DESCRIPTION OF THE CLIWOC DATABASE

G. P. KÖNNEN and F. B. KOEK

Royal Netherlands Meteorological Institute (KNMI), PO Box 201, 3730AE De Bilt, The Netherlands
E-mail: frits.koek@knmi.nl

Abstract. We developed a user-friendly database with the 1750–1854 CLIWOC data, which is suitable to be integrated with the ICOADS database. The meteorological content focuses on wind direction and wind speed. The data, stored in the IMMA format, are accessible in numerical and in their original descriptive forms. Apart from alphanumerical meteorological information, the database contains nautical information relevant to historians, and provides access to a considerable number of images of logbook pages. The construction of the database involved a number of difficulties, including language, unit conversion, terminology and zero meridian problems. We believe that this publicly accessible database can give an important contribution to the understanding of low-frequency climate variability, as it extends the current climatological ocean databases by more than a century and probes deep into the pre-industrial era.

1. Introduction

Standardized instrumental meteorological ship observations start only after 1850; the ICOADS (International Comprehensive Ocean-Atmosphere Data Set) world database, which originally contained data back to 1854 (Woodruff et al., 1987; Wallbrink et al., 2003; Worley et al., 2005), was recently extended back to the late 18th century by the incorporation of observations from the US Maury collection (Woodruff et al., this volume). For the study of low-frequency climate variability this is still rather late. Although quantitative pre-1800 instrumental observations are few, ship logbooks at the time contain detailed reports of wind direction and wind force. Despite the fact that visual wind observations are usually referred to as non-instrumental (see, for a better word, García-Herrera et al., this volume), their quantitative character is often larger than generally believed. Pre-1854 wind observations over the oceans, perhaps in combination with a few localized land surface pressure data, enable reconstruction of the large-scale atmospheric circulation or even pressure patterns with a greater accuracy than is generally thought. The European-Union (EU) sponsored CLIWOC (Climatological Database for the World's Oceans) project aims to collect, digitize, and analyze climatological data from logbooks from the open oceans 1750–1854 and to make the database available to the scientific community (García-Herrera et al., this volume).

The CLIWOC database is not a dedicated one but general, like ICOADS. This implies exhaustive digitization of the meteorological observations. Hundreds of ship logbooks from 1750–1854 originating from Spain, England and The Netherlands, as well as several logbooks from France and a few from other countries (Sweden,

Climatic Change (2005) 73: 117–130
DOI: 10.1007/s10584-005-6946-4

USA, Denmark, Germany), were collected and digitized. The union of the shipping routes of these countries covers the North and South Atlantic as well as the Indian Ocean (see García-Herrera et al., this volume, Figure 1). In this article, an outline is given of the contents and structure of the CLIWOC database.

