

Ships' logbooks from the Arctic in the pre-instrumental period

Matthew Ayre¹, John Nicholls², Catharine Ward¹ and Dennis Wheeler¹

¹Formerly University of Sunderland, Sunderland, SR1 3SD, UK

²History Department, University of Hull, Hull, HU1 7RX, UK

*Correspondence: Dennis Wheeler, Formerly University of Sunderland, Sunderland, SR1 3SD, UK.

E-mail: denniswheeler1948@msn.com

The article was funded by the Leverhulme Trust and Durham University Alumni Fund.

Ships' logbooks are now an accepted part of the repertoire of data sources in climate change studies. This article examines some of the particular issues surrounding logbooks from the Arctic region in the so-called pre-instrumental period. Attention is given to the means by which narrative descriptions of wind, weather and sea ice cover can be reliably expressed in index form. Consideration is also given to the various means by which these data can be most effectively managed for scientific analysis as in most cases they were not recorded for such purposes. Many such logbooks remain yet to be digitized and the methods described here can be applied with equal confidence in future through such undertakings using English language documents.

Dataset

Identifier: <http://www.hull.ac.uk/mhsc/ARCDoc/>

Creator: ARCDoc research team

Title: Hudson's Bay Company ships' logbooks

Publisher: University of Hull HYDRA Unit

Publication year: 2015

Resource Type: Climate data set

Version: 1.0

Geosci. Data J. 2: 53–62 (2015), doi: 10.1002/gdj3.27

Received: 16 May 2014, revised: 29 June 2015, accepted: 13 July 2015

Key words: arctic climate, documentary sources, ships' logbooks, historical climatology, air circulations

Introduction

In 2008, the Leverhulme Trust awarded a research grant to a consortium headed by the University of Sunderland. The principle objective of the exercise was to extract daily weather data from the logbooks of British whaling ships and of vessels in the service of the Hudson's Bay Company (HBC) for the period 1750–1850. These data were then to be expressed in a form that could be readily used by climate researchers and made available through an accessible website. An overview of this, the so-called ARCDoc project, can be found at <http://arcdoc.wordpress.com>. The project was one of source identification, data gathering and processing prior to making the data freely available. Interpretation was not seen as an integral part of the project, and that task was always intended to be left to those more adequately skilled in such applications of climate data. However, this identification of source

material, the determination of the nature of the data and the development of methods by which such, exclusively non-instrumental, data can be prepared for scientific analysis was thought to be of sufficient importance to warrant reporting here and to alert those who may follow on the hazards, and benefits, offered by such data. What follows is a description of the UK sources of data, their locations, the nature of the raw data and how they may be treated and expressed in a form that renders them appropriate for scientific analysis. Discussion also takes place of the database where the original and derived data can now be found and are freely available. Attention is here confined to UK logbooks; any attempt to review or make constructive comment on the worldwide collection in their various languages and dispersed archives being too large a task here to be undertaken.

Logbook studies are now commonplace and draw upon a vast reserve of documentary material.

However, most of the many thousands of surviving logbooks are from vessels in the service of the various national navies, of which those of the Royal Navy are the most abundant. From the earliest days these documents assumed the character of government papers and they were catalogued and preserved; those of the Royal Navy's captains and masters are now kept in the UK National Archives in Kew (<http://www.nationalarchives.gov.uk>). However, the logbooks of vessels negotiating the hazardous waters of the Arctic are mostly of a different provenance and principally constitute the logbooks of whaling vessels and those of the HBC. In the event, data were abstracted from all known 36 UK whaling logbooks and from some 200 HBC logbooks for the same period, the century from 1750.

There are consequently, few UK Arctic logbooks when set against the richness of items for the Atlantic and Indian Oceans. This is a reflection of the more mercantile nature of the source. Ship-owning companies and families had little need to preserve their logbooks and formalized systems such as those that were applied to Royal Navy logbooks were absent and their survival was all-too-often a matter of serendipity or family pride. Consequently, of the 10 000 or so voyages made by UK whalers into the Arctic region, few more than 200 logbooks survive (Brown *et al.*, 2008). These are scattered around local archives, often in whaling ports such as Dundee or Whitby, although the largest collection is to be found in the Hull History Centre (<http://www.hullhistorycentre.org.uk>). This matter of coverage is made more uncertain by the chequered history of UK whaling which, while beginning in the 17th century, only assumed importance in the late 18th as a result of Government bounties awarded to such enterprises (Credland, 1995).

The fortunes of the logbooks of ships in the service of the HBC was, however, more assured as this quasi-governmental organization, in a fashion similar to the English East India Company (Brohan *et al.*, 2012), was large enough to secure its own archive. The Company's ships sailed annually between London and its Hudson's Bay factories, and the logbooks are currently preserved in the Manitoba State Archives (<http://www.gov.mb.ca/chc/archives/hbca/>) with microfilm copies available at the UK National Archives. This collection embraces the period 1751–1880, with gaps only in the early 1840s (Ward and Wheeler, 2011). The Company had nevertheless been trading in the region for nearly a century by 1750 having received its Royal Charter in 1670. It should however be noted that all such ventures, whether whaling or trading, were confined to the summer months and, with only a few exceptions, the logbooks from the Arctic relate to that season alone.

