

DATA PAPER

WILEY

Geoscience
Data Journal

A historical Southern Ocean climate dataset from whaling ships' logbooks

Praveen R. Teleti¹ | W. G. Rees¹ | Julian A. Dowdeswell¹ | Clive Wilkinson²

¹Scott Polar Research Institute, University of Cambridge, Cambridge, UK

²Climatic Research Unit, University of East Anglia, Norwich, UK

Correspondence

Praveen R. Teleti, Scott Polar Research Institute, University of Cambridge, Cambridge, UK.

Email: prt31@cam.ac.uk

Abstract

Historical ship logbooks provide vital historic meteorological observations in the Southern Ocean, one of the largest climate-data deficient regions on the Earth. Christian Salvesen Whaling Company logbooks from whaling ships operating in the Southern Ocean, starting from the 1930s through the 1950s, are examined. Meteorological information contained in these logbooks has been extracted to produce a historical climate dataset. We discuss various instructions recommended by the British Admiralty to observe and record weather conditions on-board whaling ships. Statistical tests were used to flag erroneous values and corrections were made using neighbouring values. Meteorological parameters such as air pressure, air and sea temperature and wind force on the Beaufort scale were standardized, converting imperial to metric units. The data were structured according to the internationally accepted International Maritime Meteorological Archive format, which includes the most commonly reported meteorological variables, including the time, location and ship-related meta-data. Hence, a readily accessible, error-corrected and standardized historical climate dataset of the Weddell Sea sector of the Southern Ocean is presented.

KEYWORDS

Antarctica, data rescue, historical climate observations, Southern Ocean, whaling

1 | INTRODUCTION

The Southern Ocean is the least documented climatic region of the globe (Jones *et al.*, 2016). Much of the climate data on the Southern Ocean collected in the last

40 years or so is from polar orbiting satellites. Before the advent of satellites, meteorological information is drawn largely from exploratory expeditions to Antarctica and the Southern Ocean. A small number of scientific stations were built on the Antarctic Peninsula and sub-Antarctic islands starting from the early 20th century and, with the impetus of the IGY 1957-58, many more scientific stations on the Antarctic coast were established. To understand long-term climatic changes in Antarctic climate, many previous studies have turned to meteorological measurements taken at these few early stations (e.g. Turner *et al.*, 2005; Chapman and Walsh, 2007; Steig *et al.*, 2009; Nicolas and Bromwich, 2014; Fogt *et al.*, 2018). Due to near-continuous measurements taken at these stations since the 1950s, a

Dataset

Identifier: doi.org/10.17863/CAM.31530

Creator: Teleti, P. R., G. Rees, J. Dowdeswell, C. Wilkinson

Title: A historical Southern Ocean climate dataset from whaling ships' logbooks

Publisher: University of Cambridge

Publication year: 2018

Resource type: Book

Version: 1.0

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

clearer picture of climatic processes over Antarctica has emerged. However, understanding of climatic patterns and processes over the Southern Ocean remains less clear. Many multi-model ensembles generally disagree with observed changes (e.g. in sea ice) in the Southern Ocean for the common period (since 1980s), due to unresolved relationship between climate variables which results from the lack of long-term observations in this area (Zunz *et al.*, 2013; Shu *et al.*, 2015; Jones *et al.*, 2016).

To address the issue of a lack of meteorological observations over the oceans in general, many attempts have been made to systematically assemble marine observations combining different data sources taken on-board scientific, commercial and cruise vessels and from drifting/fixed ocean buoys. One of the largest such efforts is ICOADS Release 3.0 (International Comprehensive Ocean-Atmosphere Data Set, Freeman *et al.*, 2017), which is the most comprehensive dataset that combines data from in-situ marine meteorological observations, mainly from ships and buoys, and from many different national and international data sources. Individual observations are available from 1662 to 2014, while monthly summaries for the period 1800–2014 are found on $2^\circ \times 2^\circ$ grids, and on $1^\circ \times 1^\circ$ grids since 1960. The dataset comprises more than 40 core variables including many meteorological parameters, positional and ship-related meta-data. A number of specialized datasets focusing solely on one of the meteorological variables (e.g. temperature, surface and sea-level pressure) have emerged from ICOADS dataset. A dataset covering both land and ocean regions, the Hadley Centre-CRU Temperature dataset (HadCRUT4, Morice *et al.*, 2012), has been developed by the Climate Research Unit (CRU) at the University of East Anglia in conjunction with the Hadley Centre (UK Meteorological Office), in which the marine component of the dataset is derived from ICOADS. A further dataset focusing on surface and sea-level pressure observations, International Surface Pressure Databank (ISPD, version 2, Cram *et al.*, 2015), contain observations from land stations, marine observing systems and tropical cyclone best-track pressure reports. Similarly, the marine component of this dataset is derived from ICOADS.

Despite tremendous advances in the extraction and assimilation of data to produce global climate datasets, large regions of the Southern Ocean remain poorly represented, with a heavy reliance on observations from early Antarctic and Southern Ocean expeditions as primary sources for these datasets. Logbooks from these voyages do provide the first ever weather observations in these regions but are not sustained over time. Other sources of logbooks from commercial, fishing and whaling vessels traversing the Southern Ocean have been largely overlooked. The North American and European whaling industry focused on the Southern Ocean soon after over-fishing

led to the collapse of suitable stocks in the northern high-latitudes (Tønnessen and Johnsen, 1982). The first whaling fleet, the Dundee whaling expedition, is known to have visited the Falkland Islands in 1892–1893 (Headland, 2009). Since then until the 1980s, except for the years during two world wars, whaling ships hunted and caught whales almost every year in the Southern Ocean. A key advantage of using whaling logbooks as a source of meteorological data is that vessels usually visited the same whaling grounds year after year, providing sustained temporal coverage in these regions.