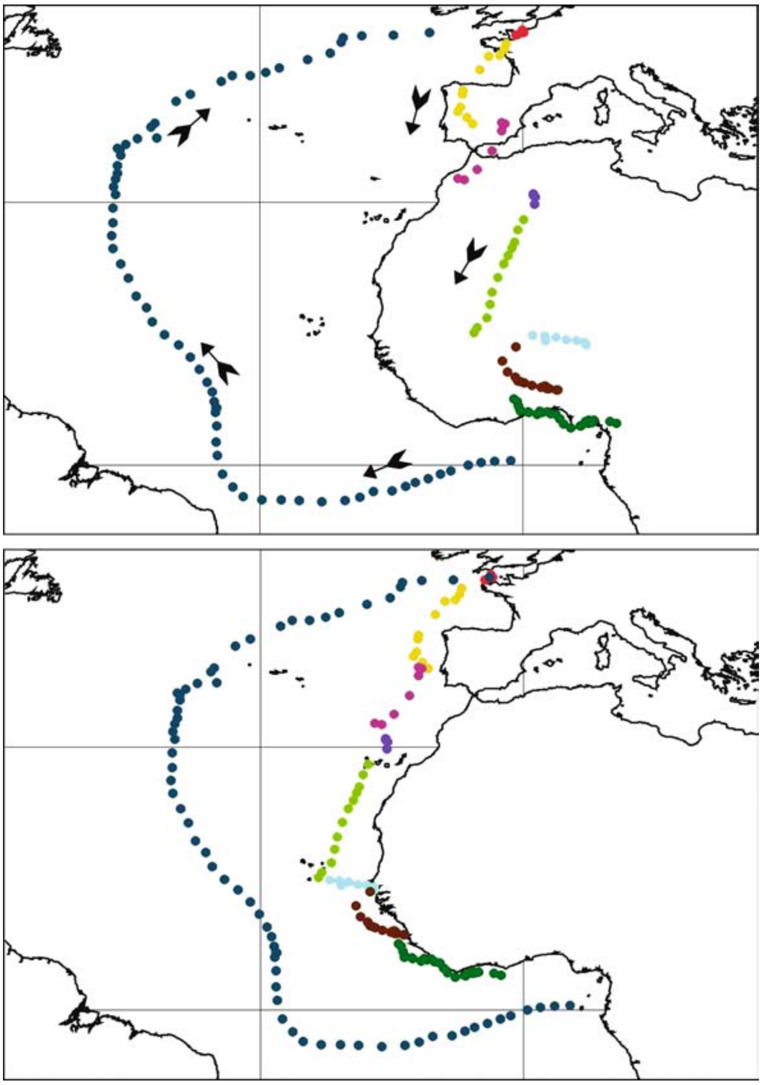


Figure 1. *Top:* the positions of HMS Surprise (1750–1751) on a round trip from England to St. Thomas (Gulf of Guinea) without correcting the longitude to the current standard, i.e. Greenwich. Every colour refers to the use of another zero meridian: Start Point (which is the name of the land tongue in SE England at 50°13′N, 3°38′W), Ushant, Cape Roxent, Madeira, Point Negro, Isle of May Bay, Cape St. Maries, Bananas and (at the start of the trip back) St. Thomas. During the voyage back no transition in zero meridian occurred by absence of land sightings. *Bottom:* the positions after converting the longitudes to Greenwich.

2. Data

2.1. EXTRACTION

Data extraction from the Spanish, English and Dutch logbooks took place in the three participating European countries, respectively. The French logbooks were mostly digitized by the Argentines; logbooks from the remaining countries by the Dutch.

All meteorologically relevant data of the midday (noon) observation were extracted, i.e. the date, geographical position, wind direction, wind force, present weather, sea state and, when available, sea ice reports, air temperature and air pressure measurements. Logbooks contain usually more than one observation per day, but the extraction of sub-daily observations was abandoned, as their information content is less than that of the noon observations (see also García-Herrera et al., this volume). Metadata, including the place where the logbook is stored, the logbook identification, the ship's name and type, the names of the original logbook keepers (the actual writers), reports of encounters at sea and, for users outside the climatological fields, recordings of notable events on board were all inserted into the database. Thus, apart from meteorological information, the database also contains miscellaneous facts like occurrences of deaths on board, punishments, illness and sightings of birds or whales. Additionally, the database allows access to 13,474 digital images of logbook pages. The access is restricted to the Dutch images only, as images from the other countries are not available. The number of accessible images comprises about 30% of the Dutch logbooks.

In total 1674 logbooks were digitized, comprising 4942 voyages and 280,195 reports. Table I shows the distribution by country. Note that the number of voyages per logbook varies widely among the countries.

TABLE I
Number of CLIWOC logbooks, voyages and observations by country

	Logbooks	Voyages[a]	Observations[b]
Spain	452	773	54,082
England	585	1,866	88,473
Netherlands	611	2,062	126,300
France	20	233	10,632
Other	6	8	708
TOTAL	1,674	4,942	280,195

[a]A round trip counted as two voyages.
[b]An observation is a note in a logbook that contains time, ship's position, and at least one meteorological element.

2.2. LONGITUDE CONVENTIONS

After keying the logbook data and initial quality controls by each CLIWOC partner, the records were sent to the Dutch CLIWOC partner. Calendar checks were performed on the English logbooks from 1750 to 1752, because for that country the Gregorian calendar was not uniformly adopted at the time. For every ship, the route was plotted, checked and adjusted according to the present-day prime meridian convention (i.e. Greenwich). In total, 646 different zero meridians were identified among all voyages. A reason for the vast number is the habit in those days to reset the zero meridian at each major landfall. In about half of the cases, the zero meridians were documented explicitly in the logbooks and the adjustments are trivial. In the remaining cases the transitions became often apparent only after plotting the ships' positions. The latitude of the break in the ship's longitude usually provides enough information to deduce the offset of the new zero meridian with respect to Greenwich.

Despite the adjustments of the zero meridians, some routes still appeared over land due to the accumulation of errors in dead reckoning. To account for that, we applied incremental adjustments to the ship's longitudes to derive the most likely track, according to the method originally developed by Jackson et al. (2000). Figure 1 shows an example of the effect of longitude adjustments on the track of a ship's voyage. Note that in this example the outward journey contained more land sightings and hence many more changes in longitude zero points than the trip back, which went over the open ocean. The figure is typical for CLIWOC in the sense that more than 50% of the ship tracks needed a major revision.

During the time span of the CLIWOC project, not all positions could be reconstructed with sufficient precision. About 10% of the 280,195 extracted positions need an advanced evaluation. Work on that is in progress. These incomplete reports are available from the database.

2.3. QUALITY CHECKS OF WIND DATA

The potential of the CLIWOC database depends critically on the accuracy and reliability of its contents. Quality checking concentrated on wind direction and wind force, being the only elements with quantitative meteorological information throughout the CLIWOC period. Attempts to fine-tune the standardization and quality checks of air pressure and temperatures are left to a later stage, although these data remain accessible in the CLIWOC database.

