In the application of EOFs to highly correlated fields such as those commonly analyzed in meteorology, this means that a large portion of inertia can be accounted for by retaining only the first few eigenvectors. This leads to define a "restricted *k* EOF model" to approximate and to study the data: $X \approx E_{.k}C_{k.}$ where $E_{..k}$ stands for the first *k* columns of E, and similarly $C_{k.}$ designates the first *k* rows of C.

The optimal properties of EOFs can be stated directly in terms of this restricted EOF model as follows: the *k*-component model forms an optimal approximation to the original matrix in the sense of least squares. That is, the minimum of

$$\left\| X - (A \bullet B) \right\|^2 = \sum_{ij} \left(X_{ij} - \sum_{l=1}^{k} A_{il} B_{lj} \right)^2 = f(A,B)$$

on all *p* by *k* matrix A and all *k* by *n* matrix B is obtained by taking $E_{.k}$ and $C_{k.}$ as A and B, respectively. Moreover, this minimum is equal to

$$\sum_{l=k+1}^{\min(p,n)} \lambda_l$$

Now let X denote a typical ship's dataset such as COADS 2° lat x 2° long trimmed monthly means for some area and historical period. In order to take into account the sampling properties of this ship's dataset while estimating a restricted *k* EOF model, we may correspondingly seek a minimum of

$$f^*(A,B) = \sum_{ij} W_{ij} \left(X_{ij} - \sum_{l=1}^{k} A_{il} B_{lj} \right)^2$$

on all *p* by *k* matrix A and all *k* by *n* matrix B.

Here, W is an *p* by *n* positive weight matrix constructed in such a way that the resulting EOFs and principal components are defined to emphasize the better-observed aspects of the data. In particular, for the extreme case of zero sample size, an entry of the data matrix should play no role in fitting the model; this can be done by assigning zero weights to such cells.

There are several ways to determine this weight matrix in order to take into account that the monthly means for each grid cell are based on samples of widely varying sizes:

a) The simplest method is to set
$W_{ij} = 1$ if X_{ij} is present
$W_{ij} = 0$ if X_{ij} is missing
This will take care of missing values, but gives the same weight to all non-missing cells in the data matrix.

b) Another choice is to fit the *k* EOF model with weights proportional to size samples $W_{ij} = \alpha N_{ij}$ where N_{ij} is the number of ship observations contributing to the cell's monthly mean X_{ij}.

c) A more elaborate strategy is to use some smooth function of the number of observations $W_{ij} = 1 - \exp(-N_{ij} / 6$ where again N_{ij} is the number of ship observations used in computing X_{ij}. For this particular weight function, W_{ij} is in the neighborhood of 1 if $N_{ij} > 10$ and near 0.5 if N_{ij} equals 6.

d) Still another strategy for constructing the weight matrix is to use the inverse of the variance or standard error associated with each grid's cell and month. This information is, for example, available in the distribution files of COADS (Woodruff et al., 1987).

After the weight matrix is constructed, we have to minimize the least-squares problem stated above in order to estimate the EOFs and their associated principal components. Note that this cannot be done by solving some eigensystem as in the traditional EOF analysis, and we have to use non-linear least-squares techniques (Gauss-Newton or Marquardt-Levenberg algorithms) to obtain a solution to our problem. The algorithms used here to minimize f*(A,B) are a generalization of the techniques described in Terray (1995). Once this is done, the last step is to compute the SVD of the A•B product in order to normalize the solution as in the traditional EOF approach. This new method will be referred to as weighted EOF analysis in the rest of this paper.

3. Examples

In order to illustrate the application of this technique and to show that this method allows us to analyze the natural variability exhibited by data of varying reliability, two classical ship's datasets were analyzed using the weighted EOF technique.

Example 1

The first example is a weighted EOF analysis of the January 1993 version of the Global Ocean Surface Temperature Atlas (GOSTA; Bottomley et al., 1990). The data in this atlas are presented as monthly anomalies on a 5° latitude x 5° longitude grid wherever data existed. The data were extracted for the period from 1900 to 1991. The weight function used in this analysis is simply: $W_{ij} = 1$ if X_{ij} is present and $W_{ij} = 0$ if X_{ij} is missing.

Note that this is not a very good choice since this will give the same weight to all non-missing data, but there is no information on the number of ship-reports used to compute monthly anomalies for individual grid cells in the distribution files of GOSTA.

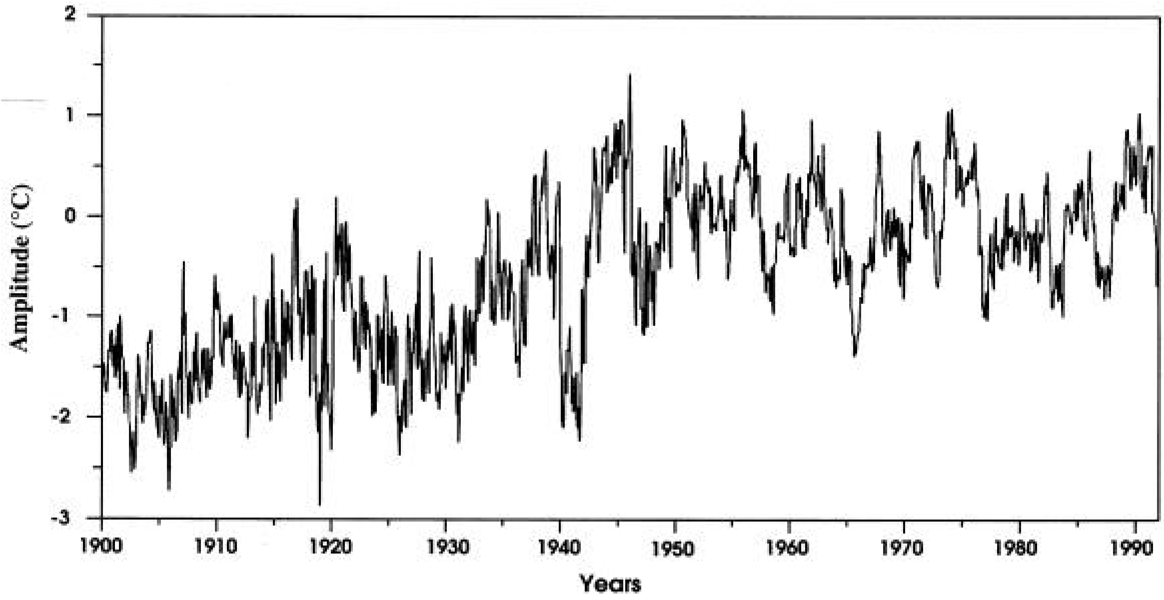


Figure 1: GOSTA global SST missing SVD analysis (rank=2). Estimated SST EOF1 amplitude.

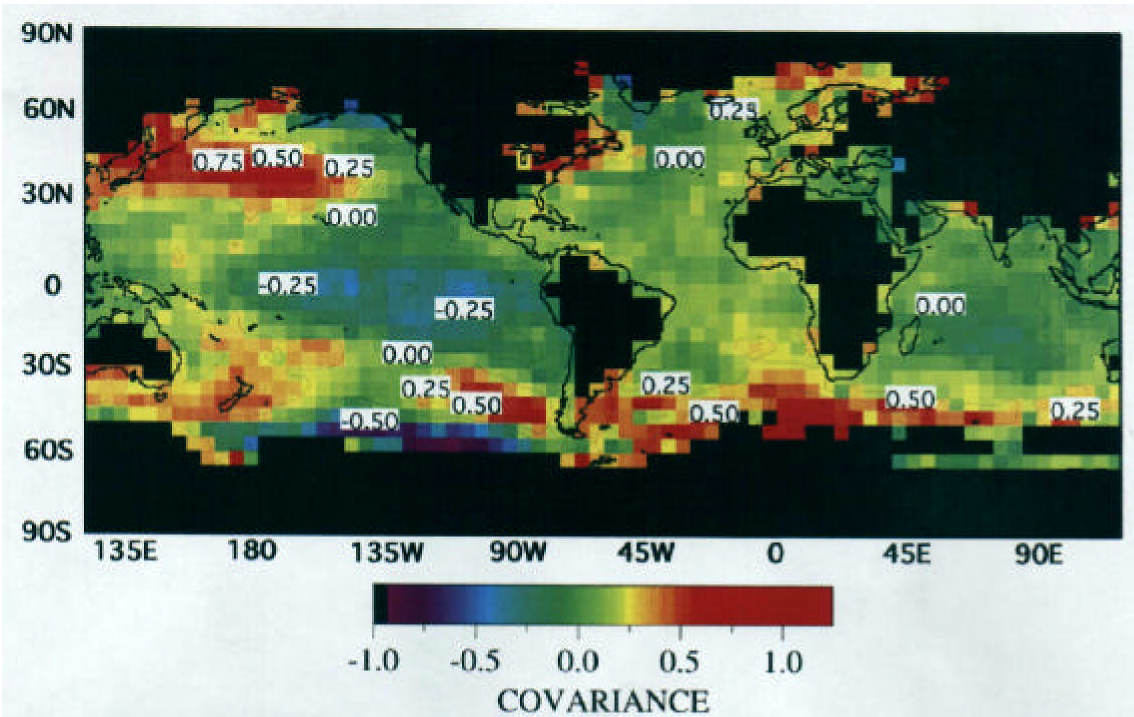


Figure 2: GOSTA global SST estimated EOF1 (10.6%, rank = 2).

With this weight function, a two-component model was estimated. At the end of the iterations, the two-component model explains a little less than 16% of the total weighted inertia of the data. Note that the norm of the gradient of the objective function has been decreased by several orders of magnitude from the initialization to the end of the algorithm.

The first estimated principal component is shown in Figure 1. This time series as presented has unit variance. The associated eigenvector is shown in Figure 2. This eigenvector has been multiplied by the square root of its associated eigenvalue. In this way, the spatial loadings depicted in Figure 2 can be interpreted as covariance coefficients between the grid's cells and the time series plotted in Figure 1.

Interdecadal changes of SST are particularly evident in this first principal component. The time series suggests a cold start of the twentieth century with a sudden warming between about 1920 and 1940.

After World War II, the time series suggests a slight cooling until 1976. After this date, a slow but regular warming took place. Indeed, this first estimated principal component is very similar to the time series of global and hemispheric temperature anomalies presented by Parker et al. (1994).

However, an important discrepancy between our time component and the estimates of Parker et al. is that recent decades are not substantially warmer than the preceding ones on Figure 1.

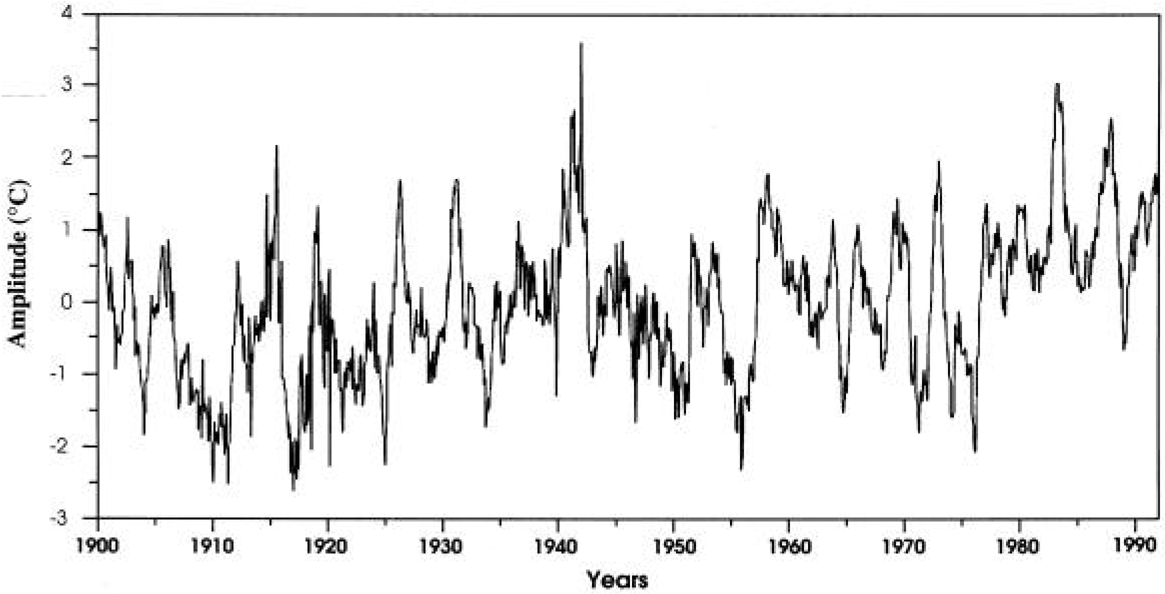


Figure 3: GOSTA global SST missing SVD analysis (rank=2). Estimated SST EOF 2 amplitude.