The reasons for under-utilization of this source may be because, as many of these logbooks are housed in different countries, private and commercial whaling logbooks are a relatively low priority compared to the more accessible national Antarctic expeditions in digitalization efforts, combined with wide usage of non-English languages in the logbooks. For this study, we have chosen to extract and assimilate a large number of observations from commercial whaling ships to create a historical meteorological dataset. Although the period of interest was from the early 20th Century to the International Whaling Commission's whaling moratorium in 1986, the spatial and temporal span of the data recovered was ultimately dictated by the data sources located and accessed. To locate such logbooks, we have consulted a report published by the RECLAIM project (<https://icoads.noaa.gov/reclaim/>), a contributory project to ICOADS, which listed all the identified archives of ships' logbooks in the Southern Ocean. It was found that the Centre for Research Collections, University of Edinburgh, contains a limited number of logbooks of the Christian Salvesen Whaling Company, a British whaling interest that operated a number of harbour and ship-based production facilities in the Southern Ocean from 1908 to 1963 (Vamplew, 1975).

Fortunately, the logbooks have been digitalized into 2,700 images by the RECLAIM project and were made available for this study (Wilkinson, 2016a). These logbooks are from whaling expeditions undertaken during 1930s and 1950s, and we use these logbook images to extract meteorological observations to construct a climate dataset of the Southern Ocean. The time period also reflects the two most prolific whaling decades in 20th Century, shaped by political and economic conditions (Jackson, 1978; Tønnessen and Johnsen, 1982). The newly extracted historical data are stored and made available in an internationally accepted format to streamline assimilation with existing datasets, preserving and extending existing international marine/climate datasets. To bring historical observations to the same standard as modern ones, historical observations are standardized and homogenized into modern units. In the following sections, we show the detailed methods of error detection–correction and standardization for each variable

or group of variables in the dataset, along with suitable storage formats. More details about observations and procedures used to make the meteorological measurements, and also efforts to convert those observations into modern units, are discussed in Section 2. After passing observational data through stringent quality-control checks and standardization processes, the resulting dataset is presented in Section 3. We then offer conclusions and future research tasks as a result of the newly created dataset.

2 | DATA AND METHODS

2.1 | Description of logbooks

The British Admiralty, with the help of the Meteorological Office's Marine Observers Handbook (MOH) (Her Majesty's Stationery Office (HMSO), 1930, 1950), issued meticulously detailed instructions on all aspects of weather

observation-taking and record keeping, including lists of instruments and observable parameters, and the methods and frequency with which these weather observations were to be taken. The observations were usually made by ships' navigating officers, first officers or other experienced scientific personnel on-board. The essential set of observed parameters include air temperature, wind conditions, sea-level pressure, sea state and a general description of weather, with time, date and position.

Logbooks in the current collection contain observations from 13 years or whaling seasons from the two decades of the 1930s and 1950s. Hand-written Chief Officers' (commonly known as deck) logbooks are the most common type of logbook in the current collection (Table 1). Deck logbooks were the principal source of information on navigation and weather observations, and were duly retained for legal and insurance purposes. A small number of catch books and H1-9 reports typed from original logbooks,

TABLE 1 List of vessels, whaling season, start–end dates, length of logbook observations and type of documents used in this study of Southern Ocean whaling logbooks from the Christian Salvesen Whaling Company

Name of the vessel	Whaling season	Start date	End date	Length of the season (in days)	Type of document
<i>New Sevilla</i>	1933–1934	29-Oct-1933	09-Mar-1934	131	H1-9
	1934–1935	23-Nov-1934	22-Apr-1935	150	H1-9
	1935–1936	26-Nov-1935	11-Mar-1936	106	H1-9
<i>Salvestria</i>	1931–1932	10-Jan-1932	25-Mar-1932	75	H1-9
	1932–1933	23-Oct-1932	03-Mar-1933	131	H1-9
	1933–1934	18-Oct-1933	04-Feb-1934	109	H1-9
	1934–1935	24-Nov-1934	03-Apr-1935	131	H1-9
	1935–1936	18-Nov-1935	17-Feb-1936	91	H1-9
<i>Sourabaya</i>	1932–1933	21-Sep-1932	06-Mar-1933	166	Catch book
	1933–1934	10-Sep-1933	06-Mar-1934	177	Catch book
<i>Southern Harvester</i>	1950–1951	30-Nov-1950	09-Mar-1951	98	Deck logbook
	1951–1952	22-Nov-1951	09-Mar-1952	107	Deck logbook
	1952–1953	11-Dec-1952	22-Mar-1953	101	Deck logbook
	1953–1954	17-Dec-1953	22-Mar-1954	95	Deck logbook
	1956–1957	18-Dec-1956	19-Mar-1957	91	Deck logbook
	1957–1958	28-Nov-1957	19-Mar-1958	111	Deck logbook
	1958–1959	19-Dec-1958	24-Mar-1959	95	Deck logbook
	1959–1960	11-Dec-1959	07-Apr-1960	118	Deck logbook
<i>Southern Venturer</i>	1950–1951	09-Dec-1950	12-Mar-1951	92	Deck logbook
	1951–1952	14-Nov-1951	13-Mar-1952	120	Deck logbook
	1952–1953	21-Nov-1952	04-Apr-1953	133	Deck logbook
	1953–1954	07-Dec-1953	19-Mar-1954	102	Deck logbook
	1954–1955	15-Dec-1954	27-Mar-1955	102	Deck logbook
	1956–1957	09-Dec-1956	21-Mar-1957	102	Deck logbook
	1958–1959	17-Nov-1958	17-Mar-1959	120	Deck logbook

issued by the British Ministry of Transport and Civil Aviation and US Hydrographic Office, respectively, are also present. The catch books recorded daily whale-catch numbers, the amount of blubber processed, and corresponding oil produced; they also include records of weather and positional information. All these documents record observations at noon except for deck logbooks, which recorded 4-hourly observations throughout the season. The logbooks also contain reports of floating ice in the form of icebergs and sea ice. Sightings of sea ice, and more often icebergs, were found in both types of logbook, more in the earlier Norwegian-language logbooks which describe different types of sea ice and icebergs in Norwegian terms which can only be approximately translated into English. Even though there were a moderate number of observations concerning sea ice and icebergs and their distance and direction relative to the ship, consistency in reporting is poor. That is unsurprising as sea ice types were not standardized at that time and type-ambiguity can produce misleading results. Hence, the sea ice and iceberg information was not included in the preparation of this dataset; however, such information can be obtained from authors separately.