Wheeler (this volume) describes in detail the quality checks applied to the CLIWOC wind data. These include, among other things, consistency checks between vessels within convoys. Simultaneous changes in observing practices between two or more countries, as well as systematic differences in these practices between countries, remain largely undetected by most quality checks. This problem is particularly relevant to wind direction reports.

Wind direction recordings are usually based on readings from a compass, of which the use was widespread in the CLIWOC period. Navigators were quite aware of the angle between the direction of true and magnetic north, called the 'magnetic variation'. The magnetic variation may easily be more than 10°, sometimes exceeding 30°. The question whether the reported wind directions were related to true north or magnetic north could not be answered adequately or consistently from literature studies.

A comparison of the average wind directions of the four nations in an area of relative constant winds (the trade wind region in the North Atlantic) with the present-day climatic values indicates that the English used magnetic wind directions throughout the CLIWOC period. The Spanish data shows a preference for wind directions relative to true north, while the Dutch seem to have made a switch from magnetic to true directions somewhere in the period 1790–1810. However, to our opinion, these results of this analysis are not sufficiently conclusive to justify the implementation of country-dependent conversion procedure in the database. In the present release, all wind directions were assumed to be with respect to the magnetic north and the conversion to true north has been made throughout. The design of the database allows for an easy implementation of another scheme in the future.

3. Geographical and Temporal Coverage of CLIWOC

Figures 2 and 3 show the geographical coverage by country of CLIWOC over its entire 1750–1854 period; the all-country version is the Figure 1 in García-Herrera et al. (this volume). Table II shows the number of reports by ocean, country and 50-year period in the CLIWOC database. The border between the Atlantic Ocean and Indian Ocean is put at 20°E (near Cape Town), the border between the Indian Ocean and Pacific Ocean at 120°E (a meridian that runs close to Manila), and the border between the Pacific Ocean and Atlantic Ocean is defined by the American continent and the 70°W meridian (Drake Passage).

The table and figures show the following three features. First, the Atlantic and Indian Oceans are well covered with data. The number of reports in the North Atlantic is roughly twice of what was digitized in either the South Atlantic or Indian Ocean. The Pacific Ocean coverage is poor. Second, the English and Dutch ships ply both the Atlantic and Indian Oceans, both in the 18th and the 19th century. Third, the Spanish and French ship tracks are predominantly in the 18th century, enhancing in particular the data density over the Atlantic and Pacific Oceans.

Figures 4 and 5 show the number of reports per year in CLIWOC by ocean and country; the all-ocean version is Figure 2 in García-Herrera et al. (this volume). Figure 4 shows that, over the Atlantic, the Dutch/UK contributions summed together produce a more or less even density through time; the Spanish and French data result in a peak in the 18th century coverage. The figure also shows that after around 1830, the main input becomes Dutch. We note, however, that this effect is just a result of