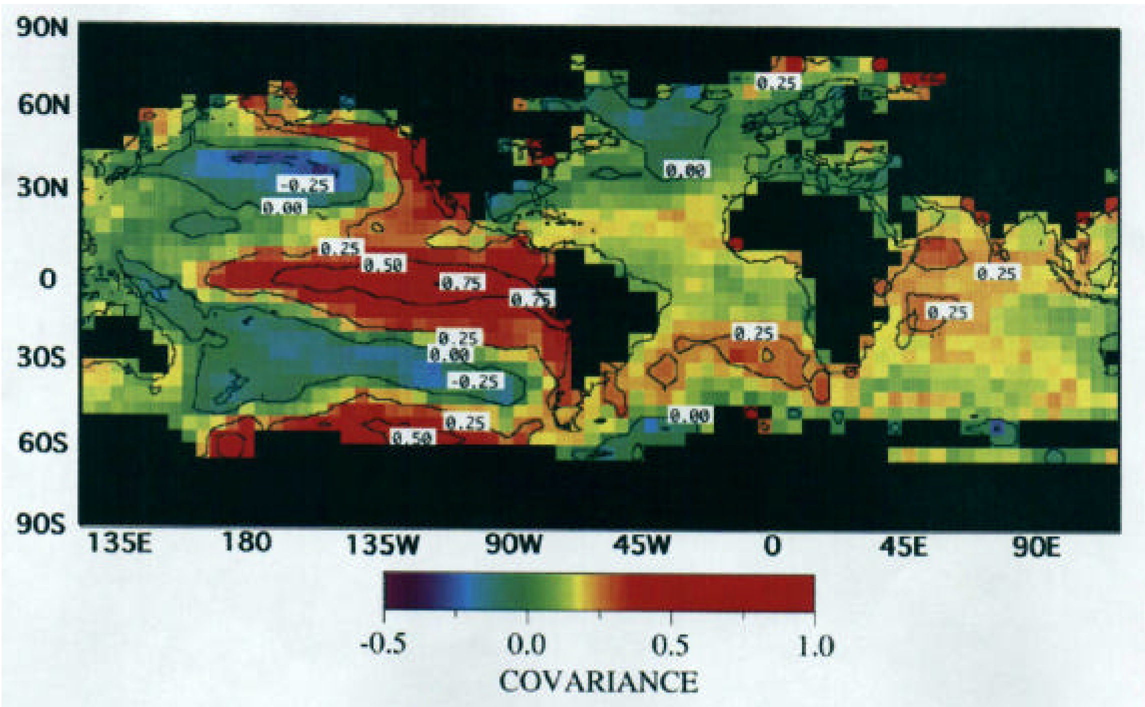


Figure 4: GOSTA global SST estimated EOF2 (6.9%, rank=2).

Note also that the first part of this time series is much more noisy than the last part; this may be due to our choice of the weight function since we gave the same weight to all data entries with non-missing values in the atlas without taking account of the number of ship-reports used to construct the anomalies.

In the same fashion, the strongly negative time coefficients during World War II are due to high and isolated positive monthly anomalies in the central and eastern Pacific which were likely computed using very few ship observations.

We hypothesize that a much better job can be done about these two problems if we use a more appropriate weight function.

The spatial pattern associated with this time series suggests that these decadal SST variations are well-marked in the midlatitude North Pacific and in parts of the middle- to high-latitude Southern Ocean (Figure 2). By contrast, the areas in the central and eastern equatorial Pacific and also in the South Indian Ocean are negatively correlated to this time series. It may be pointed that this fact is also evident in the global fields of decadal annual surface temperature anomalies presented by Parker et al. (1994).

The second estimated principal component is shown in Figure 3. A strong interannual signal seems to be present in this time series with a time-scale of about 3 to 4 years, especially in recent decades. A sudden warming may also be noticed after 1976. The estimates during World War II are again unreliable. The spatial loadings associated with this time series exhibit the well-known ENSO signature with a warm tongue in the central and eastern Pacific, and with smaller amplitudes and opposite phase in the middle latitude North and South Pacific (Figure 4).

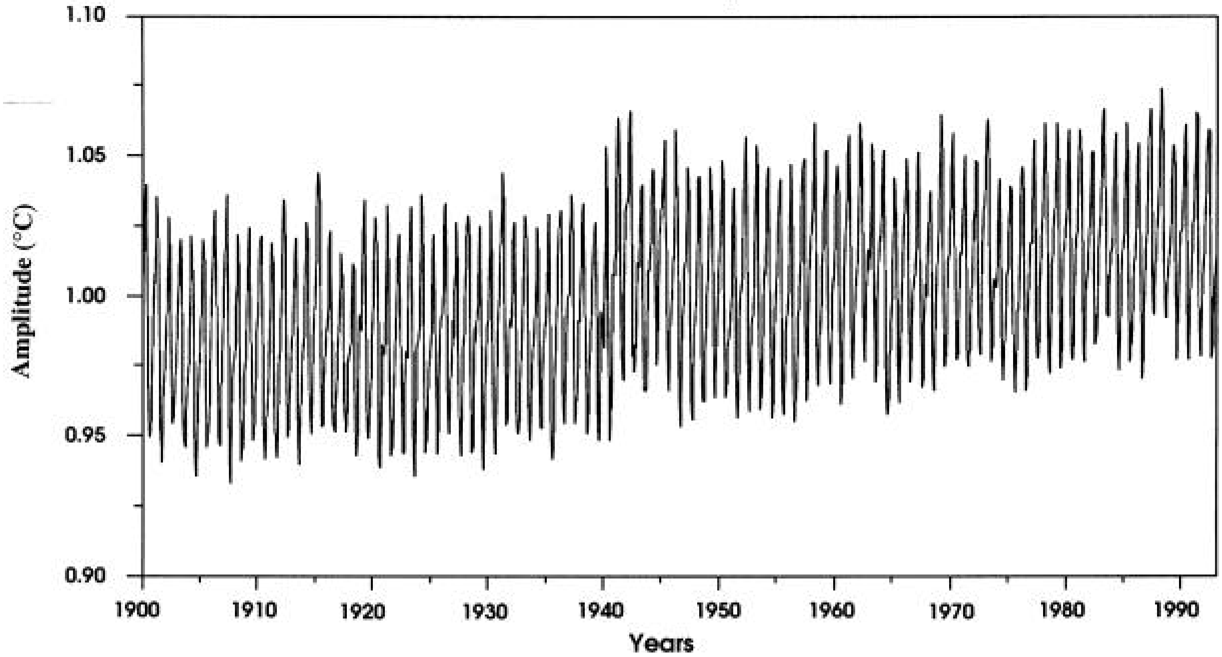


Figure 5: COADS Indian Ocean SST weighted SVD analysis (rank=2). Estimated SST EOF1 amplitude.

Some positive areas also are noticeable in the Indian Ocean. Thus, this second principal component and its associated spatial pattern suggest that recent warmings may have some connections with ENSO and a sudden change of the climate mean state which took place in the pacific regions during 1976.

Example 2

The second example is taken from the COADS trimmed monthly mean summaries (Woodruff et al. 1987). SSTs over the Indian Ocean (41°S-31°N and 29°-121°E) were extracted for the period 1900 to 1992. Note that these data are not anomalies but estimates of monthly mean SST on a 2° latitude x 2° longitude grid.

The weight matrix used in this analysis was constructed with the smooth function of the number of observations contributing to each cell's monthly mean value discussed in Section 2. Again, a two-component model was estimated from the data by the weighted EOF technique. These two components explain more than 99.8% of the total weighted inertia.

The first principal component is, to a very good approximation, sinusoidal with an annual period (Figure 5). An interdecadal trend seems also to be present in this time series with a sudden warming after 1976. The same results may be obtained by averaging the data for the whole Indian Ocean (Terray, 1994). The associated spatial pattern exhibits a north to south gradient of SST (Figure 6). Note also that SST is colder off the African coast.

The second principal component (not shown) is still marked by an annual period but its spatial pattern shows a characteristic phase difference between North and South which adds an annual modulation to the first principal component and its associated eigenvector.

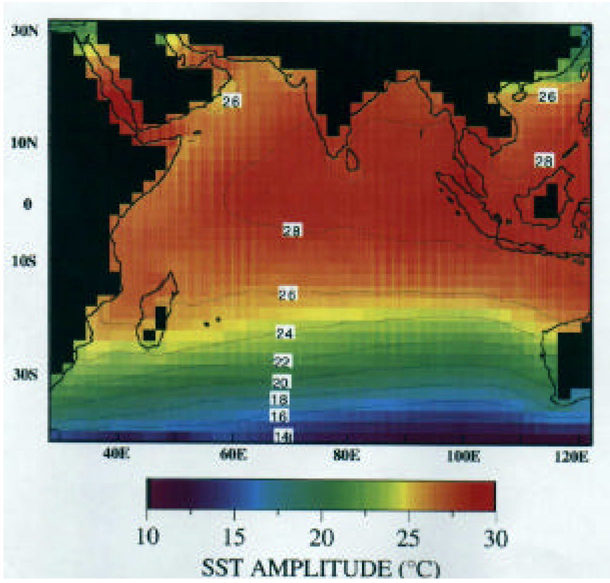


Figure 6: Indian Ocean SST estimated EOF1 (98.6%, rank=2).

In order to present in a more traditional manner the annual signal described by these two time series and their associated spatial loadings, a climatology of Indian Ocean SSTs was computed from the rank 2 weighted approximation of the data given by the two-component model. This climatology may then be compared to a traditional climatology obtained from an objective analysis in order to show the coherence of the results.

The mean SST fields for January and July obtained from the rank 2 weighted approximation of the data are shown in Figures 7 and 8.

SST patterns in the January mean field are dominated by highest temperatures (28°C) in the eastern Indian Ocean between the equator and about 15°S and also near Madagascar. Strong SST gradients are evident over the higher latitudes of the southern Indian Ocean.

The July mean SST fields show the effect of upwelling and monsoon cooling near the African coast associated with the Somali jet and to the south of Peninsular India while other parts of the North Indian Ocean are still dominated by warmer SST. All these patterns are found in classical atlases (Hastenrath and Lamb, 1979; Bottomley et al., 1990).

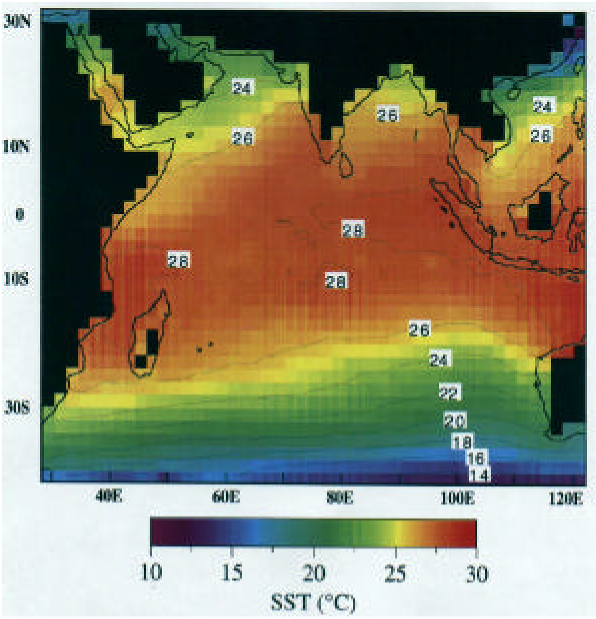


Figure 7: SST mean (COADS) January.

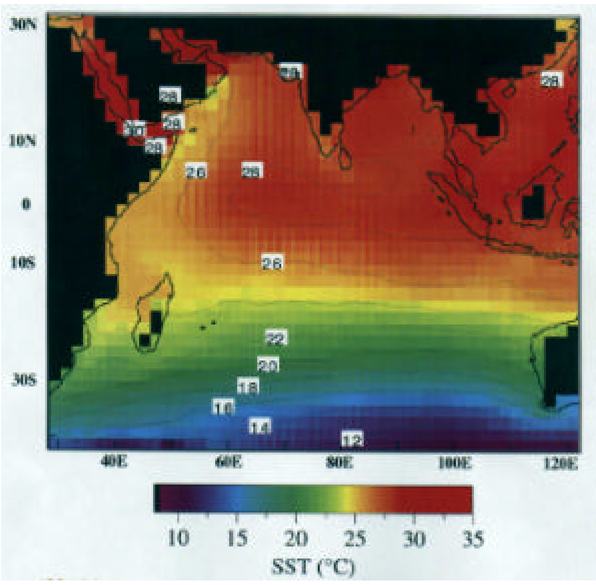


Figure 8: SST mean (COADS) July.