The opportunity to use automated Optical Character Recognition (OCR) software to extract data from these records is severely restricted due to the frequent use of cursive writing, the presence of irregular abbreviations and the tabular structure of the logbook pages themselves. Trial OCR runs produced very high misread rates, even with training data, in a format that was very time-consuming to edit manually. In view of the quality of data acquired, and the time required to produce it using OCR methods, manual extraction was the preferred method for obtaining data from the logbooks. We have extracted close to 9,000 unique data records. Each record contains a number of positional, meteorological and meta-data fields, including latitude/longitude, air temperature, wind conditions, sea-level pressure, vessel name and port of registry, among others. All extracted raw data were added to a relational database for simple and systematic access.

2.2 | Instructions and error detection/correction

In the following sections, we consider the instructions issued by the Admiralty/MOH to better understand the procedures followed in acquiring these shipboard meteorological observations. The use of these instructions is twofold; first, instructions can point to sources of error and, secondly, they preserve the meta-data of observations, which could be used in future for comparing observations taken using different methods and adjusting them as necessary. Manually extracted data are not without their own challenges. The most common type of error is gross or

observational error; that is, misreading the numbers on instrument scales, together with faulty recording in logbooks which leads to typographical and transcriptional errors. To ensure internal consistency and temporal coherence, statistical tests were employed to identify cases of zero-variance ('consecutive identical values') or high variance ('outliers') for intra-day and consecutive inter-day observations. Once such cases were identified, they were either made missing or replaced by suitable value (usually the mean of temporally neighbouring values).

2.2.1 | Positional Information

Each meteorological record must contain a valid position to be useful. The ship's position (latitude and longitude) was observed and recorded once a day at noon as a usual practise; hence, noon position is assigned to all the observations taken during the preceding and following 12 hr. The Admiralty Manual (Chapter I, British Admiralty, 1938) instructs that positional observations were to be taken with reference to true north rather than magnetic north. Latitudinal and longitudinal position was recorded in degrees and minutes of four cardinal directions, which has been now converted into the degree and decimal system to facilitate further processing. A spatial plot of all raw data points (Figure 1a) shows that some locations have unrealistically large latitudinal and longitudinal differences from one day to the next, suggesting erroneous locations. Distance travelled between two positions (taken at mid-day of consecutive days) is considered a good indicator of spurious jumps in positions. To this end, all observations belonging to a 'Ship ID', which is a unique combination of vessel name and season, were queried from the database, and distance between neighbouring positions was calculated. The distance was measured in nautical miles (nm) following 'rhumb lines', as it was common practise to travel along rhumb lines at a constant compass bearing, rather than following arcs of a great circle. Assuming distance data follow a normal distribution, most of the values would be within three times the standard deviation (σ) from the mean (x); any values outside this bracket were treated as suspicious.

According to the central-limit theorem, the higher the difference between an individual value and the data mean, the more likely the value is to be an artefact of gross and/or transcriptional errors. We used the Generalized ESD (extreme studentized deviate) test (Rosner, 1983) to flag outliers for each unique Ship ID group. The test removes the observations that maximize $R_i = \max_j |x_j - \bar{x}| / \sigma$, which is the spread of individual values away from the mean value. The test then re-computes a number of R -values depending on estimates of the maximum number of outliers in the dataset. The exact number of outliers and their positions within

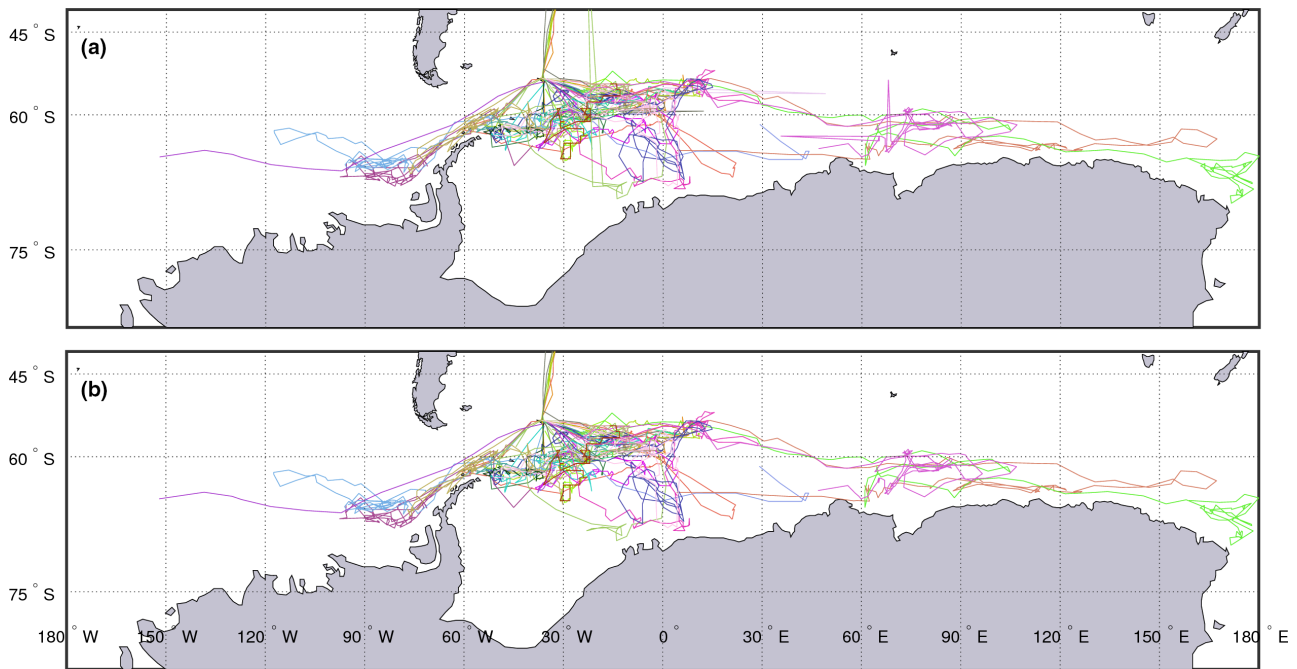


FIGURE 1 Ship tracks from the Christian Salvesen Whaling Co. logbooks separated into ‘Ship ID’ groups and denoted by different colours; raw (a) and corrected by statistical testing (b)

the dataset is determined by finding the largest i such that $R_i > \lambda_i$, λ_i is the critical t -value of R .