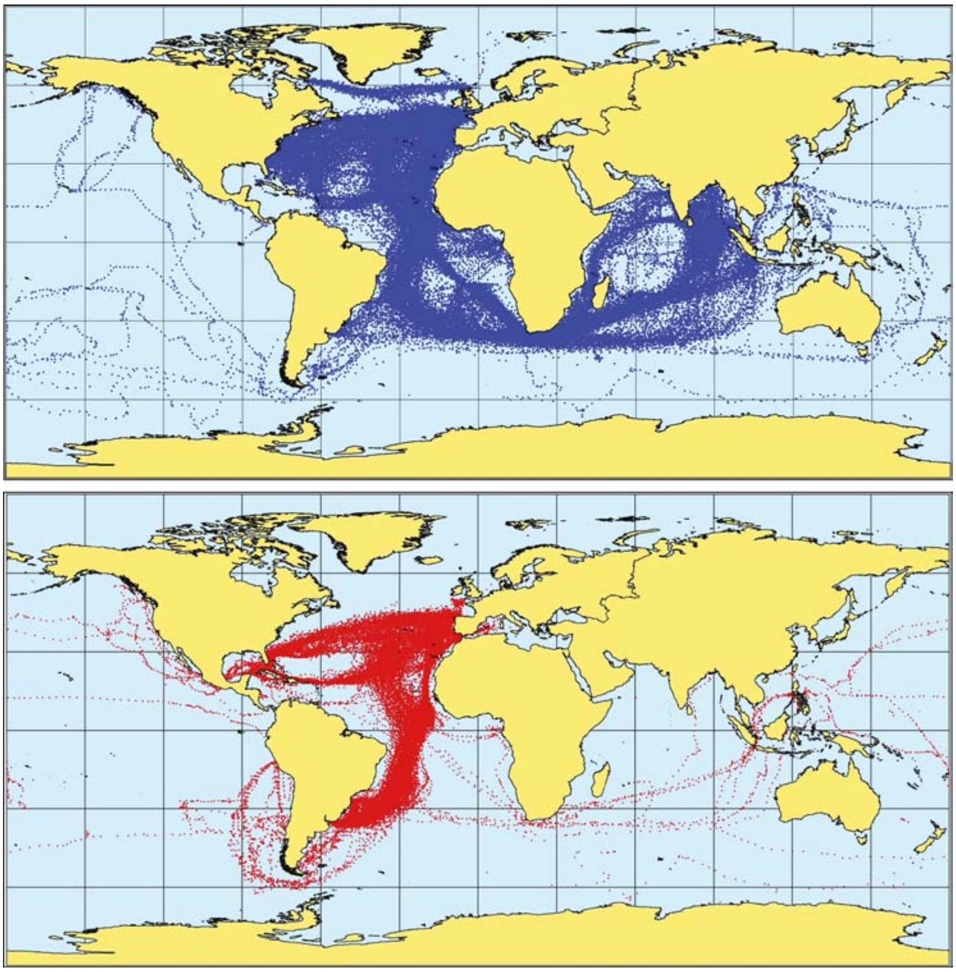


Figure 2. All ships' positions in CLIWOC 1750–1854. *Upper panel:* English ships; *lower panel:* Spanish ships.

the selection of the UK data in the CLIWOC project: many observations in the UK archives are waiting to be digitized (García-Herrera et al., this volume). Over the Indian Ocean, the data come primarily from the Dutch and UK. Table III shows the seasonal distribution of the reports per ocean. The numbers concerning the North Atlantic and Indian Oceans, show why imprints from the North Atlantic Oscillation (NAO) may be more distinctly apparent in the CLIWOC database than imprints from the El Niño Southern Oscillation (ENSO). Scientific analyses of CLIWOC are discussed in Jones and Salmon (this volume).

Table IV shows the number of wind, pressure and temperature observations per ocean for the period 1800–1854. Pre-1800 numbers are not included in the table, as there were hardly any instrumental observations at the time (see Figures 3 and

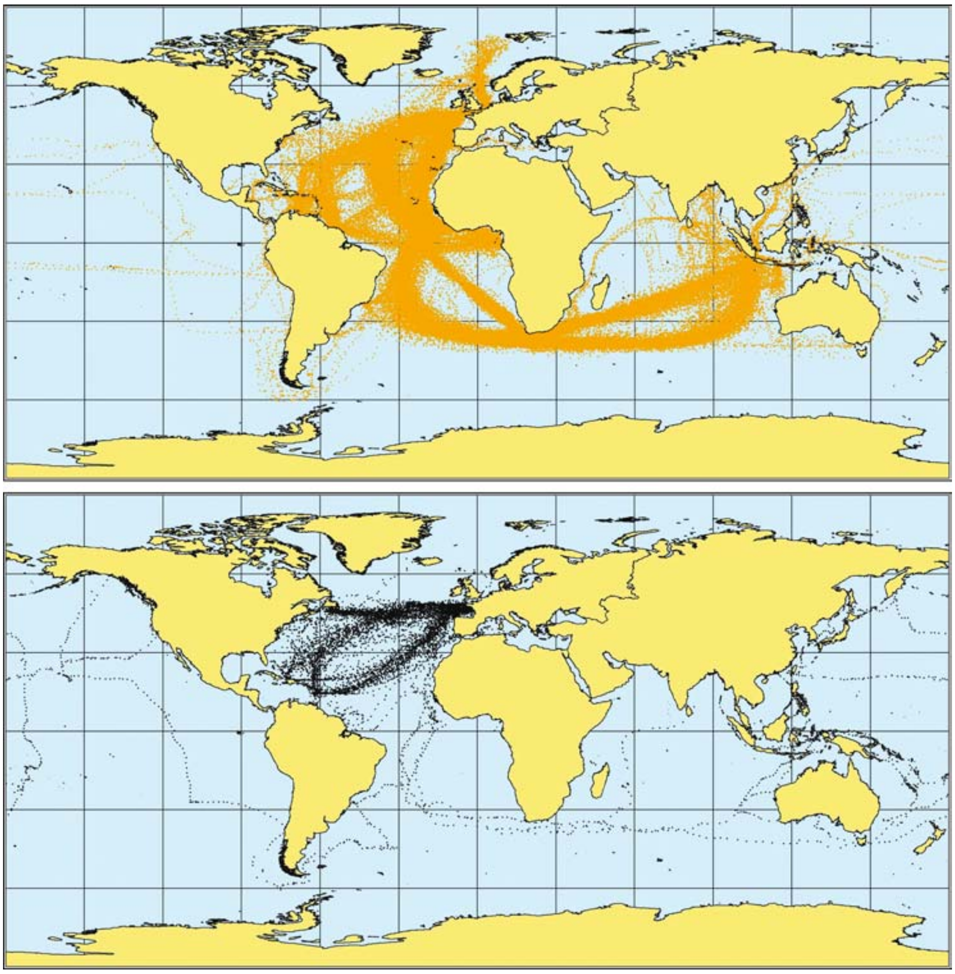


Figure 3. All ships' positions in CLIWOC 1750–1854. *Upper panel:* Dutch ships; *lower panel:* French ships.