4. Conclusions

EOF analysis has been widely used to explore the spatial and temporal relationships within large geophysical datasets. This success can be explained by the ability of EOFs to compress the main modes of variability in the original dataset into a few time series and associated spatial patterns. Indeed, the EOF technique can be thought of mathematically as a method for approximating data matrices by matrices of lower rank since conventional EOFs provide the best approximation of a data matrix in the sense of least squares under the assumption that all the data entries have the same weight equal to one.

This paper presents an extension of conventional EOFs when weights are assigned to data entries in the original dataset. The new method which may be termed weighted EOF analysis is designed to fit a lower rank least squares approximation to a data matrix with a general choice of positive weights. If the weight matrix is carefully constructed, this new tool allows us to analyze the natural variability exhibited by data of varying reliability. It must also be emphasized that the proposed method directly takes care of missing values by assigning zero weights to such data entries.

Indeed, there are many situations in which weighted EOF analysis is more appropriate than conventional EOFs. In particular, weighted EOF analysis is shown to be a useful tool for extracting climatic signals from ship's datasets which are characterized by a strong irregular space-time sampling. In the context of ship datasets with irregular space-time sampling, weighted EOF analysis can be particularly useful for the following purposes:

- accurate and robust detection of climate signals (annual, interannual and multidecadal) on a grid-mesh, directly from the ship observations;
- blended analysis of marine and land datasets;
- interpolation of missing values;
- derivation of climatologies and smooth oceanic fields;
- sensitivity experiments (e.g. by using various weight matrices with the same dataset).

Other applications of weighted EOF analysis will be reported in detail elsewhere.

References

Bottomley, M., Folland, C.K., Hsiung, J., Newell, R.E., and Parker, D.E., 1990: *Global ocean surface temperature atlas "GOSTA"*. Joint project of the UK Meteorological Office and Massachusetts Institute of Technology, 20 pp., 313 plates, HMSO.

Hastenrath, S., and Lamb, P.J., 1979: *Climatic atlas of the Indian Ocean. Part I: Surface climate and atmospheric circulation*. The University of Wisconsin Press, Madison, 109 pp.

Kutzbach, J.E., 1967: Empirical eigenvectors of sea level pressure, surface temperature, and precipitation complexes over North America. *Journal of Applied Meteorology*, **6**, 791-802.

Parker, D.E., Jones, P.D., Folland, C.K., and Bevan, A., 1994: Interdecadal changes of surface temperature since the late nineteenth century. *Journal of Geophysical Research*, **99** (D7), 14,373-14,399.

Terray, P., 1994: An evaluation of climatological data in the Indian Ocean area. *Journal of the Meteorological Society of Japan*, **72**, 359-386.

Terray, P., 1995: Space-time structure of monsoon interannual variability. *Journal of Climate*, **8**, 2595-2619.

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

Drifting and Moored Buoy Metadata Issues

Val R. Swail
Environment Canada, Toronto, Ontario

The proliferation of moored and drifting buoys over the past two decades, and particularly in the 1990's, represents a significant improvement in the quality and quantity of surface marine observations. This increase is especially marked near the continental margins of North America, and to a lesser extent of Europe.

As with any database, the proper use of the data requires investigation of the metadata associated with the observations. When analyzing historical ship data, for example, it is essential to refer to WMO Publication No. 47, *International List of Selected, Supplementary and Auxiliary Ships* (e.g. WMO, 1973-1994) for information on the country of origin, call sign, and information about the observing program and its changes over time.

Similarly, to make proper use of the vast quantity of buoy data, either in analysis of long term climate trends and variability, or in operational forecasting, it is essential to have quick and easy access to a complete global database of metadata for all moored and drifting buoys. A good example of how such a meta-database can be used in conjunction with the actual meteorological data is given by Kent et al. (1997). In order to be most useful (and therefore used) the buoy metadata must be available in a form such as the WMO-47 meta-database for ships. This issue was addressed by the WMO/CMM Sub-Group on Marine Climatology in April 1996, and identified as an area for further work.

In terms of the present status of buoy metadata there is both good news and bad news. The good news is that there is a wealth of metadata available, particularly for the moored buoy programs. The bad news is that it is not readily usable in analysis of large climatological data sets. It is neither consistent nor complete, nor concentrated in one document such as WMO-47 into a one-line summary of all required metadata which can then be appended to the actual data record, but is spread about in a number of different locations, most often in *html* files on one or more Web sites.

This may be acceptable for operational weather forecasting applications, but becomes extremely tedious and time-consuming in climatological studies of huge amounts of data. Also, the type and extent of metadata information varies from source to source. The U.S. National Data Buoy Center has a set of very comprehensive metadata concerning its moored buoy program. The Canadian Ocean Data Acquisition System has a considerable description of its metadata, but the content is not consistent with the U.S., and they certainly aren't in compatible formats. Metadata on European buoy programs may exist, but were not easily found in this investigation, which is really the point. Similarly, metadata information on all drifting buoy programs was scarce. The following paragraphs will propose that a buoy metadata standard be established, similar to WMO-47, and provide a suggested list of contents.

Table 1 lists the information which should be included in a buoy meta-database. Most of the items are self-explanatory.

Table 1
Information proposed for inclusion in a buoy meta-database.

Country
Station name
WMO identifier
GOES number, ARGOS number, Hex number
Water depth, site elevation (not all buoys are at sea level, e.g. Great Lakes)
Mooring type (rope, chain, combination, inverse catenary)
Watch circle
GOES transmit time* and frequency
Position of mooring
Date of mooring position
Hull type, hull number
Payload type*
Digital picture
Dimensions (length, breadth)
Type of station (moored, drifting, ice)*
Degree of automation*
Communication type (GOES, ARGOS, RF)
Date of last service
Date of receipt of metadata information into the meta-database
Observing program (ship code (COADS) bulletins)*

* Additional information or explanation is provided in the text.

"GOES transmit time" refers to the time slot assigned for the message to be transmitted from the buoy, and can be anywhere in the hour. This has relevance to the precise time of certain measurements as described in later paragraphs dealing with the observing program.

"Payload type" is typically described in existing meta-databases in acronyms, such as DACT, VEEP, GSBP and ZENO. While the experienced user will know the characteristics of each of these in terms of sampling periods and frequency, averaging methods and so on, the information is cumbersome to use in processing vast amounts of buoy data. For some payloads the information on processing is not readily available. Therefore, while it is useful to have an identification of the payload type in the metadata, it is also important to include separately and specifically information pertaining to the characteristics of the payload.

The "type of station" field is self explanatory; it is highlighted here to reflect the possibility that other types of data may be included in this data set, such as the U.S. Coastal-Marine Automated Network (C-MAN) stations, or offshore drilling platforms. In the case of the latter, the field on "degree of automation" becomes relevant; if the meta-database is restricted to automated stations only, then this field is unnecessary.

The “observing program” field described in the meta-database should list all of the elements reported by the buoy. It must be considered that there are often more data available in the message sent by the buoy through GOES than appear in the ship code messages which find their way into marine data banks such as COADS. In listing the elements observed, there should be one set of indicators for data in the ship code, and a set of indicators for supplementary data which may be available in other archives such as at NDBC, or the Canadian Marine Environmental Data Service. Such additional data would include second anemometer or pressure sensor data, additional wave information such as the maximum wave, vector mean wind speed (where scalar mean is reported in the ship message), continuous wind measurements, etc.

Further detailed information is required on the observing program. Elements observed may include atmospheric pressure, pressure tendency, air and sea temperature, humidity, wind speed and direction, peak wind gust, wave height, period and direction, peak wave height, 1-D or 2-D wave spectra, subsurface temperature, ocean currents, salinity, visibility (if manned stations are included in the database) and compass heading. For all elements the information is required is shown in Table 2.

The “time of the observation relative to hour” may be important for certain applications, such as satellite sensor validation. For example, for some Canadian buoys the meteorological sample is taken immediately prior to the GOES transmission time. Therefore, if the GOES time slot is H+45 minutes, then the 10-minute wind, air and sea temperature samples will be over the period 35-45 minutes after the hour, while the pressure will be 45 minutes after the hour. For other buoys the meteorological sample is taken 5 minutes either side of the hour. Wave information is similarly non-standard in terms of its time within the hour, although it usually precedes the meteorological sample, whenever that is taken.

The “averaging method for winds” must be specified, both for mean wind speed and gust speed. Either wind speed may be reported as a vector or scalar mean; the gust speed is usually reported as the highest running mean value over the gust interval (e.g. 5 seconds) during the full wind sample (say 10 minutes) but there are other possible computation methods.

In addition to this relatively <u>static</u> meta-database there is also a more <u>dynamic</u> meta-database, which may include information on a daily or weekly basis on completeness of the data record, quality of the observations, and in some cases calibration and drift correction applied to sensors.

Due to the potentially frequent changes in this information it may be more reasonable to leave some or all of this metadata in the hands of the individual national agencies for access as necessary, with a pointer in the main meta-database to the source. In order for a meta-database system to be effective it must not create a prohibitive workload for national Port Meteorological Officers (or buoy co-ordinators) or for the World Meteorological Organization. With the present quarterly updates of WMO-47 it may be possible to include some of the information in Table 3 in the static data set; however, many of the status fields would need to be accompanied by associated date fields, thus increasing both the reporting burden and the record length.

There are a number of other miscellaneous buoy data/ metadata issues which arise in processing and analysis of buoy information.

Confusion often results from the convention of assigning buoys and offshore drilling platforms similar numbers. On the east coast of Canada, for example, buoys are numbered 44137, 44138, etc., while a platform near Sable Island is assigned 44144. Similar cases exist for the North Sea. A proper metadata system including both buoys and platforms would help alleviate this situation; ideally, however the two distinct types of station would be assigned different identifiers.

Table 2

Information required on observing program.

Instrument type (propeller/vane, cupwheel, sonic)
Manufacturer and model number
Location (distance from bow, distance from centre line, on foremast, aftmast, yardarm)
Height above (below) water line
Original units of measurement
Calibration date
Averaging period
Sampling frequency (e.g. 2 Hz for wind measurements, 1 Hz for temperature)
Time of observation relative to hour*
Averaging method (winds)*
Compass type, model number (winds)
Processing method (waves; e.g. NDBC directional, ZENO, Watchman, etc.)

* Additional information or explanation is provided in the text.

Table 3

Information proposed for a dynamic meta-database.

For each observed element a code should be specified from the following:
element observed
element not observed
data under evaluation, not reported
data unreliable, but still reported (e.g. direction occasionally bad)
sensor/system failure
sensor status unknown
Active sensor where there is more than one (winds, pressure)
Known bias in direction of anemometer
Sensor drift (temperature particularly, but possibly also wind speed)
Buoy adrift and date
Dates deployed, retrieved, adrift and serviced
Periods when buoy adrift but still reporting
Periods when buoys reporting to ARGOS not GOES

In some instances moored buoys have been relocated to different sites, often hundreds of miles away, while still using the same number. This has caused problems for unsuspecting users who may not have easy access to the metadata.

Buoy type and identification are often miscoded in the COADS database, particularly for non-U.S. buoys. Many Canadian buoy reports, especially in recent years, are coded as “SHIP” in place of the true buoy identifier, and the platform type is often incorrectly described as “platform” or “ship”.

Many moored buoys now have GPS systems on board for determining position. This information should be reported in the position information in the actual data report rather than the mooring location from a lookup table; there are numerous cases where the position assigned to individual reports has been in error due to late updates to position tables. In one known case, a buoy reported positions 120 km away for a period of 10 months. The mooring position could be used as a backup position assignment in the event of a GPS failure, similar to switching active anemometer or pressure sensors.

Timely updates of buoy metadata are very important. Information should be provided at least quarterly, as for the ship data in WMO-47. Extensive use of date fields within the metadata system is required to pinpoint when changes have occurred, particularly for the dynamic metadata variables.

The metadata system proposed above should be relatively simple to incorporate for future buoy (and possibly platform) deployments. It will require some effort, however, to compile the desired information for historical data, even where that information is generally available, such as the NDBC metadata files. For other buoy networks it may not be possible to describe the observing programs in any detail. As for WMO-47, it is important to include whatever details are possible in the historical metadata, and strive to improve for the future.