When outliers were plotted (not shown), most clustered together except for a few with very large deviations. Upon closer inspection, it was found that un-clustered outliers were indeed erroneous values; hence, they were replaced by the average of temporally neighbouring values. However, clustered groups of outliers showed an interesting pattern; almost all were from the days either at the start or end of the logbook entries. This could be explained by the fact that ships covered large distances to and from whaling grounds and their resupply base on South Georgia at the beginning and close of the whaling season (Figure S1). For the rest of the season, ships were usually drifting within whaling grounds and covering smaller amounts of daily mileage. We re-performed the statistical test on the corrected data, and no outliers were flagged barring values at the start and end of whaling seasons which confirms that distance data were a heterogeneous mixture of large and small distances. Corrected positions are shown in Figure 1b.

2.2.2 | Time information

Time information is also a part of the set of variables that must be present for each data point. Time-keeping instruments were supplied to all British whaling ships for navigational purposes from the Royal Observatory, Greenwich, or from the nearest chronometer depot (Chapter XI, British

Admiralty 1938). Each ship was allowed three chronometers and one deck watch to be used for day-to-day record keeping and navigation purposes. Navigating officers on-board carefully installed and maintained the chronometers in a designated room. Great effort was made to keep correct GMT time as an essential aid to navigation. Each of the three chronometers was compared with the other two, and readings from two chronometers showing near-identical time were taken as the correct GMT time and used to set the deck watch. If possible, other methods of time-keeping (e.g. wireless time signals, telegraphic time signals and astronomical observations) were also used to correct the on-board chronometers.

The longitudinal position was computed by following Admiralty Navigation Manual instructions (Misc. Chart 86, British Admiralty, 1938). The globe was divided into 24 time zones, each spanning 15° of longitudes. Each ship's local noon-time was set as the time when the sun reached its highest point in the sky. The difference between deck (GMT) time and local time was measured, and, if it was positive, then longitude was calculated to be 15° or its fraction East for each hour of difference and vice-versa if it was negative. Conversely, GMT time can be determined if the longitude of the ship's position and local time are known.

Whaling ships operated in two different time frames: deck time and factory or kitchen time, for operational reasons. The fact that ships remained in a time zone for many days or weeks on end while catching whales required the whaling operations to align to the particular local time

zone. Crew shifts, meal times and other on-board activities followed local time; however, all ships were recommended to use four principal hours 0000, 0600, 1200 and 1800 GMT (deck time) for the observation and recording of meteorological parameters (Chapter XIV, British Admiralty, 1938). Hence, it is assumed that the times recorded in deck logbook are GMT whereas catch book entries followed local time, which changed when the ship passed from one time zone to another. Each recorded time entry in the catch book was placed in its respective time zone according to the ship's longitudinal position. Local time was then converted into GMT using the procedure outlined earlier.

2.2.3 | Wind conditions

Before the widespread use of instrumental anemometers, the direction and force of the wind were estimated visually. Wind direction is specified as a point of the true compass from which wind blows, and is observed to the nearest true compass point (Chapter XIV, British Admiralty, 1938).

TABLE 2 Beaufort wind scale with description and visual criteria (Chapter XIV, British Admiralty, 1938)

Beaufort number	Wind speed (NM per hr or knots)	Wind description	Visual criteria
0	Less than 1	Calm	Sea mirror-smooth
1	1–3	Light air	Small wavelets like scales; no foam crests
2	4–6	Light breeze	Waves short and more pronounced; foam has glassy appearance
3	7–10	Gentle breeze	
4	11–16	Moderate breeze	Waves are longer and white
5	17–21	Fresh breeze	Waves more pronounced and long; white
6	22–27	Strong breeze	Larger waves form
7	28–33	Strong wind	Sea heaps up; wind blow foam in streaks
8	34–40	Fresh gale	Height of waves increases visibly; foam is blown in dense streaks
9	41–47	Strong gale	
10	48–55	Whole gale	High waves with long over hanging crests
11	56–65	Storm	Waves so high that ships within sight are hidden in the troughs; air filled with spray
12	Above 65	Hurricane	

Wind force is expressed by means of the Beaufort wind scale (Simpson, 1906), a 13-point numerical scale devised in 1808 and named after Admiral Sir Francis Beaufort. The Beaufort scale was used to record the wind force in the deck logbooks and catch books. A companion table was printed in the preface of each logbook, which supplied conversion scales and visual criteria to aid observers on a ship's deck (Table 2).

In our present collection, some early logbooks use wind terms to describe wind strength. All of the encountered wind terms were resolved to one of the levels of the Beaufort scale. Once wind strength is established in terms of the Beaufort scale, it can be expressed easily in knots or m/s. Thereafter, wind force and direction data were separately tested for outliers using the Generalized ESD test. A small fraction (less than 0.5%) of the values for each Ship ID was found to be outliers. Erroneous values were replaced by the average of neighbouring values, if available, or otherwise made null in the database.

2.2.4 | Sea state and swell

The sea state and swell are closely related to wind conditions. The difference is that sea state is defined as those waves caused by ambient wind conditions, whereas swell is produced by waves formed by past wind action, or by wind blowing at a distance. Careful observation of sea state and swell was vital to the detection of weather systems forming and passing by at a distance from a ship (Table 3). A *short swell* means a swell where the length or distance between each successive wave crest is relatively small. A *long swell* means a swell in which the length or distance is large. A *low swell* means a swell where the height between the lowest and highest part of the swell is small. A *heavy swell* means a swell where height is great.