4 in García-Herrera et al., this volume). The numbers of pressure and temperature measurements in Table IV are strikingly similar. This originates from the fact that ships that were instrumented, usually carried both a thermometer for recording the air temperature, and a barometer (with a second thermometer attached to it, to record the temperature of the barometer).

4. Output Format; Conversions to SI; Availability

The motivation for CLIWOC implies easy accessibility of the data. Since a long-term aim of the project team was incorporating the data into the ICOADS database,

TABLE II
Number of reports per ocean, by country and period

	North Atlantic	South Atlantic	Indian Ocean	Pacific Ocean	All Oceans
Spain					
≤1800	29,560	12,306	324	2,143	44,333
>1800	535	194	342	154	1,225
England					
≤1800	31,977	12,599	16,270	1,310	62,156
>1800	9,315	5,221	7,059	212	21,807
Netherlands					
≤1800	20,706	5,194	5,221	0	31,121
>1800	35,181	18,544	26,914	1,504	82,143
France					
≤1800	6,450	165	160	918	7,693
>1800	158	56	89	0	303
Others					
≤1800	160	121	108	0	389
>1800	101	41	48	0	190
Total	134,143	54,441	56,535	6,241	251,360

Note. Excluding the 28,835 extracted observations with uncertain positions (see Section 2.2).

TABLE III
Number of CLIWOC data per ocean by season

	North Atlantic	South Atlantic	Indian Ocean	Pacific Ocean	All Oceans
Dec–Jan–Feb	27,566	13,210	15,434	1,487	57,697
Mar–Apr–May	35,711	18,146	16,379	2,152	72,388
Jun–Jul–Aug	40,094	13,714	12,493	1,434	67,735
Sep–Oct–Nov	30,772	9,371	12,229	1,168	53,540
Total	134,143	54,441	56,535	6,241	251,360

early contacts were made with the ICOADS group. It was agreed that the format for the final CLIWOC data would be according to the International Maritime Meteorological Archive (IMMA) standards (Woodruff, 2004). This format allows the use of the present-day unit conventions in a core record along with the original reports in attachments to the core. The core contains the basic coordinates and meteorological elements transformed into standard SI (‘Système International’) units, together with pointers to attachments where all data are stored in the original languages and units. The attachments also give access to metadata, miscellaneous observed data and pathways to the available digital images. A reference table links the old geographical names to the modern English names and positions.

TABLE IV
Number of CLIWOC data 1800–1854 per ocean by element

	North Atlantic	South Atlantic	Indian Ocean	Pacific Ocean	All Oceans
Wind	42,183	21,986	31,158	1,775	97,102
Air temperature	20,056	11,487	16,062	1,057	48,662
Air pressure	20,220	11,409	16,233	917	48,779

Note. The numbers of pre-1800 pressure and temperature observations are negligible.

For wind speed, the transformation from the original terminology to SI units involves two steps. The first step converts the old descriptive wind force terms into Beaufort using the multilingual dictionary as lookup table (CLIWOC Team, 2003, see also Prieto et al., this volume; Koek and Können, this volume; Wheeler and Wilkinson, this volume). In the second step average m/s values are assigned to the Beaufort forces according to the WMO code 1100 scale (WMO, 1947). We note that the use of the WMO code 1100 midpoint and scale values is in accordance with the ICOADS practice (Slutz et al., 1985). To non-integer Beaufort class estimations we assigned the midpoint values of the upper or lower half of the wind speed range associated to the Beaufort class in question; to reports indicating a change in wind force toward a next class, the value of the class boundary was used (see Table V). To allow for updating, the transformation tables are included as dynamical modules within the database. In the database, it is always possible to revert to the original descriptions and units.

For the other weather elements, the transformations were as follows. Wind directions were converted from magnetic to true using the software that was kindly made available by Andrew Jackson of the Leeds Geophysical Research Group. Air temperature readings were converted from archaic units (Réaumur, Fahrenheit) to Celsius. Barometer readings were first converted into millimeters (using 25.4 and 27.0699534 millimeters for English and French inches, respectively) and then into hPa after which the pressure was reduced to standard gravity. No correction for the temperature or the height above the sea surface was applied to the pressure.

To facilitate a large group of users, the data are stored in ASCII (IMMA format), as well as in Microsoft Access format. Both Access97 and Access2000 are downloadable from the CLIWOC website. In due course, the data will be incorporated in the monthly summaries of ICOADS and integrated with the Maury data (see García-Herrera et al., this volume).