Another important issue involves getting more historical buoy data into the COADS database. For many years non-U.S. buoy data were not archived hourly in COADS due to operational decisions at the U.S. National Centers for Environmental Prediction.

While all U.S. buoy data was archived hourly, Canadian data, for example, was archived only at the main synoptic hours, with some intermediate synoptic observations. On the east coast of Canada, for several years, buoy data were archived on a very irregular, non-synoptic basis. While this problem seems to have lessened in the past year or two (subject to miscoding as described above), there remains the problem of getting the missing historical data into the COADS database.

The final issue may be the daunting obstacle. In these times of budget and staff reductions, who will take the lead role in marine climate data and metadata issues, and where will the money come from to carry out the work required? Certainly the WMO/CMM Working Group on Marine Climatology and the COADS Project are obvious leadership candidates, but the issue of funding must be addressed at both the national and international levels.

References

Kent, E.C., S.A. Josey and P.K. Taylor, 1997: Use of WMO47 Metadata in a global flux climatology (this volume).

World Meteorological Organization, 1973-1994: *International list of selected, supplementary and auxiliary ships*. WMO Publication No. 47, Geneva.

---oooOooo---

Use of WMO47 Metadata in a Global Flux Climatology

Elizabeth C. Kent, Simon A. Josey, and Peter K. Taylor
Southampton Oceanography Centre, Southampton, UK

Abstract

The SOC Flux Climatology (Josey et al., 1998) is unique in its use of Voluntary Observing Ship (VOS) metadata to improve the quality of ocean surface flux estimates obtained from merchant ship data within the Comprehensive Ocean-Atmosphere Data Set Release 1a (COADS, Woodruff et al. 1993). Corrections to the ship data combined two WMO-sponsored data sources: firstly the VOS Special Observing Project - North Atlantic (VSOP-NA) which identified errors in merchant ship weather observations and secondly the metadata collected by Port Meteorological Officers around the world and published annually by the WMO (WMO47, e.g. WMO 1994). The metadata were merged with COADS individual reports giving instrument types and heights for most VOS weather reports. These metadata allowed the identification of those reports that required correction following the recommendations of the VSOP-NA.

The resulting fluxes are now starting to be validated and the total heat flux from the SOC climatology agrees well with high quality research data from buoys where that is available.

1. Introduction

Climatological estimates of the ocean surface heat balance, climatologies, can be calculated from COADS. In past climatologies, global adjustments have been made to the flux fields to balance the global heat budget. Authors have justified these modifications as potentially compensating for the effect of ship measurement errors on the fluxes, and for uncertainties in the bulk formulae used to calculated the surface fluxes from the VOS reports.

The approach taken to balance the heat budget has been to appeal to external information to constrain the fluxes. For example, daSilva et al. (1994) used inverse analysis to simultaneously tune the COADS-based heat and fresh water fluxes to conform with oceanographic estimates of meridional heat and fresh water transports. This resulted in a 13% increase in the latent heat flux (with a compensatory increase in the precipitation to balance the fresh water transport) and a 8% decrease in the incoming solar flux. The sensible heat flux and the longwave flux were changed by smaller amounts. Even larger adjustments have been made in other studies.

Following the identification of sources of error in VOS data in the VSOP-NA project we were able to test whether the heat imbalance (about 30 Wm^{-2} excess heating of the ocean) found in VOS-derived flux climatologies is due to errors in the data. If the VSOP-NA corrections applied to the COADS data are similar in size to the global adjustments used by daSilva et al. (1994) then we might assume that the latter can be justified on the grounds of ship errors. If not, we will have to look for other methods of balancing the heat budget.

2. Correcting the VOS data

2.1 The metadata

The metadata for merchant ships in the Voluntary Observing Fleet have been published annually by the WMO since the 1950's (e.g. WMO 1994) and are available in electronic format from 1973 onwards. The metadata consist of, for each ship, the ship's name and callsign followed by a coded list of instrument types and heights. The instrument types for pressure, air and sea temperature and humidity are listed along with information about more specialized instrumentation installed on the ship. Anemometer heights are listed where the wind report is instrumental rather than a visual observation of sea state and also the height of the observing platform which we have used as a proxy for the height of air temperature and humidity measurement.

The ascii version of the metadata for the years 1973 to 1994 reformatted as part of the SOC flux climatology project can now be found on the Internet:

ftp://ftp.cdc.noaa.gov/Public/coads/metadata/wmo47

2.2 The ship corrections - VSOP-NA

The VSOP-NA (Kent et al. 1991, Kent and Taylor 1991) consisted of the detailed analysis of 2 years of meteorological reports from 46 ships selected because they reported regularly in the North Atlantic. Port Meteorological Officers gathered detailed information about the ships and the instruments carried. Photographs or plans of the ships and of the instrumentation sites were collected where possible along with information on observing practices. In addition extra fields were added to the ship's weather log to identify the conditions at the time of the observations.

The information from the logbooks, over 33 thousand records, was keyed into ASCII format by the Deutscher Wetterdienst in Hamburg. The reports were then merged with the output of a numerical weather prediction model by the UK Meteorological Office. The model was used to provide a consistent standard to allow ship reports separated in space and time to be compared.

The results of the VSOP-NA suggested that:

a) Sea surface temperature (SST) measurements made using engine intake thermometers were biased warm (Kent et al. 1993a). This correction could be applied for the log book reports which contain the method of SST measurement but for reports received by radio the method of SST measurement needs to be found from the WMO47 metadata.

b) Air temperature measurements were affected by solar radiation. The warm bias caused by the solar heating of the ship superstructure could be removed on average using a formula depending on the incoming solar radiation and the relative wind speed (Kent et al. 1993b). Both these parameters can be calculated from information in the normal weather report, no external metadata is required.

c) Humidity measurements were unaffected on average by solar heating (Kent and Taylor, 1996).

d) Humidity measurements from screens were biased high when compared with those from psychrometers, presumably due to their poorer ventilation. This bias could be removed on average by reducing the humidity from screens using an empirical formula (Kent et al. 1993a). The method of humidity measurement needs to be found from WMO47.

e) Height correction of instrument based wind speed measurements to 10 meters should be carried out to homogenize the wind data (Kent et al., 1993a). The height of the anemometer is contained in the WMO47 metadata.

2.3 Combining the data and metadata

The link between the COADS data and the WMO47 metadata is made using the ship callsign which is present in both datasets. Since a given call sign can be transferred from one ship to another, and because changes may occur in the metadata, the matching must be done on a year by year basis. For each ship meteorological report in COADS the WMO47 database was searched to find the metadata for the ship with the appropriate callsign in the particular year. About 10% extra reports were matched if reports unmatched with the correct year were checked with metadata for the following year, indicating there is sometimes a time lag between a ship being recruited to make weather reports and its details being collected for WMO47. Figure 1 shows the success rate for matching a report in COADS with the ship metadata.

The figure shows that although the number of reports in COADS has increased slightly over the period the number of reports from ships has declined. The deficit is made up from reports from fixed platforms and moored and drifting buoys. The composition of COADS is therefore very different towards the end of the 12-year period than at the beginning. This also leads to reduced spatial coverage as the data from platforms and buoys is usually restricted to fixed locations near the coast.

The matching rate increased from less than half the ship reports at the beginning of the period, largely due to the lack of callsign information in COADS to more than 80% by the end of 1992. The step increase in the match rate in 1982 is due to the inclusion of the ship callsign in WMO's format for exchange of ship logbook data at that time.

2.4 Use of the enhanced dataset: Beaufort Scale evaluation

The conversion of visual observations of sea state to wind speed data requires the use of a Beaufort equivalent scale to assign a wind speed to a Beaufort interval. Older scales, for example WMO1100 or CMMIV (WMO, 1970) used simultaneous observations of anemometer wind speed and sea state from a few selected sites or ships in

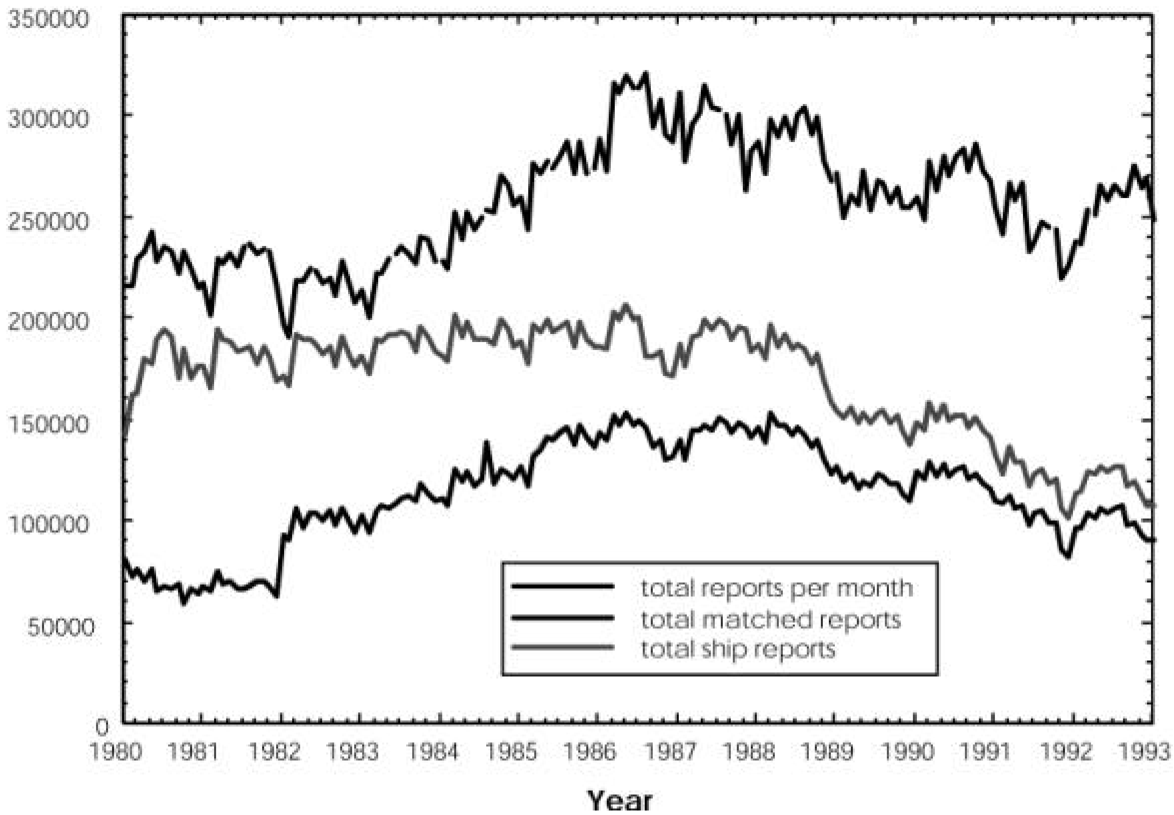


Figure 1: The number of reports per month in COADS Release 1a, top line. The number of reports from ships, centre line. The number of COADS reports for which metadata was found in WMO47, lower line.

their derivation. It is therefore not certain that these scales will work well with data from the VOS as a whole. This uncertainty led daSilva et al. (1995) to derive a Beaufort scale using data from COADS. However Lindau (1995) suggested that the uncertainty in VOS anemometer heights would cause error and compared anemometer measured Ocean Weather Ship winds of known height with VOS visual winds. The metadata-enhanced COADS has allowed us to determine the global distribution of anemometer heights.

It was already known that the Atlantic was dominated by visual wind reports (as many of the European Meteorological Services favor this method of observation) and that the Pacific was dominated by anemometer wind reports. Kent and Taylor (1997) show that the mean anemometer height in the North Pacific is greater than 30 meters over large areas; the Atlantic region reports come from lower anemometers, 20 to 25 meters on average. This is due to large Japanese ships with high anemometers dominating the reporting in the Pacific.

This large difference in anemometer height between the two ocean basins will lead to biases in the data if a uniform height correction is made to the wind data.

The merged dataset has also allowed us to use VOS anemometer winds corrected from known anemometer heights to directly compare VOS anemometer and visual winds (Kent and Taylor, 1996). The Lindau (1995) scale came out best in this comparison and is used in the SOC climatology. The daSilva et al. (1995) scale also performed well. The WMO1100 scale gave equivalent wind speeds which were too low at low wind speeds and slightly too high at high wind speeds. The CMMIV scale was found to be biased in the opposite sense, too high at low wind speeds and too low at high wind speeds.