Both sea state and swell are recorded in the deck log by means of the adjusted Douglas Sea and Swell scale (WMO Code table 3700, Manual on Codes, No. 306, part A) (Table 3). The direction of the swell is specified in a similar way to that of wind direction; that is, the point of the 16-point compass from which the swell travels. Sea-state observations were converted into the numerical height of wave fields using Table 3. Wave height data were passed through the Generalized ESD test to flag outliers, and flagged observations were replaced by the mean of temporally neighbouring values, if available, or otherwise made null. Both sea-state and swell wave heights were standardized to metres.

2.2.5 | Air and sea-surface temperature

One of the main meteorological parameters of interest is temperature and all whaling ships were equipped with mercury thermometers to measure it. Most of the logbooks

TABLE 3 Ten-point Sea state scale (left) and Swell scale (right), adapted from Douglas Sea and Swell Scale (WMO Code table 3700)

Sea state			Swell scale		
Scale No.	Description	Height of waves – crest to trough (in ft.)	Code No.	Description	Swell state
0	Calm (Glassy)	0	0	None	Low
1	Calm (Rippled)	0.00–0.33	1	Short or average length	
2	Smooth	0.33–1.64	2	Long	
3	Slight	1.6–4.1	3	Short	Moderate height
4	Moderate	4.1–8.2	4	Average length	
5	Rough	8.2–13.1	5	Long	
6	Very rough	13.1–19.7	6	Short	Heavy
7	High	19.7–29.5	7	Average length	
8	Very high	29.5–45.9	8	Long	
9	Precipitous	Over 45.9	9	Confused	

record only air temperature, which was measured using the dry bulb of a psychrometer. The psychrometer structure was placed about 5 feet above the upper deck in the open air, as free as possible from the sun's radiation or warm air from galleys, engine and boiler rooms (Chapter XIII, British Admiralty, 1938). To maximize the exposure to the ambient weather, the psychrometer was usually hung towards the windward side of the ship. At least 15 min were allowed to pass after placing it in position for temperatures to settle before taking observations.

Ship-design changes from sail to diesel had more impact on the methodology of sea-surface temperature (SST) measurements than air temperatures. Originally a wooden or a canvas bucket was lowered to sample and draw up sea water, before thermometers enclosed in metal cases were put into the bucket to measure SSTs. The water was drawn clear of any discharge from the ship and the thermometer was kept inserted in the water for several minutes to obtain stable readings. If a canvas bag was used, it was not placed in a draught as evaporation could lower the measured temperature. When ships became diesel-engine driven, SSTs were measured directly from seawater that was let in to cool the engines in the engine room. Measurements were taken with a mercury thermometer placed as close as possible to the water inlet (Kent *et al.*, 2010).

As it was more usual to measure air temperatures than SSTs, if logbooks were ambiguous about the source of measurement, it was considered to be air temperatures. Where available, both temperatures were passed through a Generalized ESD test to flag outliers, separated by Ship ID groups and parameters. The flagged outliers were replaced with reference to neighbouring values. Almost all observations were recorded in degrees Fahrenheit, which were converted to degrees Celsius in the database.

2.2.6 | Sea-level pressure

Along with wind and swell observations, atmospheric pressure measurements were vital in keeping ships safe from hazardous storms in the high seas. Vessels in the path of powerful storms could capsize or be severely damaged by the waves and sea swell. Hence, great care was taken to observe and record changes in atmospheric pressure as radio/wireless weather forecasts were not available for the Southern Ocean in the 1930s and 1950s. All British whaling ships were supplied with Kew-pattern marine mercurial barometers (Chapter XIII, British Admiralty, 1938). Each barometer was fitted with a Gold slide which could offset the pressure-column reading for latitude, height of instrument above sea-level and temperature. It was recommended to note the Gold slide offset before reading the barometer, as heat from the body of the observer could affect the Gold slide more quickly than the barometer.

The above corrections were necessary to reduce the barometer observations to sea-level and a latitude of 45° (standard unit of atmospheric pressure; Chapter XIII, British Admiralty, 1938). When a ship moved violently, the mercury jerked up and down along the scale, and the mean of the highest and lowest heights was taken as the value of the barometric pressure. Logbooks in the earlier years of our collection recorded pressure in inches of the mercury column. By the 1950s, this had changed to millibars. Once again, we performed a Generalized ESD test to determine extraction and typographical errors. We found no outliers in the data; indeed, air pressure readings appeared to have been very carefully and meticulously observed and recorded. To bring uniformity to the observations in the dataset, a standard formula was used to convert mercury-column inches into hecto-Pascals (hPa or millibars).

2.3 | Standardization

The aim of producing a readily accessible dataset cannot be achieved without adhering to some generally accepted format. The International Maritime Meteorological Archive (IMMA) format (Woodruff, 2007) comprises a comprehensive set of marine climate variables, including the most commonly reported meteorological variables with the time, location and ship-related meta-data, among others. IMMA1 (IMMA version 1, the latest version), adopted in the ICOADS 3 dataset, is an internationally accepted format to integrate historical weather information from diverse platforms and national/international sources. A number of earlier formats exist, ranging from 1853 Maritime Conference conventions to the WMO's early International Maritime Meteorological (IMM) punched-card format (Yoshida, 2004, WMO 1952 publication), to Global Telecommunication System (GTS) codes used to transmit weather reports to land-based stations from ships; and many other reporting practises. With the myriad of conventions and formats used to record, transfer and archive historical and near-contemporary weather data, it was necessary to devise a common format that would make different data sources compatible with each other. Hence, the IMMA format was produced by retaining the best features and concepts of previously used formats, but providing a new format that is better aligned with modern electronic data services and storage.

The IMMA format is designed to be flexible in terms of the number of meteorological and meta-data fields to the level of individual records in the datasets. The IMMA format record consists of an essential set of parameters called Core, followed by number of different attachments (*atms*). The Core is divided into locational and meteorological sections, incorporating many of the most commonly used parameters in a standardized form (fields listed in IMMA, Supp. D, more information can be found in IMMA report). We are using Ship meta-data (Meta-*vos*) *atm* (C7) to store meta-data from the logbooks, including recruiting country and the country of registration of the ship, the types of thermometers, barometers and other details. We have constructed the dataset following the recommendations regarding format, conventional codes (e.g. country codes, variable indicators etc.) and precision for each variable in the core (C0) and Meta-*vos* *atm* (C7) fields.