The Spanish partner will maintain the CLIWOC database. The data will also be accessible through KNMI and ICOADS as a separate data set. In the future we will investigate whether the methodologies developed during the production of the CLIWOC database are applicable to other data sets from similar sources.

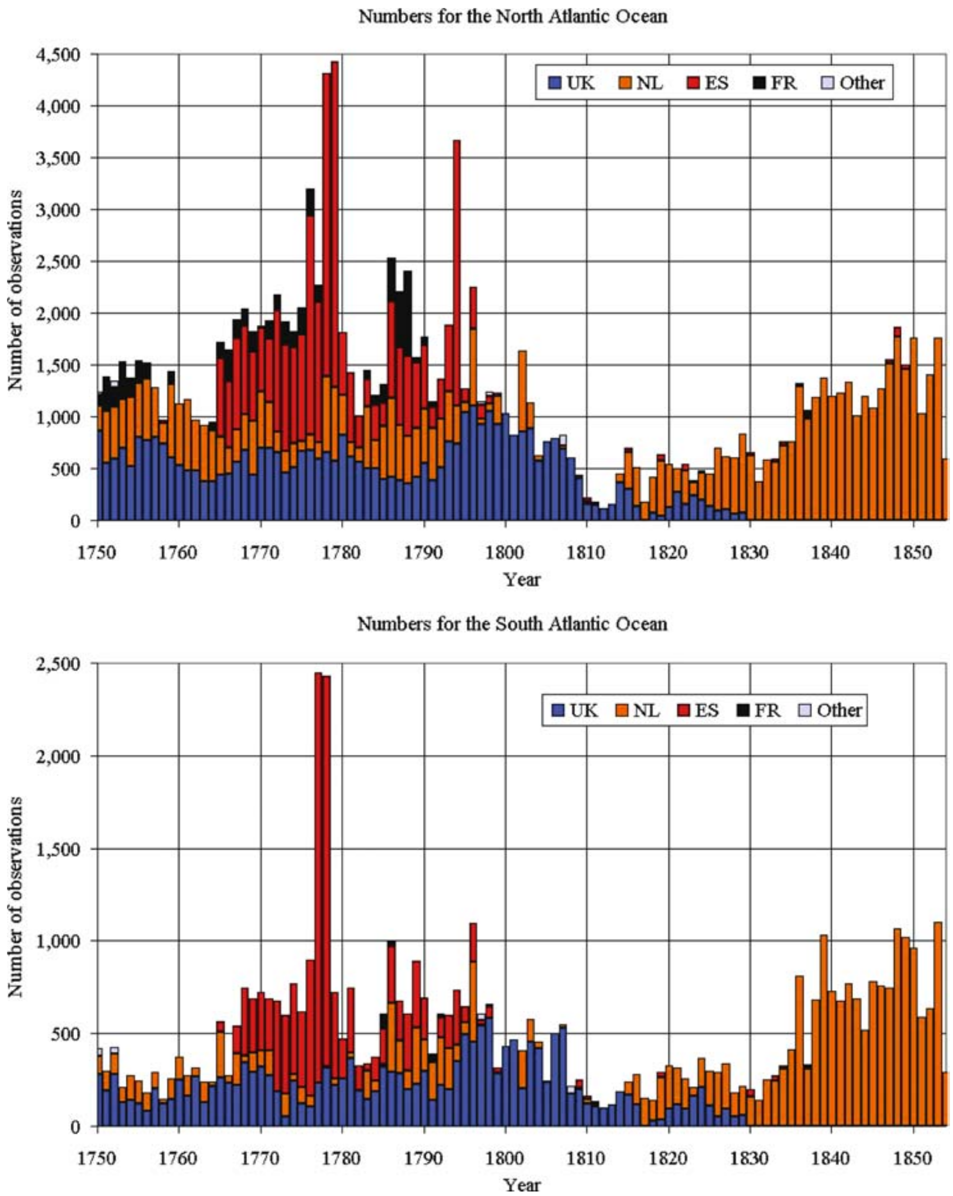
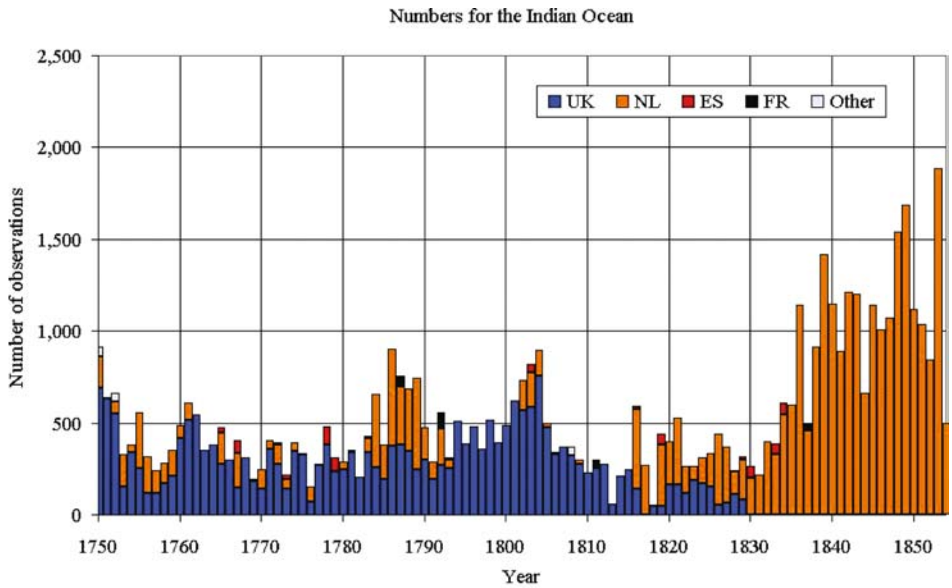