3. Results

3.1 Effect of the VSOP-NA corrections

The effect of the corrections described in Sections 2.2 and 2.4 on the flux fields can now be determined (for a full description of the corrected climatological fields see Josey et al., 1998). As an example we shall take the latent heat flux. Figure 2 shows the effect of the VSOP-NA corrections, the individual height corrections (as opposed to a single assumed height of 25 meters for anemometer winds) and the visual scale of Lindau (1995) on the latent heat flux in January 1990.

The latent heat flux is reduced in the North Pacific, in some regions by more than 15 Wm^{-2}. This is due to the correction to SST from engine intakes. The SST is reduced in these cases, which reduces the saturation humidity at the sea surface and hence the sea-air humidity difference resulting in a decrease in the latent heat flux. An additional, but smaller, decrease in the latent heat flux in this region arises from the use of individual anemometer heights. In contrast the correction to the screen humidities decreases the air humidity, hence increasing the sea-air humidity difference and leading to an increase in the latent heat flux.

It is largely this effect which leads to the increased latent heat flux values over much of the Atlantic where most of the screen measured humidities are reported.

The effect of the solar radiation correction to the air temperature only affects the calculation of stability and the effects on the latent heat flux are therefore small. A larger effect of this correction is seen in the sensible heat flux field (not shown). The overall effect of the individual height corrections and the Lindau (1995) scale is patchy.

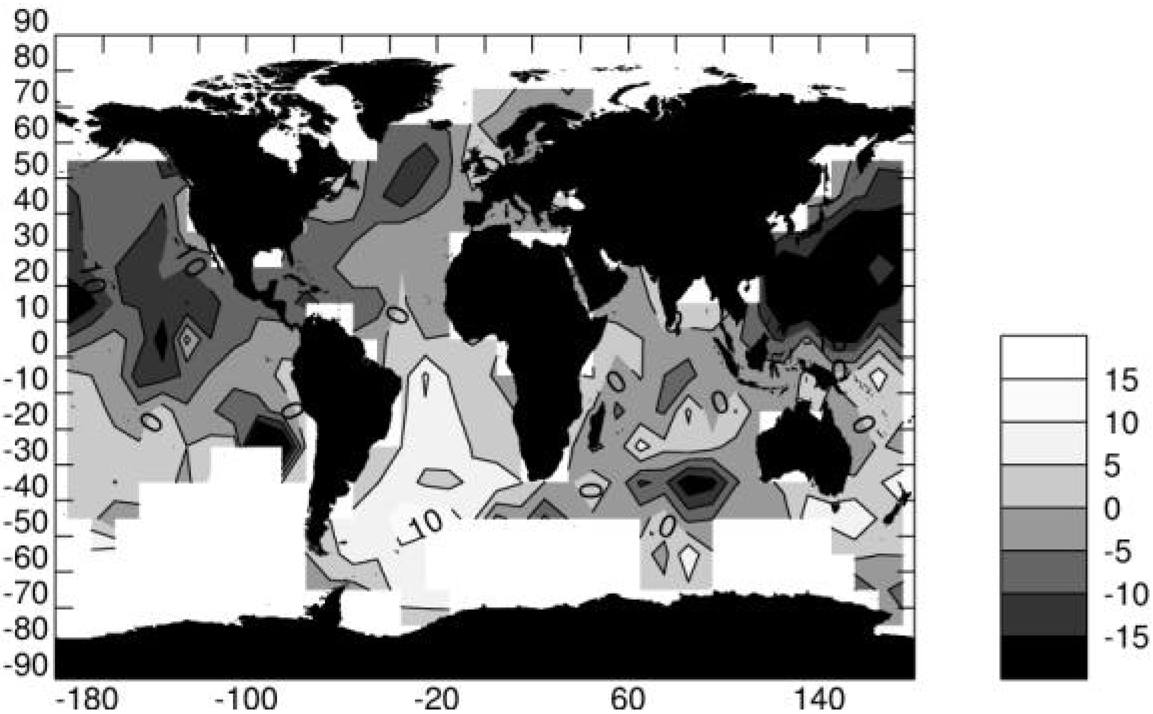


Figure 2: The effect of corrections applied to the VOS data on the latent heat flux. Differences are plotted in Wm^{-2} and a negative difference represents a decrease in the heat loss from the ocean due to corrections.

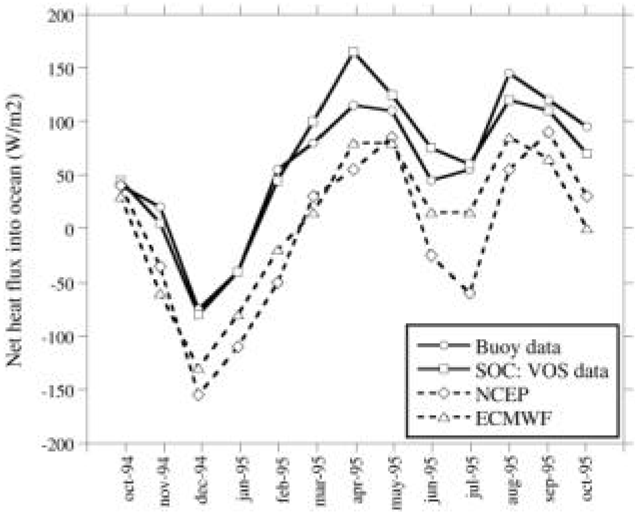


Figure 3: Comparison of monthly mean total heat flux data from a research buoy (Woods Hole IMET system; Weller et al., 1997) with monthly mean total heat flux from the SOC climatology for the same period. Also shown are total heat flux data from the numerical weather prediction models of NCEP and ECMWF which in this region agree much less well with buoy data than do the SOC data. (Adapted from Weller et al., 1998.)

The effect of the ship corrections on the latent heat flux is thus complex and regional, and has little correlation with the magnitude of the latent heat flux. This implies that a global increase of latent heat flux in order to match the fluxes with the ocean heat transport estimates cannot be justified on the grounds of errors in the ship data.

3.2 Example of heat flux validation

The surface fluxes in the climatology require validation against independent data. An example of this is shown in Figure 3 in which the total heat flux from the climatology is compared with data from a research buoy in the Arabian Sea (Weller et al., 1998).

The buoy and the climatology agree well; this is particularly pleasing as the region is not well sampled in COADS.

The agreement between the buoy and the SOC total heat flux is seen to be much better than the agreement between the buoy and two commonly used numerical weather forecast models; those of ECMWF and NCEP. In addition, the authors found that the adjusted fluxes of daSilva et al. (1994) gave a poorer representation of the heat exchange in this region; again suggesting that global adjustments are not appropriate. (Although the daSilva et al. (1994) fluxes cover a different period than the SOC fluxes, analysis of the SOC fluxes suggested that the period of the buoy deployment is climatologically typical.) Comparisons have been carried out with independent data in other regions and preliminary results from these analyses suggests that, in general, global adjustments to the fluxes will worsen the agreement.

4. Conclusions

The use of metadata with a global dataset has enabled the correction of certain errors in ship's meteorological reports. A good success rate for the matching of reports with the correct instrument information was achieved: over two thirds of the ship reports were merged with metadata. The metadata were required to apply the results of the VSOP-NA project to COADS to remove biases from the data. The corrected data were used to calculate heat fluxes from COADS and resulted in flux estimates which differed from previous estimates. For example the latent heat loss from the ocean was decreased by up to 15 Wm^2 in the North Pacific but increased in the South Atlantic.

The SOC flux fields calculated from the corrected COADS data compare well with co-incident data from research quality buoys. These results suggest that global adjustments of the flux fields are not appropriate and that the heat budget should be balanced by local adjustments to the flux fields, for example in the data sparse southern oceans (see Josey et al., 1998 for discussion). Further comparison with other data is however required.

References

daSilva, A.M., Young, C.C., and Levitus, S., 1994: *Atlas of Surface Marine Data Vol. 1: Algorithms and Procedures*, NOAA Atlas NESDIS 6, 83 pp.

daSilva, A.M., Young. C.C., and Levitus, S., 1995: Toward a Revised Beaufort Equivalent Scale. *Proceedings of the International COADS Winds Workshop*, Kiel, Germany 31 May-2 June 1994, Diaz, H.F., and Isemer, H.-J. (Eds.), NOAA Environmental Research Laboratories, Boulder, Colorado, and Institut für Meereskunde, Kiel.

Josey, S.A., Kent, E.C., and Taylor, P.K., 1998: New insights into the ocean heat budget closure problem from analysis of the SOC air-sea flux climatology. *Journal of Climate* (submitted).

Kent, E.C. and Taylor, P.K., 1991: *Ships observing marine climate: A catalogue of the Voluntary Observing Ships participating in the VSOP-NA*. WMO Marine Meteorology and Related Oceanographic Activities Report, No. 25, 123 pp.

Kent, E.C. and P. K. Taylor, 1996: Accuracy of humidity measurements on ships: Consideration of solar radiation effects. *Journal of Atmospheric and Oceanic Technology*, **13**(6), 1317-1321.

Kent, E.C. and Taylor, P.K., 1997: Choice of a Beaufort Equivalent Scale. *Journal of Atmospheric and Oceanic Technology*, **14**(2), 228-242.

Kent, E.C., Taylor, P.K., Truscott, B.S., and Hopkins, J.S., 1993a: The accuracy of voluntary observing ship's meteorological observations - Results of the VSOP-NA. *Journal of Atmospheric and Oceanic Technology*, **10**(4), 591-608.

Kent, E.C., Tiddy, R.J., and Taylor, P.K., 1993b: Correction of marine air temperature observations for solar radiation effects. *Journal of Atmospheric and Oceanic Technology*, **10**(6), 900-906.

Kent, E.C., Truscott, B.S., Hopkins, J.A., and Taylor, P.K., 1991: *The accuracy of ship's meteorological observations: Results of the VSOP-NA.* WMO Marine Meteorology and Related Oceanographic Activities Report, No. 26, 86 pp.

Lindau, R., 1995: A New Beaufort Equivalent Scale. *Proceedings of the International COADS Winds Workshop*, Kiel, Germany 31 May-2 June 1994, Diaz, H.F., and Isemer, H.-J. (Eds.), NOAA Environmental Research Laboratories, Boulder, Colorado, and Institut für Meereskunde, Kiel.

Weller, R.A., Baumgartner, M.F., Josey, S.A., Fischer, A.S., and Kindle, J., 1998: Atmospheric forcing in the Arabian Sea during 1994-1995: Observations and comparisons with climatology and models. *Deep Sea Research* (accepted).

WMO, 1970: *The Beaufort Scale of wind force (Technical and operational aspects)*, Reports on Marine Science Affairs, No. 3, WMO Geneva.

WMO, 1994: *International list of selected, supplementary and auxiliary ships*, WMO/OMM Publication No. 47, Geneva.

Woodruff, S.D., Lubker, S.J., Wolter, K., Worley, S.J., and Elms, J.D., 1993: Comprehensive Ocean-Atmosphere Data Set (COADS) Release 1a: 1980-92. *Earth System Monitor*, **4**(1), 4-8.

---oooOooo---

Incorporation of the Norwegian Ship Logbook Collection into COADS

Scott Woodruff,[1] Erik Wishman,[2] Federico Silva,[2] Sandy Lubker,[1] and Marianne Nitter[2]

(1) Climate Diagnostics Center, NOAA, USA; (2) Arkeologisk museum, Stavanger, Norway

Abstract

A Norwegian Ship Logbook Collection consisting of approximately 600 "Meteorological Diaries" (containing ~600K individual marine reports) spanning 1867-1899 was rediscovered in 1992 in the archives of the Norwegian Meteorological Institute (DNMI). The Collection was partially digitized (200K marine reports, and 2K voyage metadata records) during 1993-96 by the Archeological Museum in Stavanger (AmS), with US cooperation to help design the keying format, and to archive the digital data and metadata. As an important additional source of metadata, two complete versions of the Norwegian Instructions for Carrying out Meteorological Observations at Sea *were translated into English, and are planned for availability in digital form on the Web. Resources have not yet been located for completion of digitization at AmS or elsewhere, but the already digitized data and metadata will form a valuable new resource for enhancement of COADS and other global surface marine data archives, in support of climate and global change research.*

1. Background

The Norwegian Logbook Collection of about 600 "Meteorological Diaries" belongs to the Norwegian Meteorological Institute (DNMI), and is presently archived at the Archeological Museum in Stavanger (AmS). The Collection contains approximately 600,000 individual marine meteorological reports. It represents a Norwegian contribution under the Brussels agreement of 1853 to organize the international collection of marine meteorological observations with the help of merchant ships (Maury, 1854; Wishman, 1995). The Collection comprises two slightly different logbook formats: version 1, published during 1867-1868; and version 2, which appears to have come into use a year or two after 1868. In both versions, up to six 4-hourly marine reports were made per day (04, 08, 12, 16, 20, 00 hours local time). Each report may contain meteorological observations such as sea surface and air temperatures, wind, weather, and barometric pressure (Figure 1). Voyages were made predominantly in European and North Atlantic waters, and to a more limited extent in many other regions of the globe (Figure 2).