3 | RESULTS AND FUTURE STEPS

We have created a readily accessible, standardized and quality checked IMMA-compliant dataset of meteorological observations from the Southern Ocean. This dataset is a result of the first ever study to extract meteorological observations solely from whaling logbooks in the Southern Ocean. Each record contains a number of positional,

meteorological and meta-data parameters found in the Christian Salvesen Co. whaling logbooks of the 1930s and 1950s. Each parameter was manually extracted from logbooks and stored in a relational database. All data points were passed through statistical tests to flag and correct erroneous values. To make the dataset accessible and interoperable with other marine datasets, data were homogenized and standardized. Each record in our dataset was produced according to recommendations in the IMMA format, bringing them to a level similar to that of existing international datasets relating to historical meteorological records such as ICOADS.

In total, the assembled dataset contains close to 9,000 observations recorded during 1870 observation-days spanning two decades. It contains 71 variables in total, including 48 and 23 variables for Core and Ship meta-data sections respectively. All data records have positional and time fields due to our quality systems' insistence on non-null data for these fields. In addition, wind conditions, air pressure and temperature fields are more populous than other climatic fields. The meta-data fields are collected from various sources and are provided alongside the meteorological observations. For example, the dimensions of the observing platform (e.g. ship) are not recorded in the logbooks, but by searching the UK Shipping Registry against the name of the vessel mentioned in the logbook, the Official Number (ON), type of vessel, number of engines, dimensions and other ancillary information are obtained. In the following section, we inspect the spatial and temporal characteristics of the dataset.

3.1 | Spatial and temporal spread of dataset

A preliminary exploration of the new dataset was undertaken. All observations were divided into separate seasons and plotted (Figures S2 and S3). We have collated data from four and nine seasons in 1930s and 1950s, respectively. These two decades (1930s and 1950s) were the only time period represented in the current collection of Salvesen logbooks. Interestingly, this period could contain far more observations than any other period in the whaling history in the Southern Ocean, due to the increased number of whaling expeditions at this time (Jackson, 1978; Tønnessen and Johnsen, 1982). The collected data from 1930s show that whaling activity was much more confined to South Georgia and the northern and western edges of Weddell Sea in that pre-WWII period (Figure S2). The highest number of observations for any season in the 1930s decade comes from the 1933–1934 season, yet the longest season was the 1932–1933 season (Table 1; Figure S2). The whaling season would typically start between mid-September and early-November and continue until March of the next year, making the length of the season in this part of the

Southern Ocean between about 150 and 170 days (Table 1).

On the other hand, whaling positions in the 1950s are much more spread out, covering almost the whole circumpolar Southern Ocean (Figure S3). This could be partly due to the scarceness of whaling stocks around South Georgia and the larger number of competitors in the 1950s than in the 1930s. The whaling seasons started much later in the 1950s as compared with the 1930s but the length of season is longer (typically about 180–200 days). The start and length of whaling season were dictated by weather conditions and international regulations (Jackson, 1978; Tønnessen and Johnsen, 1982). Even though the length of seasons in the 1930s and 1950s are fairly comparable (Table 1), 70% of the meteorological data in the dataset comes from 1950s. This is due to the fact that, on average, there were six observations per day in the 1950s as compared to 1.46 observations in the 1930s.

Next, to better understand the spatial density of the dataset, all positional data were binned in a 0.5° latitude-longitude grid box (Figure 2). The average number of observations in a grid box is weighted by the cosine of the central latitude of the grid box. As shown in the spatial plots for each season (Figures S2 and S3), the Weddell Sea sector is the most densely populated region, followed by the Amundsen and Bellinghousen seas (ABS) sector and the Indian Ocean sector. The data density reflects the whaling practices of that time; for example, the marginal sea ice zones around the Weddell Sea and much of the Antarctic coastline were popular whaling grounds due to abundance of whale pods and shelter provided by the ice pack, which dampens waves, during bad weather (McLaughlin, 1976; Tønnessen and Johnsen, 1982).

3.2 | Planned use of the historical dataset and data-rescue efforts

Observational meteorological datasets such as ICOADS3 and ISPD contain the world's largest collection of global marine observations; however, significant gaps exist in the Southern Ocean. These datasets are routinely used to constrain climate model runs to produce long-term climate datasets, also known as climate reanalysis (e.g. the Twentieth Century Reanalysis version 2 (20CRv2); Compo *et al.*, 2011). The meteorological observations derived from historical logs of whaling vessels in this study, compliment these existing observational datasets by filling a considerable data void in the Southern Ocean. The observations can also be fed into the growing number of numerical reanalysis climate models used to generate increasingly accurate pre-modern climate datasets. Furthermore, the relatively dense set of observations comprising the whaling dataset can be used independently to study and to

reconstruct circumpolar or regional climate by using appropriate spatio-temporal gridding techniques. Such gridded dataset can enhance our understanding of the background climatology of the Southern Ocean from the pre-satellite era, with which it is important to compare more recent observations and trends. For example, the MSLP observations can be used to reconstruct the Southern Annular Mode (SAM; Marshall, 2003). Similarly, storminess and frequency of depressions tracking through the Southern Ocean is another example of evidence concerning synoptic climatology that can be derived from these historic datasets.

In addition, the RECLAIM project has identified a large cache of largely unknown Christian Salvesen Whaling Co. logbooks at the Sea Mammal Research Unit (SMRU), University of St. Andrew's, Scotland (Wilkinson and Wilkinson, 2018). In total 14,900 new images have been captured, with most logbooks from British-flagged whale factory ships (together with one Norwegian and two Japanese vessels). These logbooks contain the usual positional, weather and sea-conditions and ice-condition records. Time and resource permitting, data extracted from the logbooks found at St. Andrew's, together with the dataset presented in this study, can close substantial gaps in our knowledge of Southern Ocean climate in the pre-satellite period. A further collection of whaling logbooks at the Vestfold Archives in Sandefjord, Norway, has also been identified and photographed, producing more than 30,000 digital images of logbooks, catch books and day reports (Wilkinson, 2016b). This study has shown that whaling logbooks can be used to extract valuable meteorological observations from little known area of the World and points to their large untapped potential for data-rescue efforts and the construction of historical meteorological datasets.