Figure 4. Annual number of observations included in CLIWOC by ocean and country. *Upper panel:* North Atlantic; *lower panel:* South Atlantic.

5. Current and Previous CLIWOC Releases

The EU CLIWOC project ran from 1 Dec 2000 until 1 Dec 2003. Before Oct 2003 the observational density in the database was not high enough to allow for the first scientific analyses (García-Herrera et al., 2004). The version described here is the



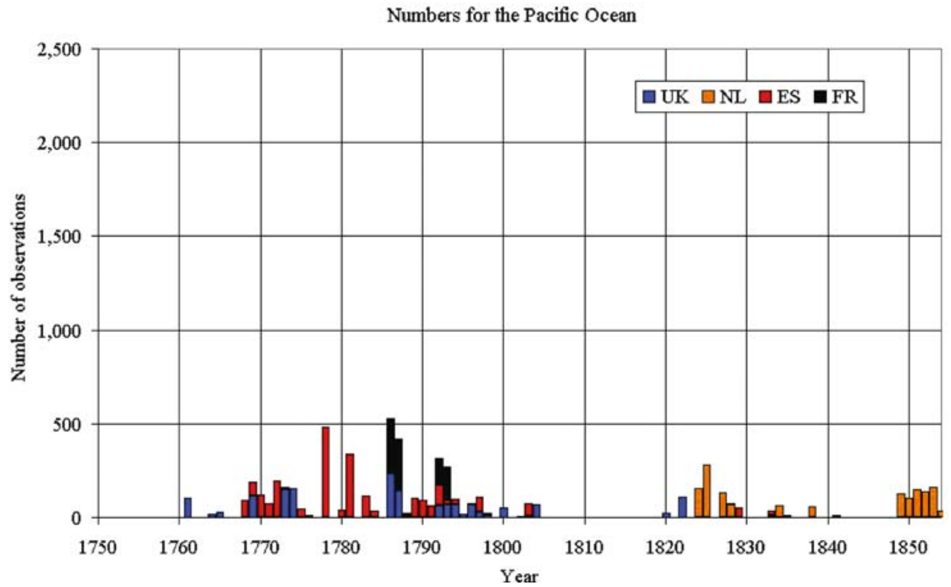


Figure 5. Annual number of observations included in CLIWOC by ocean and country. *Upper panel:* Indian Ocean; *lower panel:* Pacific Ocean.

fourth release (April 2004). It is published as CD-ROM and can be considered as the official outcome of CLIWOC. This official version is called Release 1.5. The current scientific results from CLIWOC (Jones and Salmon, this volume) are based on the third version, Release 1.1. As Table VI shows, the number of reports differs only little between Release 1.1 and the final version, Release 1.5.

TABLE V
Beaufort to m/s convention in the CLIWOC database, as based on WMO code 1100 scale (WMO, 1947)

Beaufort	Class boundary (m/s)	Lower-half value (m/s)	Midpoint value (m/s)	Higher-half value (m/s)	Class boundary (m/s)
0	0	0.0	0.0	0.1	0.2
1	0.3	0.6	1.0	1.2	1.5
2	1.6	2.0	2.6	2.9	3.3
3	3.4	4.0	4.6	4.9	5.4
4	5.5	6.1	6.7	7.3	7.9
5	8.0	8.7	9.3	10.0	10.7
6	10.8	11.6	12.3	13.1	13.8
7	13.9	14.7	15.4	16.3	17.1
8	17.2	18.1	19.0	19.8	20.7
9	20.8	21.7	22.6	23.5	24.4
10	24.5	25.5	26.8	27.4	28.4
11	28.5	29.5	30.9	31.6	32.6
12	32.7	33.9	35.0		

Note. In most cases (95%) the midpoints of the Beaufort classes were applied. Where the wind description allows for a subdivision within a Beaufort class, conversion took place via the lower half values viz. the upper half values. In cases that a report indicates a reduction towards the next-lower class, the lower class boundary value (second column) was applied; where the report indicates an increase towards the next-higher class, the upper class boundary value (last column) was applied.