In 1992, the existence of the Collection was brought to the attention of the UK and US by Erik Wishman (now retired) at AmS. Spot-checks indicated that these Norwegian data likely do not already exist in digital form in either the US Comprehensive Ocean-Atmosphere Data Set (COADS *; Woodruff et al., 1997), or the UK Main Marine Data Bank (MDB). Specifically, we searched for the *Gefion*, which sailed from Hull, England across the Norwegian Sea and around the North Cape to Onega, Russia (6 reports/day for 20 May-14 June 1871 = 150 reports). In COADS, for example, a search of that time period and general area found only 18 reports, all located in the North Sea (ship names were unavailable in these reports, as is frequently the case in early ship records).

2. The digitization project at AmS

A cooperative Norwegian-US digitization project was initiated in 1992, in coordination with DNMI, with the goals to make the Norwegian Ship Logbook Collection widely available for climate and global change research as part of COADS and MDB, as well as to support historical European climate research activities at AmS. As an additional impetus for the project, analyzed historical pressure data were contributed to AmS for research purposes by the US National Center for Atmospheric Research (NCAR).

The keying format designed by the US National Oceanic and Atmospheric Administration (NOAA) was similar to that implemented by China for keying the Maury Collection of early US ship data (Elms et al., 1993; Elms, this volume). The keying format comprised two types of records: i) "header" records containing metadata about the ship and its instrumentation, extracted from the initial pages of each logbook (e.g., Figure 3), and ii) "observational" records containing the individual 4-hourly meteorological observations (e.g., Figure 1).

The two record types can be linked by means of a control number for each voyage (or leg of a longer journey). The header records also include voyage starting and ending locations, to help enable track-checking of the ship's movements in time and space. Voyage starting and ending positions could be missing, e.g., when a ship was in port; thus during digitization, missing port names and positions were taken from London Times (1990). The header records may also include other metadata when available such as the type of ship (brig, bark, schooner, corvette, steamer, etc.).

During 1993-96, a total of about 200,000 observational (data) records and 2,000 voyage header (metadata) records were keyed by AmS, and supplied to NOAA for permanent archival (Figure 4). Two editions (c. 1867-68 and 1869) of the *Instructions for Carrying out Meteorological Observations at Sea*, preface logbook versions 1 and 2, respectively.

* COADS is the result of a continuing cooperative project between the National Oceanic and Atmospheric Administration (NOAA)--specifically its Envionmental Research Laboratories (ERL), National Climatic Data Center (NCDC), and 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).

North Atlantic / Arctic Sea

Farvand Nord Atlanterhavet Ishavet, — Misvisning = 20°-19-17 15. 10° (9-8-?)

paa Reise fra Dunkerque til Arkangel

1873 Juli Maaned.	Kl.	Paaværende Plads. Bredde.	Længde.	Barometer. Therm.	Højde.	Temperatur. Luft. Th. No 1	Vand. Th. No. 2	Vind. Retning.	Styrke 0-6.	Skydække 0-klart 10-overskyet.	Regn, Sne, Slud, Taage, Hagel.	Søgang. 0-9	Anmerkninger.
	4			64	29.74	10.3	9.	SSO	2	5		3	Tåge i Syd
	8	68° 6'	10° 19'	64	29.74	10.5	9.	SSV	1	5		3	Fog to the south
d. 20	MD.	68° 13	10° 38'	64	29.74	10.2		SV	1	7	R	3	Misvisning 19°
	4			63	29.73	10.		SV	1	5		3	Magn. deviation 19°
	8	68° 32'	11° 29	63	29.72	10	9.?	SOtS	1	5		3	
	MN.			62	29.68	10		OtN	1	6	R	3	
	4			61	29.68	9.5		NtV	1	10	R	3	Heavy "rollers" (waves) from NE
	8	68° 42'	12° 14'	61	29.68	9.5	9	NNO	1	5		5	Megen Dønning fra NO
d. 21	MD.	68° 44'	12° 20	61	29.72	10.2		NV	2	5	T	5	Misvisning 17°
	4			63	29.76	10.		VNV	3	4		5	
	8	69° 16'	13° 40'	62	29.78	9.5	8.5	VNV	3	5		4	
	MN.			61	29.84	9.4		Vest	3	6		5	
	4												
	8	69° 45'	16° 55'	60	30.02	9.7	8.5	SV	4	5		5	
d. 22	MD.	70° 0 obs	17° 56	61	30.1	10.3		SV	4	6		4	Misvisning 15°
	4			62	30.14	10.		VSV	4	5		4	
	8	70° 29'	19° 14'	61	30.2	8.7	7.5	VSV	3	5		4	
	MN.												
	4			59	30.27	8.5		VtS	2	5	T	3	
	8	70° 50'	21° 30'	59	30.28	9.2	7.3	VtS	1	5	T	3	Strong current from WNW
d. 23	MD.	70° 44' obs	22° 5 obs	61	30.28	10.8		NV	0-1	3		3	Sterk Strøm fra VNV
	4			66	30.22	10.5		Ost	1	1		3	Misvisning 10° (continent)
	8	70° 56	22° 25'	65	30.2	8.9	7.5	Ost	1	4		3	
	MN.												
	4												
	8	71° 18'	23° 30	62	30.16	11°	7.5	Syd	1	1		3	
d. 24	MD.	71° 22 obs	24° 0	68	30.14	11.5		SV	1	2		3	Misvisning 9°
	4			73	30.1	11.7		VSV	1	1		3	
	8	71° 22	25° 45'	72	30.1	11.2	7.6	Vest	2	1		3	
	MN.												
	4			65	30.15	9.3		NV	1	1		2	Tågebank i Nord
	8	71° 20'	26° 50	64	30.18	10.7	7.0	0	0	1	T	2	Klart til Veirs
d. 25	MD.	71° 22 obs	27° 10' observ	65	30.16	9.5		SO	1	1		2	Misvisning 8°
	4			68	30.12	7.8		SOtO	2	2	T	3	
	8	71° 39'	28° 6'	67	30.12	8.2	7.	SOtO	1	1		3	Klart til Veirs

Figure 1: Example of a logbook page of meteorological data from version 2 (ca. 1869) of the Norwegian Collection (English translations by EW). Columns are for day, hour, location (latitude, longitude), barometer (thermometer and height), air and water temperature, wind direction and force, cloud cover, weather (rain, snow, sleet, fog, hail), sea state, and remarks. Version 1 (1867-1868) lacks columns for cloud cover and sea state, and contains a column labeled simply “weather”, but otherwise closely resembles this format.

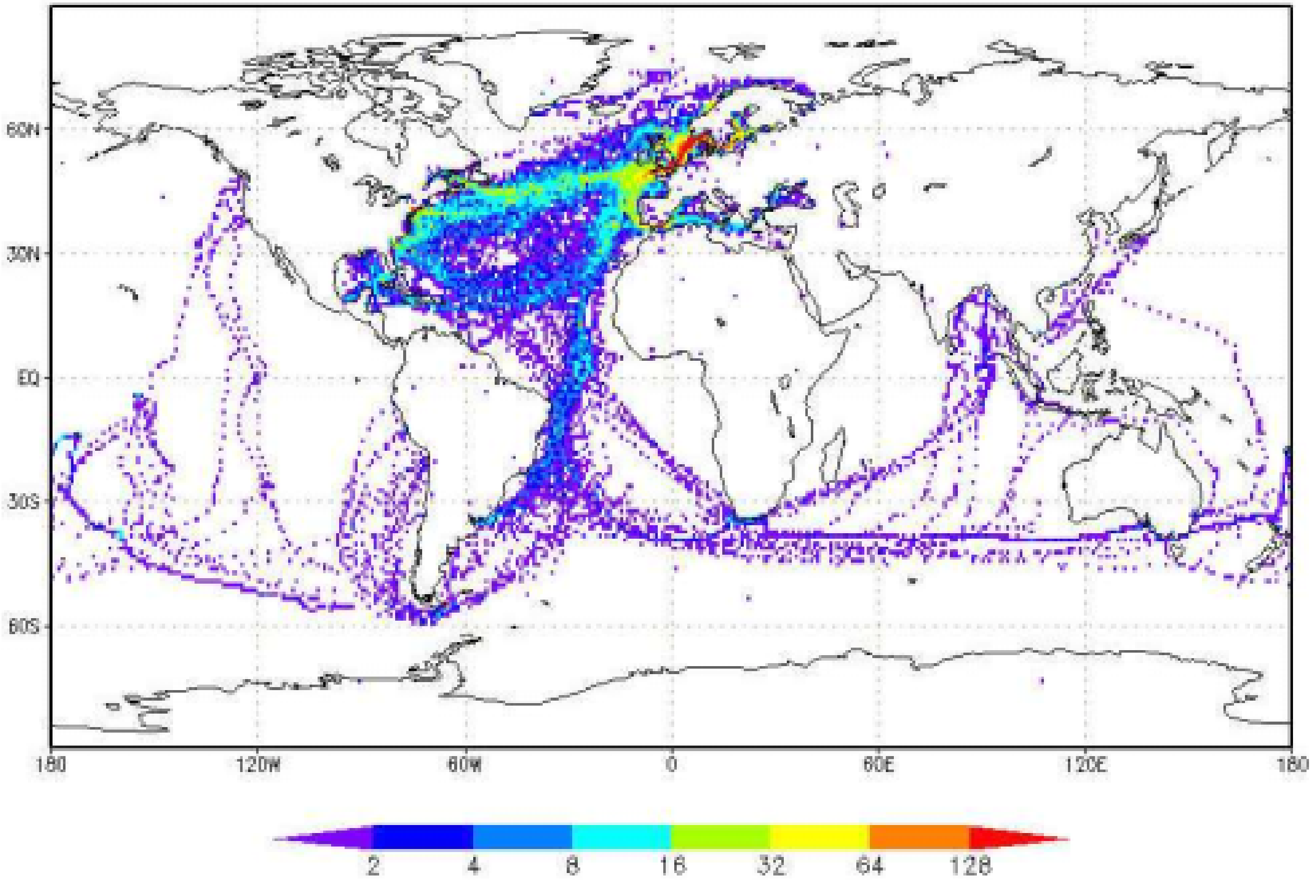


Figure 2: Global coverage of reports (~40K) containing latitude and longitude in the digitized portion (200K reports) of the Norwegian Logbook Collection (due to early navigational limitations, position typically was observed only at local noon). Reports containing observational data without location were also keyed; the missing locations will be interpolated during post-processing, with a flag to distinguish original from interpolated positions.