A review of data-rescue priorities and practises undertaken by Brönnimann *et al.* (2018), states that to make these data-rescue efforts more effective, they should be continuous and coordinated with a long-term goal. One of the largest international data-rescue initiatives, Atmospheric Circulation Reconstructions over the Earth (ACRE; www.met-acre.org/) coordinates over 50 organizations in various countries. The ACRE initiative both undertakes and facilitates the retrieval of historical instrumental surface, terrestrial and marine weather observations. A few examples of ACRE supported projects are ISPD, the U.K. Colonial Registers and Royal Navy Logbooks project (www.corral.org.uk), ICOADS3, RECLAIM, the International Environmental Data Rescue Organisation (IEDRO; www.iedro.org), and NOAA's NCDC Climate Database Modernization Program (CDMP; www.ncdc.noaa.gov/oa/climate/cdmp/cdmp.html). Other, similar initiatives are Euro-Climhist database (www.euroclimhist.unibe.ch), Tabora (www.tabora.org) and Old Weather (oldweather.org). In the

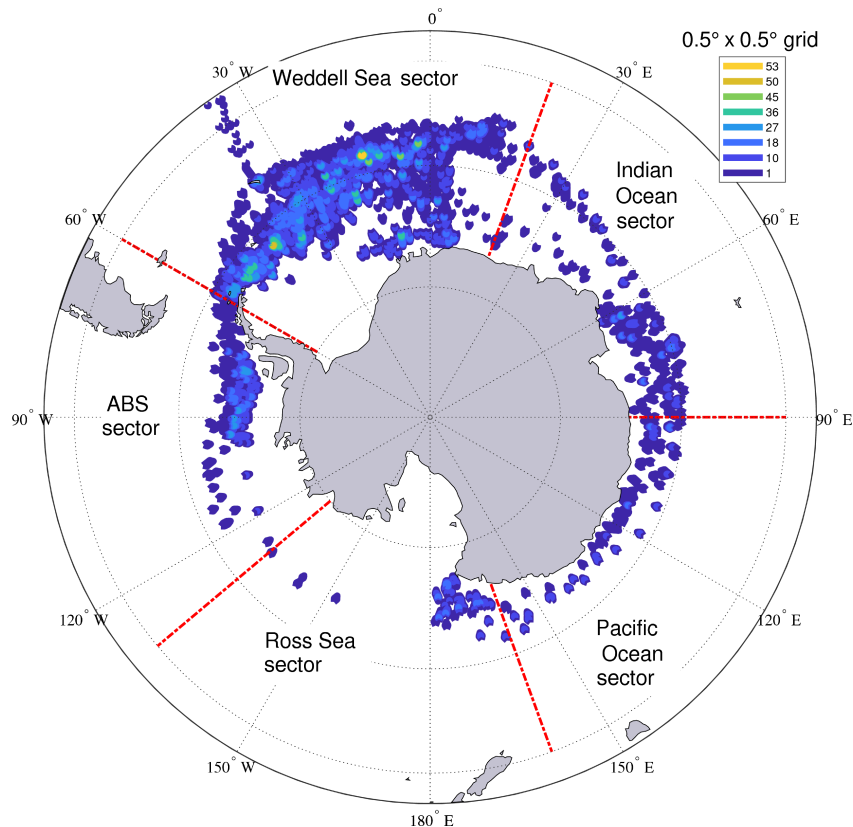


FIGURE 2 Map of positional data from Christian Salvesen Whaling Co. ships operating in the Southern Ocean in the 1930s and 1950s. All data points are binned in 0.5° grid boxes and all grid boxes shown contain at least one observation from the whole period

Southern Ocean region, the Southern Weather Discovery project, part of ACRE Antarctica (www.zooniverse.org/projects/drewdeepsouth/southern-weather-discovery), is at the forefront of data-rescue endeavours by using the collective efforts of many citizen scientists to decipher, translate and extract meteorological information from station observations, weather diaries and ship logbooks.

The current dataset is made available in CSV (comma separated values) format file and an accompanying description file. The dataset is held in the *Apollo* digital database, University of Cambridge Data Repository (doi.org/10.17863/CAM.31530), with free access to the dataset through a request made at the dataset webpage.

OPEN PRACTICES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at <https://doi.org/10.17863/CAM.31530>. Learn more about the Open Practices badges from the Center for Open Science: <https://osf.io/tvyxz/wiki>.

ACKNOWLEDGEMENTS

The RECLAIM project (<http://icoads.noaa.gov/reclaim/>) is in part facilitated by ACRE-Oceans, a sub-chapter of ACRE

initiative, that deals with marine records. C.W. acknowledges funding received from the UK Meteorological Office, Hadley Centre to undertake scanning of logbooks at University of Edinburgh. Prof. Vegard Nergård of the Arctic University of Norway, Tromsø, is thanked for help with translation and extraction of weather observations from Norwegian logbooks. The authors thank two anonymous reviewers for the helpful comments and suggestions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Praveen R. Teleti  <https://orcid.org/0000-0003-2691-8488>
W. G. Rees  <https://orcid.org/0000-0001-6020-1232>

REFERENCES

- British Admiralty. (1938) *Admiralty Navigation Manual Volume-I*. London: Admiralty.
- Brönnimann, S., Brugnara, Y., Allan, R., Brunet, M., Compo, G., Crouthamel, R. *et al.* (2018) A roadmap to climate data rescue services. *Geoscience Data Journal*, 5(1), 28–39.