TABLE VI
CLIWOC Releases June 2003–April 2004

Release #	Number of records	Valid by	Used for	Remarks
0.4	178,040	7 October 2003	Preliminary analyses	Limited availability
1.0	181,027	20 November 2003	Final presentation to EU	Also made available on Internet
1.1	239,853	23 January 2004	Jones and Salmon (2005)	Update on Internet; documented in García-Herrera et al. (2004)
1.5	280,195	15 April 2004	Final EU product on CD-ROM	Update on Internet

Acknowledgments

This work was supported by the European Commission, DG XII, Programme 2.4.1: ‘Better exploitation of existing data and adaptation of existing systems’, EC Framework V Project EVK2-CT-2000-00090 (CLIWOC). C. Koninckx kindly provided us with the Swedish logbooks. We thank Andrew Jackson of the Leeds Geophysical Research Group for making available the software for calculating the magnetic variation as a function of place and time. The CLIWOC database is accessible via Internet via http://www.ucm.es/info/cliwoc/, from which pointers are included for future consultation of the imaged logbooks.

References

CLIWOC Team: 2003, *CLIWOC multilingual meteorological dictionary, An English-Spanish-Dutch-French dictionary of wind terms used by mariners from 1750–1850*, KNMI publication 205, HISKLIM 5 (Available from the UK or Dutch partners).

García-Herrera, R., Wheeler, D. A., Können, G. P., Koek, F. B., Jones, P. D., and Prieto, M. R.: 2004, *CLIWOC final report*, UE contract EVK2-CT-2000-00090. Available from Dto. Física de la Tierra II, Facultad de Físicas, Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid, Spain.

García-Herrera, R., Können, G. P., Wheeler, D. A., Prieto, M. R., Jones, P. D., and Koek, F. B.: 2005, ‘CLIWOC: A climatological database for the world’s oceans 1750–1854’, *Climatic Change*, this volume.

Jackson, A., Jonkers, A. R. T., and Walker, M.: 2000, ‘Four centuries of geomagnetic secular variation from historical records’, *Philos. Trans. Roy. Soc. London, Philos. Trans. Math. Phys. Eng. Sci.* **358**, 957–990.

Jones, P. D. and Salmon, M.: 2005, ‘Preliminary reconstructions of the North Atlantic Oscillation and the Southern Oscillation index from wind strength measures taken during the CLIWOC period’, *Climatic Change*, this volume.

Koek, F. B. and Können, G. P.: 2005, ‘Determination of terms for wind force/present weather. The Dutch case’, *Climatic Change*, this volume.

Prieto, M. R., Gallego, D., García-Herrera, R., and Calvo, N.: 2005, ‘Deriving wind force from nautical reports through content analysis. The Spanish and French cases’, *Climatic Change*, this volume.

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, Climate Research Program, Boulder, CO, 268 pp.

Wallbrink, H., Koek, F. B., Können, G. P., and Brandsma, T.: 2003, ‘Sea-level pressure observations from Dutch ships 1854–1938 incorporated in COADS Release 1c climatology’, *Int. J. Climatol.* **23**, 471–475.

Wheeler, D. A. and Wilkinson, C.: 2005, ‘Understanding wind force and weather terms from ships’ logbooks: The English case’, *Climatic Change*, this volume.

Wheeler, D. A.: 2005, ‘An assessment of CLIWOC data quality and reliability’, *Climatic Change*, this volume.

WMO: 1947, *Comité Météorologique International, Procès-Verbaux de la Session de Paris, 1–12 Juillet 1946*, WMO, Publication No. 55, Lausanne.

Woodruff, S. D., Slutz, R. J., Jenne, R. L., and Steurer, P. M.: 1987, ‘A comprehensive ocean atmosphere data set’, *Bull. Am. Meteorol. Soc.* **68**, 1239–1250.

Woodruff, S. D. (ed.): 2004, *Archival of data other than in IMMT format. Proposal: International Maritime Meteorological Archive (IMMA) Format*. Update of JCOMM-SGMC-VIII/Doc. 17, Asheville, NC, USA 10–14 April 2000. Available from http://www.cdc.noaa.gov/coads/e-doc/imma/imma.pdf

Woodruff, S. D., Diaz, H. F., Worley, S. J., Reynolds, R. W., and Lubker, S. J.: 2005, 'Early ship observational data and ICOADS', *Climatic Change*, this volume.

Worley, S. J., Woodruff, S. D., Reynolds, R. W., Lubker, S. J., and Lott, N.: 2005, 'ICOADS Release 2.1 data and products', *Int. J. Climatol.* **25**, 823–842.

(Received 22 March 2005; in revised form 6 April 2005)