These were translated from Norwegian to English by Erik Wishman, and provided to NOAA in digital form. We plan eventually to make these valuable documents available via the COADS Webpage:

http:// www.cdc.noaa.gov/coads/

3. Incorporation into COADS

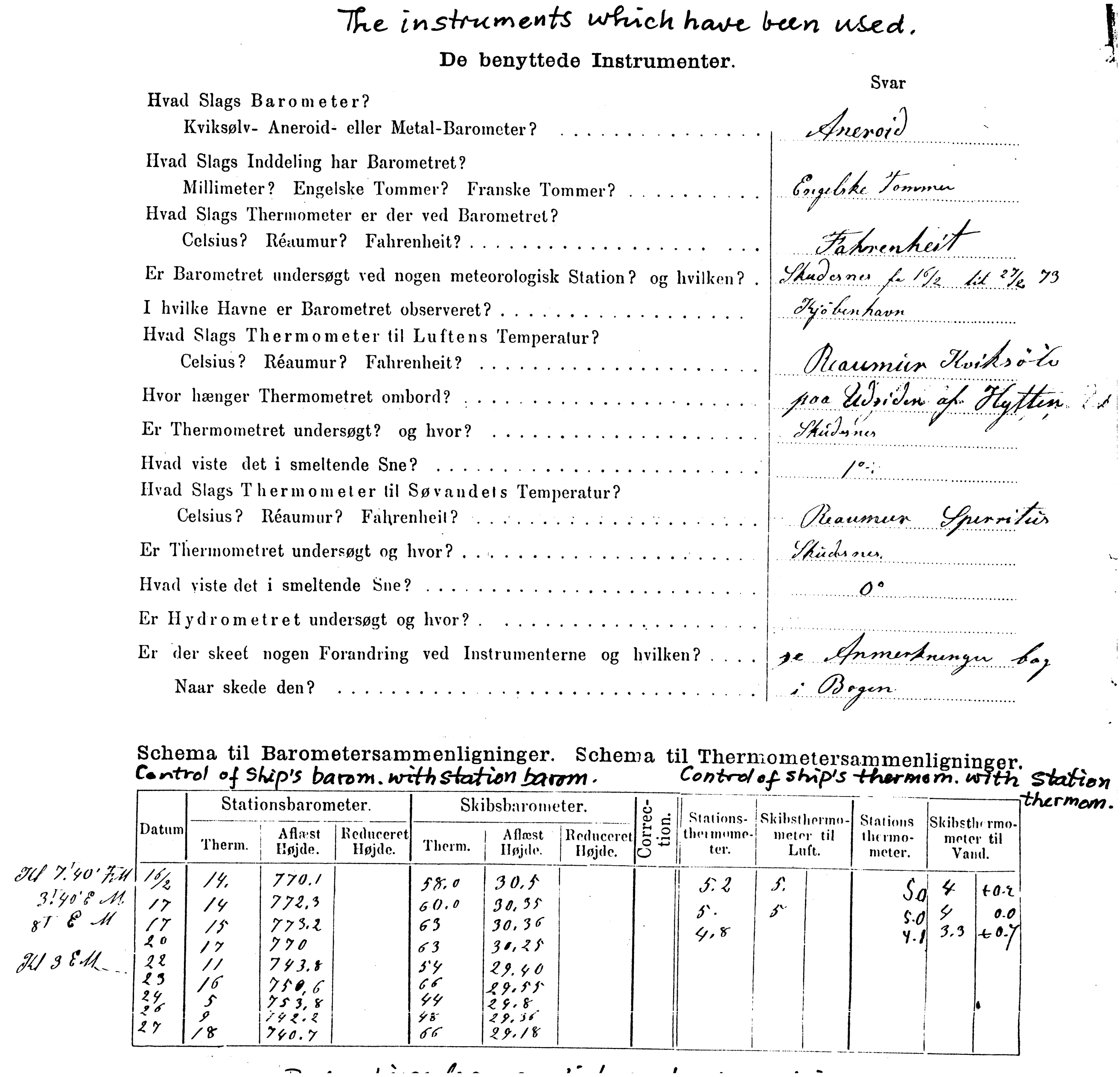
Whenever practical, the approach chosen for keying the Collection was to digitize each data element literally as it appeared in the logbook. This approach was used because we have found that conversion into modern standardized units (if even desirable) is best accomplished during "post-processing" after the original records have been keyed, since an error made in the post-processing logic can be more readily corrected, provided the original digital values are permanently archived. For example, temperatures were keyed in Celsius, Réaumur, or Fahrenheit, but an indicator was carried in the header record indicating the appropriate scale.

Incorporation of the available digital data into COADS (and later MDB) will involve a complex merger of the data and metadata, possibly in conjunction with US-UK efforts to blend COADS with MDB. Using as input the digitized observational and header records, the post-processing will implement units conversions, instrumental corrections, and other adjustments to help assure data continuity (while permanently retaining the original digitized values). Two general aspects of the post-processing are discussed below; different approaches under consideration for the more general problem of creating integrated data-metadata records throughout COADS are described by Worley (this volume).

3.1 Instrumental bias corrections

The initial pages of the logbooks included space to enter instrument characteristics (e.g., Figure 3; digitized into the voyage header records). These included the barometer type (mercurial, aneroid, or "metal-barometer") and scale (millimeters, or English or French inches), and the scales of the air and water thermometers (Celsius, Réaumur, or Fahrenheit). Space was also set aside for any corrections noted when the instruments were checked during stays in major harbors. The *Instructions* specifically indicated that observers were to enter uncorrected observations in the logbooks. The uncorrected data were digitized without change into the observational records, thus an important part of post-processing will involve application of the instrumental corrections to the observational data.

Quality control issues may arise when the instrumental corrections are missing, since likely we will be unable to determine whether the observer neglected to enter them, or the instruments did not require corrections.



The instruments which have been used.

De benyttede Instrumenter.

	Svar
Hvad Slags Barometer? Kviksølv- Aneroid- eller Metal-Barometer?	Aneroid
Hvad Slags Inddeling har Barometret? Millimeter? Engelske Tommer? Franske Tommer?	Engelske Tommer
Hvad Slags Thermometer er der ved Barometret? Celsius? Réaumur? Fahrenheit?	Fahrenheit
Er Barometret undersøgt ved nogen meteorologisk Station? og hvilken?	Skudesnes fr 16/2 til 27/2 73
I hvilke Havne er Barometret observeret?	Kjøbenhavn
Hvad Slags Thermometer til Luftens Temperatur? Celsius? Réaumur? Fahrenheit?	Reaumur Kviksølv
Hvor hænger Thermometret ombord?	paa Udsiden af Hytten
Er Thermometret undersøgt? og hvor?	Skudesnes
Hvad viste det i smeltende Sne?	1°
Hvad Slags Thermometer til Søvandets Temperatur? Celsius? Réaumur? Fahrenheit?	Reaumur Spiritus
Er Thermometret undersøgt og hvor?	Skudesnes
Hvad viste det i smeltende Sne?	0°
Er Hydrometret undersøgt og hvor?	
Er der skeet nogen Forandring ved Instrumenterne og hvilken?	se Anmerkninger bag
Naar skede den?	i Bogen

Schema til Barometersammenligninger. Control of ship's barom. with station barom.

	Datum	Stationsbarometer. Therm.	Aflæst Højde.	Reduceret Højde.	Skibsbarometer. Therm.	Aflæst Højde.	Reduceret Højde.	Correction.
Kl 7.40' F.M.	16/2	14	770.1		58.0	30.5		
3.40' E.M.	17	14	772.3		60.0	30.35		
8 E.M.	17	15	773.2		63	30.36		
	20	17	770		63	30.25		
Kl 3 E.M.	22	11	743.8		54	29.40		
	23	16	750.6		66	29.55		
	24	5	753.8		44	29.8		
	26	9	742.2		48	29.36		
	27	18	740.7		66	29.18		

Schema til Thermometersammenligninger. Control of ship's thermom. with station thermom.

Stations-thermometer.	Skibsthermometer til Luft.	Stationsthermometer.	Skibsthermometer til Vand.	
5.2	5	5.0	4	+0.2
5.	5	5.0	5	0.0
4.8		4.1	3.3	+0.7

Figure 3: Example from logbook header pages from the same logbook used for Figure 1 (English translations by EW). These areas of the header pages provide information about the instruments (types, scales, and exposures) and how well they agreed with harbor station equipment.

3.2 Other data-continuity adjustments

The adjustment of visually estimated meteorological elements (clouds, sea state, and weather) from early data units into modern standardized units represents another significant challenge for post-processing. The *Instructions* specified that wind force, for example, was to be recorded using a number from a 0-6 scale (Table 1).

In addition, when the wind was gusty, the gustiness was to be indicated as a second number after the first (steady wind force) number; e.g., 4-5 means strong wind with storm-strength gusts. As opposed to the conventional 0-12 scale, this 0-6 scale has not previously been encountered, to our knowledge, so the selection of appropriate mappings to equivalent values in meters per second represents an open question.

Table 1

0-6 wind force scale. Numbers represent one half of the Beaufort wind force scale numbers.

0	Calm	
1	Weak or light	
2	Moderate: fresh breeze; topgallant sail to top sail	Close-hauled or by the wind
3	Fresh wind; main sail to single topsail	"
4	Strong; single topsail to reefed topsail	"
5	Storm; reefed topsail or storm-sail	"
6	Hurricane; no sail can be flown	"

Figure 4: Bars (left-hand scale): Numbers of logbooks received at the Norwegian Meteorological Institute (DNMI) each year 1867-1899 (source: DNMI Yearbooks, 1867-1899). The number for 1867 is approximate, numbers for 1868-1869 were not available, and those for 1884, 1885, 1890, 1895, and 1899 may include receipts from the previous year (e.g., 1884 represents 1883/1884). Curve (right-hand scale): Numbers of digitized observational records per year (the left- and right-hand scales are aligned to reflect the assumption that each logbook contains about 1,000 observational records).

4. Future plans

At present, no economic resources are available at AmS to complete keying, although we plan to contact DNMI about possible future support for the project or explore other options to key the balance of the undigitized logbooks (436 logbooks; approximately 400K reports). Provided sufficient resources can be obtained to support the very complex conversion activities that will be required, as discussed in sec. 3, we hope to incorporate the digitized portion (about 1/3) of the Collection into COADS as part of a complete update, Release 2, scheduled for availability in the year 2000 (Woodruff et al., 1998).

5. Acknowledgments

Past funding was received from the Norwegian Research Council that helped initiate the digitization effort at AmS. The DNMI was helpful with technical and professional assistance. The NOAA portion of COADS is currently supported by the NOAA Climate and Global Change (C&GC) Program and the NOAA Environmental Services Data and Information Management (ESDIM) Program. We are grateful to J. Elms for initial design efforts on the keying format, to R. Jenne and S. Worley for NCAR's contributions in support of the project, and to U. Radok for the translation included in this paper of the half (0-6) Beaufort scale.

References

Elms, J.D., Woodruff, S.D., Worley, S.J., and Hanson, C., 1993: Digitizing Historical Records for the Comprehensive Ocean-Atmosphere Data Set (COADS). *Earth System Monitor*, **4**; No. 2, 4-10.

Maury, M.F., 1854: Maritime Conference held at Brussels for devising a uniform system of meteorological observations at sea, August and September, 1853. *Explanations and Sailing Directions to Accompany the Wind and Current Charts*, 6th Ed., Washington, D.C., pp. 54-88.

London Times, 1990: *Times Atlas of the World*, Comprehensive Edition, Times Books, London.

Wishman, E.H., 1995: Digitizing Norwegian logbooks 1867-1899: A final round to reach the objectives of the international conference of 1853 in Brussels on obtaining meteorological data from the ocean. In Pirkko Heikinheimo (Ed.): *International Conference on Past, Present and Future Climate: Proceedings of the SILMU conference held in Helsinki, Finland 22-25 August 1995*. Publications of the Academy of Finland 6/95.

Woodruff, S.D., Diaz, H.F., Elms, J.D., and Worley, S.J., 1998: COADS Release 2 Data and Metadata Enhancements for Improvements of Marine Surface Flux Fields. *Physics and Chemistry of the Earth,* **23**, 517-527.

---oooOooo---

Incorporation of Early Historical Data and Metadata into the Comprehensive Ocean-Atmosphere Data Set (COADS)

Steven J. Worley
National Center for Atmospheric Research, Boulder, Colorado, USA

1. Introduction

COADS is a collection of marine surface observations from numerous sources spanning, presently, 1854 to 1995. Incorporation of newly digitized and previously unavailable historical observations can be useful for improving the climate record in COADS. As is typical with most long-term observational archives, COADS contains biases not associated with climate signals, e.g., biases due to varying sampling practices and instrumentation changes. Metadata are critically important in the evaluation and correction of the known and yet to be discovered biases within COADS. This is true for all data sources, and is particularly important for collections made prior to the establishment of "standard" recording methods. This paper reviews our plans for enhancing the usefulness of COADS by adding historical data and metadata to the archive.

2. Historical additions for COADS

The spatial coverage in COADS is relatively sparse for years prior to 1960. Incorporation of early historical data is an important element in our plan to reduce the limits imposed by this data sparsity. By adding more historical data we are also in a better position to do cross-comparisons and data validations, which are especially important for old data collected when recording standards were non-existent or inconsistent. Permanent archival of the historical data and metadata in digital archives such as COADS also prevents their ultimate loss (e.g., through deterioration of paper records).

There are a number of datasets that are already available for inclusion into COADS. Table 1 lists several major examples. There are many other datasets that would be valuable to include in COADS, as listed in Table 2.

Table 1
Major historical datasets available for inclusion into COADS.
Under status, "digitized" means that the data have been transcribed from hard-copy form, but are not yet available in the common format used within COADS.

Collection/Archive	Observations	Date Range	Status	Comments
U.K. Marine Data Bank	9% unique to COADS	1854-recent	Digitized	Merger project planned
Maury Collection	1 million	1796-1900	Digitized	Digitized in China Available at NCDC
Russian Marine Met.	26 million	1888-1992	Ready	1950-1992 in COADS Available at NCAR
World Ocean Atlas	4.5 million	1900-1993	Ready	1950-1979 in COADS 1990-1993 in COADS World Ocean Database (WOD98) forthcoming
Norwegian Logbook	0.6 million	1867-1890	Digitized	Only 0.2 million digitized Available at NOAA/CDC
U.S. Merchant Marine	3 million	1912-1946	Digitized	Available at NCDC
Russian Makorov	3.5 thousand	1804-1891	Digitized	Available at NCAR
Arctic Ice Island		1893-1991	Digitized	Available at NCDC

Table 2
Examples of additional historical datasets proposed for eventual inclusion into COADS.