- Chapman, W.L. and Walsh, J.E. (2007) A synthesis of Antarctic temperatures. *Journal of Climate*, 20, 4096–4117. Available at: <https://doi.org/10.1175/jcli4236.1>
- Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X. *et al.* (2011) The twentieth-century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137, 1–28.
- Cram, T.A., Compo, G.P., Yin, X., Allan, R.J., McColl, C., Vose, R.S. *et al.* (2015) The international surface pressure databank version 2. *Geoscience Data Journal*, 2, 31–46. Available at: <https://doi.org/10.1002/gdj3.25>
- Fogt, R.L., Clark, L.N. and Nicolas, J.P. (2018) A new monthly pressure dataset poleward of 60°S since 1957. *Journal of Climate*, 31 (10), 3865–3874.
- Freeman, E., Woodruff, S.D., Worley, S.J., Lubker, S.J., Kent, E.C., Angel, W.E. *et al.* (2017) ICOADS release 3.0: a major update to the historical marine climate record. *International Journal of Climatology* (CLIMAR-IV Special Issue), 37, 2211–2237. Available at: <https://doi.org/10.1002/joc.4775>
- Headland, R.K. (2009) A Chronology of Antarctic Exploration: A Synopsis of Events and Activities from the Earliest Times Until the International Polar Years, 2007-09. London: Bernard Quartich, pp. 772, ISBN 978-0-9550852-8-4.
- Her Majesty's Stationery Office (HMSO). (1930) *Marine Observer's Handbook*, M.O. 218, 5th edition. London: HMSO.
- Her Majesty's Stationery Office (HMSO). (1950) *Marine Observer's Handbook*, M.O. 522, 7th edition. London: HMSO.
- Jackson, G. (1978) *The British Whaling Trade*. Haven, Connecticut: Shoe String Press Inc.
- Jones, J.M., Gille, S.T., Goosse, H., Abram, N.J., Canziani, P.O., Charman, D.J. *et al.* (2016) Assessing recent trends in high-latitude Southern Hemisphere surface climate. *Nature Climate Change*, 6, 917–926 ISSN 17 58-678X.
- Kent, E.C., Kennedy, J.J., Berry, D.I. and Smith, R.O. (2010) Effects of instrumentation changes on sea surface temperature measured *in situ*. *Wiley Interdisciplinary Reviews: Climate Change*, 1, 718–728. Available at: <https://doi.org/10.1002/wcc.55>
- Marshall, G.J. (2003) Trends in the Southern Annular Mode from observations and reanalyses. *Journal of Climate*, 16, 4134–4143.
- McLaughlin, W. (1976) *Call to the South*. London: White Lion Publishers.
- Morice, C.P., Kennedy, J.J., Rayner, N.A. and Jones, P.D. (2012) Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset. *Journal of Geophysical Research*, 117, D08101. Available at: <https://doi.org/10.1029/2011JD017187>
- Nicolas, J.P. and Bromwich, D.H. (2014) New reconstruction of Antarctic near-surface temperatures: multidecadal trends and reliability of global reanalyses. *Journal of Climate*, 27, 8070–8093. Available at: <https://doi.org/10.1175/jcli-d-13-00733.1>
- Rosner, B. (1983) Percentage points for a generalized ESD many-outlier procedure. *Technometrics*, 25(2), 165–172.
- Shu, Q., Song, Z. and Qiao, F. (2015) Assessment of sea ice simulations in the CMIP5 models. *The Cryosphere*, 9, 399–409.
- Simpson, C.G. (1906) *The Beaufort Scale of Wind Force*. London: Report of the Director of the Meteorological Office, Official No. 180.
- Steig, E.J., Schneider, D.P., Rutherford, S.D., Mann, M.E., Comiso, J.C. and Shindell, D.T. (2009) Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature*, 457, 459–462.
- Tønnessen, J. and Johnsen, A. (1982) *The History of Modern Whaling*. Berkeley, California: University of California Press.
- Turner, J., Colwell, S.R., Marshall, G.J., Lachlan-Cope, T.A., Carleton, A.M., Jones, P.D. *et al.* (2005) Antarctic climate change during the last 50 years. *International Journal of Climatology*, 25, 279–294.
- Vamplew, W. (1975) *Salvesen of Leith*. Edinburgh, UK: Scottish Academic Press.
- Wilkinson, C. (2016a) “Report on the Imaging of Sources of Historic Ice, Meteorological and Oceanographic Data in Antarctic Waters Christian Salvesen Archive University of Edinburgh”, Recovery of Logbooks and International Marine Data, RECLAIM Project.
- Wilkinson, C. (2016b) “Report on the Imaging of Sources of Historic Ice, Meteorological and Oceanographic Data in Antarctic Waters Vestfold Archive, Sandefjord”, Recovery of Logbooks and International Marine Data, RECLAIM Project.
- Wilkinson, C. and Wilkinson, S. (2018) “Report on the Imaging of Sources of Historic Ice, Meteorological and Oceanographic Data in the Southern Ocean – Archive of the Sea Mammal Research Unit, University of St. Andrews, Scotland”, Recovery of Logbooks and International Marine Data, RECLAIM Project, 2018.
- Woodruff, S.D. (2007) Archival of Data Other Than in IMMT Format: The International Maritime Meteorological Archive (IMMA) Format. In Second Session of the JCOMM Expert Team on Marine Climatology (ETMC), Geneva, Switzerland, 26–27 March 2007. JCOMM Meeting Report No. 50, 68–101. Available at: <http://icoads.noaa.gov/e-doc/imma/> [Accessed 24 October 2018].
- Yoshida, T. (2004) History of the Marine Ship Code. Submitted to the First Session of the JCOMM Expert Team on Marine Climatology (ETMC-I/Doc. 4.2). [Note: Available at: icoads.noaa.gov/etmc/. and JCOMM (2004a).]
- Zunz, V., Goosse, H. and Massonnet, F. (2013) How does internal variability influence the ability of CMIP5 models to reproduce the recent trend in Southern Ocean sea ice extent? *The Cryosphere*, 7, 451–468.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Teleti PR, Rees WG, Doa wdeswell JA, Wilkinson C. A historical Southern Ocean climate dataset from whaling ships' logbooks. *Geosci Data J.* 2019;6:30–40. <https://doi.org/10.1002/gdj3.65>