Collection/Archive	Observations	Date Range	Status	Comments
Kobe Collection	5 million	1890-1933	Under Development	1933-1960 in COADS
U.S. Marine Journal	1.8 million	1879-1893	Planned	Digitization proposed in China
Deutscher Wetterdienst	19 million	1829-1944	Under Development	
Dutch State Archives	10+ million	1812-1900 1612-1795	Undigitized	Maybe many more observations

These tables are not an exhaustive list and it is likely that other undiscovered archives are held by the world's navies and archive centers.

3. Metadata to mitigate marine parameter biases

Long-term observational data archives are likely to have parameter biases. Varying sampling procedures and methods, changes in instrumentation, heterogeneous sampling of different data sources, and inconsistent or erroneous translations of data values, are primary sources of parameter biases.

Sea surface temperature (SST) measurements made with buckets and thermometers have well-documented differences when compared to SST measurements made with sensors located in a ship's seawater intake port. Systematic differences also appear to exist between visual (Beaufort scale) estimates of sea surface wind speed and direction, versus wind measurements from ship-mounted or hand-held anemometers.

Transitions over time from mercurial to aneroid barometers, accompanied by correction and calibration problems, is another case where instrument changes can introduces parameter biases.

Merchant ships traveling along standard shipping routes, drifting or moored environmental buoys, and ships from fishing fleets all collect observations with somewhat different characteristics. For example, fishing fleet samples can be biased toward lower wind speeds relative to the wind speeds collected from merchant vessels. One of the most unfortunate and avoidable biases is where the digital data are erroneously manipulated. Examples of this are errors made in numerical truncation and rounding within computer software, and the usage of incorrect or inaccurate conversion factors when numbers are translated from one set of units to another.

Metadata can be used to mitigate biases in marine data archives. With sufficient additional information about the methods used for sampling, the instruments, the types of platforms, and the procedures used in manipulating the data, adjustments and corrections can be made such that biases will be minimized. Even with sufficient metadata, however, care must be taken, because the metadata are just like data, they are not error free.

Figure 1: (upper) Schematic depiction of three COADS marine reports in the working (archival) LMR format. Two quality control attachments ("QC" and "trimming") are currently used in COADS, plus each report may have an attachment for "supplemental" data and an "error" attachment to store the original form of fields for which translation errors were encountered. Note that the length of the supplemental attachments may vary by data source, and the second report has no error attachment. (lower) Depiction of three COADS reports in the LMRF format typically distributed to the public. Only the trimming attachment information is retained, resulting in a readily usable fixed-length record.

4. Incorporation of metadata into COADS

New technical developments are needed to more effectively incorporate additional and earlier historical data and metadata into COADS archive formats, and to make these new products readily available to the user community. In some cases, the capabilities available in the present COADS observational data formats are configured too closely after codes prescribed for reporting of modern marine data. In these cases, fields may need to be extended, or new capabilities designed, to incorporate and efficiently distribute earlier data and metadata.

For example, a field is available in the observational formats to store wind speed in meters per second (converted from other units, if necessary). A "wind speed indicator" is also available, with one setting (among several) defined to indicate that the original wind speed was reported as a Beaufort number.

Presently, however, we do not have a separately defined field to store original Beaufort codes (to enable reconversion to meters per second via alternative Beaufort equivalence scales).

Figure 2: As in Figure 1 (lower), plus a software module to integrate the LMRF with metadata and historical data, producing a user-controlled observational output.

Further complications include the use of a 0-6 (half Beaufort) scale in the Norwegian Logbook Collection, and of poorly documented text descriptions (e.g., gale, wind, breeze, trade, and airs) in the Maury Collection.

The most basic observational format used for COADS is the Long Marine Report (LMR) format. LMR is flexible and can accommodate very diverse data collections. The LMR records are variable in length. The beginning of each record is fixed-length and contains approximately 70 “standard” data fields, generally based on modern marine surface formats and code definitions. Most modern data easily translate into these standard fields. In addition, the LMR format allows for additional data and metadata (including quality control information) to be added to each record in the form of “attachments.” These attachments can be free-form in layout, and thus used to store “supplemental” data (e.g., original Beaufort codes) that do not easily translate into the standard fields. It may be necessary to define new attachments to incorporate additional historical data and metadata into COADS.

Traditionally, the LMR format has been an archival and working format used during the development phase of COADS before it is released to the public. Normally, the public is provided a fixed-length version of the LMR format, called LMRF, which contains all the standard data fields and a condensed selection of quality control data contained in the “trimming” attachment. This format is convenient for public distribution because the records are easily read on most computing systems and the data volume is about one-third less than the LMR format. Figure 1 provides a graphical depiction of the LMR archival structure, in comparison to that of the publicly distributed LMRF format.

Tentatively, to provide users with better access to the newly available historical data and metadata we plan to augment the data distributed in the LMRF format.

Since the LMRF format alone is sufficient for large temporal segments of COADS and has been distributed to the public for about five years, we do not plan to significantly change its structure. The details for this approach have not yet been fully specified, and only after more experience with the historical data and metadata will the necessary features become apparent. However, one proposed solution would be to use a user-controllable software interface to integrate the LMRF data with associated metadata and portions of the historical data that do not fit well into LMRF, as shown schematically in Figure 2.

There are advantages and disadvantages in this approach. One disadvantage would be significant added complexity in the read-access software. However, in simple applications where users do not need complex metadata or historical data, they could be be provided with the established LMRF and relatively simple access software as is now the case. Conversely, for studies that choose to use metadata and historical data, it should be possible to format the desired fields into a unified output dataset tailored to specific user interests. That way studies of data biases, and attempts at corrections, could be systematically approached through the software interface. Also cooperating research groups could take advantage of using the data and metadata as a common starting point.

---oooOooo---

ANNEX

List of Workshop Participants

ARGENTINA
Dr. Maria del Rosario Prieto
Instituto de Nivologia, Glaciologia y
Ciencias Ambientales (INAGLIA)
Avda. Adrian Ruiz Leal s/n
Parque General San Martin
5500 Mendoza Argentina
e-mail: charopri@planet.losandes.com.ar

CANADA
Val R. Swail
Climate Research Branch
Atmospheric Environment Service
4905 Dufferin Street
Downsview, Ontario M3H 5T4 Canada
e-mail: Val.Swail@ec.gc.ca
Phone: (416) 739-4347
Fax: (416) 739-5700

CHINA
Guo Fengyi
NMDIS/National Oceanographic Data Center
93 Liuwei Road, Hedong District,
Tianjin 300171 China
e-mail: webmaster@netra.nmdis.go.cn
Fax: 8622 24304408

FRANCE
Pascal Terray
Laboratoire d'Oceanographie Dynamique et de
Climatologie, tour 26
4 place Jussieu
75252 Paris, France
e-mail: terray@ccr.jussieu.fr

GERMANY
Dr. Volker Wagner
Deutscher Wetterdienst
Postfach 30 11 90
D-20304 Hamburg Germany
e-mail: vwagner@dwd.d400.de
Fax: 49-40-3190 8942

JAPAN
Ms. Teruko Manabe
Maritime Meteorological Division
Climate and Marine Department
Japan Meteorological Agency
1-3-4 Otemachi, Chiyoda-ku
Tokyo 100-8122 Japan
e-mail: teruko.manabe@met.kishou.go.jp
Phone: +81 3 3212 8341 Ext. 5146
Fax: +81 3 3211 6908

NETHERLANDS
Günther P. Können
Royal Netherlands Meteorological Institute
(KNMI)
PO Box 201, NL-3730
AE de Bilt Netherlands
e-mail: konnen@knmi.nl
Phone: +31 30 2206451
Fax: +31 30 2210407

RUSSIA
Dr. Yuri V. Gemish
Russian Federal Service for Hydrometeorology
and Environmental Monitoring
ARRI of Hydromet. Information—WDC-B
6, Korolev St., Obninsk, Kaluga Reg.
249020 Russia
e-mail: gemish@wdc.meteo.ru
Phone: +7 095 08439 74602
Fax: +7 095 255 2225

SPAIN
Dr. Ricardo García
Departamento de Astrofísica y Ciencias de la
Atmósfera
Facultad de CC. Fisicas de la U.C.M.
Ciduad Universitaria
28040 Madrid Spain
e-mail: rgarcia@6000aire.fis.ucm.es
Phone: +34 1 394 44 90
Fax: +34 1 394 51 95

Mr. Luis Gimeno
Instituto Nacional de Meteorología
Cno. de las Morenas, s/n
Ciudad Universitaria
28040 Madrid Spain

Emiliano Hernández
Departamento de Astrofísica y Ciencias de la
Atmósfera
Facultad de CC. Físicas de la U.C.M.
Ciduad Universitaria
28040 Madrid Spain

Dr. Antonio Labajo
Instituto Nacional de Meteorología
Cno. de las Morenas, s/n
Ciudad Universitaria
28040 Madrid Spain

Antonio López
Archivo General de Indias
Sevilla Spain

Dr. Javier Martin-Vide
Department of Physical Geography
University of Barcelona
08028 Barcelona Spain

UKRAINE

Alexander Polonsky
Marine Hydrophysical Institute
2 Kapitanska Str.
Sevastopol 335000 Ukraine
e-mail: ocean@mhi2.sebastopol.ua

UNITED KINGDOM

Elizabeth C. Kent
Room 254/31 James Rennell Division
Southampton Oceanography Centre
European Way
Southampton, SO14 3ZH United Kingdom
e-mail: Elizabeth.C.Kent@soc.soton.ac.uk
Phone: +44-1703-596409
Fax: +44-1703-596400

Chris Millard
Reader Information Services
Public Record Office
Ruskin Avenue
Kew, Richmond, Surrey, TW9 4DU UK
e-mail: chris.millard@PRO.GOV.UK
Phone: +44-181-392-3444 x2120
Fax: +44-181-392-5286

David E. Parker
Room H001
Hadley Centre for Climate Prediction and
Research, Meteorological Office
London Road, Bracknell
Berkshire RG12 2SY UK
e-mail: deparker@meto.gov.uk
Phone: +44-1344-856649
Fax: +44-1344-854898

UNITED STATES

Dr. Michael Chenoweth
6816 Ducketts Lane
Elkridge, MD 21227 USA
e-mail: mlcheno@smart.net
Phone: 1410 379-1508

Dr. Henry F. Diaz
NOAA/ERL Climate Diagnostics Center (R/E/CD1)
325 Broadway
Boulder, CO 80303 USA
e-mail: hfd@cdc.noaa.gov
Phone: +1303 497-6649
Fax: +1303 497-7013

Mr. Joe D. Elms
National Climatic Data Center
151 Patton Avenue
Asheville, NC 28801-5001 USA
e-mail: jelms@ncdc.noaa.gov
Phone: +1828-271-4436
Fax: +1828-271-4328

Rolando Garcia
NCAR
P.O. Box 3000
Boulder, CO 80307-3000 USA
e-mail: rgarcia@ncar.ucar.edu
Phone: +1303 497-1446
Fax: +1303 497-1400

Robert D. Gelfeld
National Oceanographic Data Center
1315 East West Highway, 4th Floor
Silver Spring, MD 20910-3282 USA
e-mail: rgelfeld@nodc.noaa.gov
Phone: +1301 713-3295 ext. 179
Fax: +1301 713-3303

Mr. Scott D. Woodruff
NOAA/ERL Climate Diagnostics Center (R/E/CD)
325 Broadway
Boulder, CO 80303 USA
e-mail: sdw@cdc.noaa.gov
Phone: +1303-497-6747
Fax: +1303-497-7013

Mr. Steven J. Worley
Data Support Section
National Center for Atmospheric Research
P.O. Box 3000
Boulder, CO 80307-3000 USA
e-mail: worley@ncar.ucar.edu
Phone: +1303-497-1248
Fax: +1303-497-1298

WORLD METEOROLOGICAL ORGANIZATION

Mr. Fernando Guzman
Ocean Affairs Division
World Weather Watch Departrment
Case Postale No. 2300
CH-1211 Geneva 20 Switzerland
e-mail: fguzman@www.wmo.ch
Phone: +41 22 730.84.49
Fax: +41 22 730.80.21