This notable source of evidence has been the subject of earlier research (Moodie and Catchpole, 1975; Catchpole, 1985; Catchpole and Halpin, 1987; Catchpole, 1992 and Ball, 1985, 1992). These valuable undertakings were however confined to the conditions

in Hudson Bay itself and attention was not given to the situations in the Davis Straits or wider North Atlantic: areas of primary interest in the current undertaking.

Although this collection might be small, it assumes significance in respect of its regional coverage: the Arctic is a key driver in global climate, reflecting not only the warmings and coolings over the years, but serving also to exercise its own control on the wider global climate system in a vital two-way relationship. The anxieties about, and sensitivities of, this region are reviewed in Chapter 28 of the Intergovernmental Panel on Climate Change, 2014 AR5. Yet, and despite these often-reiterated concerns, it remains, at least from the 'historical' point of view, one of the least well-understood of all climatic regions.

1. Logbook studies in historical climatology

Historical climatology has developed over the past two decades and is now very much part of the wider field of climate change studies (reviewed in Jones and Bradley, 1992 and in Brázdil *et al.*, 2005, but good specific examples are seen in Garcia-Herrera *et al.*, 2008 and Leijonhufvud *et al.*, 2008). While natural proxy records using ice cores, preserved pollen, tree rings and marine and terrestrial sediment series have done much to shed light on past climates, it is only more recently that attention has turned to documentary sources. Such material is inevitably diverse and includes personal diaries, official, civic and ecclesiastical records, agricultural returns, estate papers and newspaper items (Lamb, 1982; Jones *et al.*, 2001 and Brázdil *et al.*, 2010). This diversity sets this source apart from the natural proxies and poses particular problems of homogenization and the expression of narrative accounts of the weather – instrumental records form no part of this article – into number. Different languages, often archaic, and different writers with contrasting reasons for committing their experiences to paper, offer a kaleidoscope of raw evidence that needs to be organized with notable caution. However, despite such reservations success has been achieved in this field, good examples being Pfister *et al.* (2006), Camuffo *et al.* (2010), Leijonhufvud *et al.* (2010), Brázdil *et al.* (2014) and Dobrovolný *et al.* (2010).

Ships' logbooks are now part of this evidence source for past climates (Wheeler and Garcia-Herrera, 2008) and those from the UK offer a number of advantages over the above-cited documents:

1. they survive in notable quantities, measured in their tens of thousands
2. they tend, within the limitations of time and the natural evolution of language, to employ a common vocabulary, reducing thereby the problem of homogenization between different documents and writers

- they cover the seas and oceans, thereby filling a notable 'void' represented by some two thirds of the planet!
- they provide observations at a daily, sometimes hourly, scale of resolution and are precisely dated.
- they extend, albeit with diminishing abundance, back to the mid-17th century and therefore embrace important climate episodes and events such as the closing years of the Little Ice Age, volcanic events such as Tambora and the Dalton Solar Minimum.

In recent times, Oliver and Kington (1970) singled logbooks out as of value for climate studies, while Lamb (1982) was an advocate of this source, but without realising its potential (Wheeler, 2014). While a number of efforts were made at the small scale to employ ships' logbooks before that time (Wheeler, 1985, 1987, 1991, 1995; Lamb, 1991), it was only with the EU-funded CLIWOC project (Climatological Database for the World's Oceans: 1750–1850: <http://pendientedemigracion.ucm.es/info/cliwoc/>) that a full appreciation of the source was offered and methods of handling and managing the wealth of weather information that they contain was firmly established (Wheeler *et al.*, 2006).

Logbooks record much information on the management and operation of the ship, but the daily weather records always consist in three basic categories:

- observations of wind force
- observations of wind direction
- notes on the weather of the day, with rain, snow, fog etc., all being dutifully recorded

Figure 1(a, b) are typical of logbook presentations for the period. All such observations are non-instrumental in character, with thermometers, barometers and anemometers becoming part of a vessel's standard equipment only towards the close of the 19th century following the International Conference held in Brussels in 1853 under the auspices of Matthew Maury. This is not the arena in which to rehearse the now well-established nature of logbook information and the means by which it was gathered, nor the motives for doing so. These can be found in the special edition of *Climatic Change* devoted to the CLIWOC project (*Climatic Change*, 2005).

Figure 2 indicates the coverage of logbook data from the CLIWOC project. Although CLIWOC used only some 6000 of 120 000 logbooks from the pre-instrumental period, the map reflects the available coverage: the gap for the Pacific Ocean represents a genuine lack of data for an area that did not form part of the arteries of communication for the Imperial European powers of the age – Great Britain, France, the Netherlands, Spain and Portugal. The other notable gaps are the polar region. For the Antarctic there was nothing from this period and the situation is irredeemable, but attention might be drawn to the availability of logbooks for the Arctic latitudes.

2. Arctic logbook data

One of the advantages of logbooks as a documentary source is their consistency of content, layout and

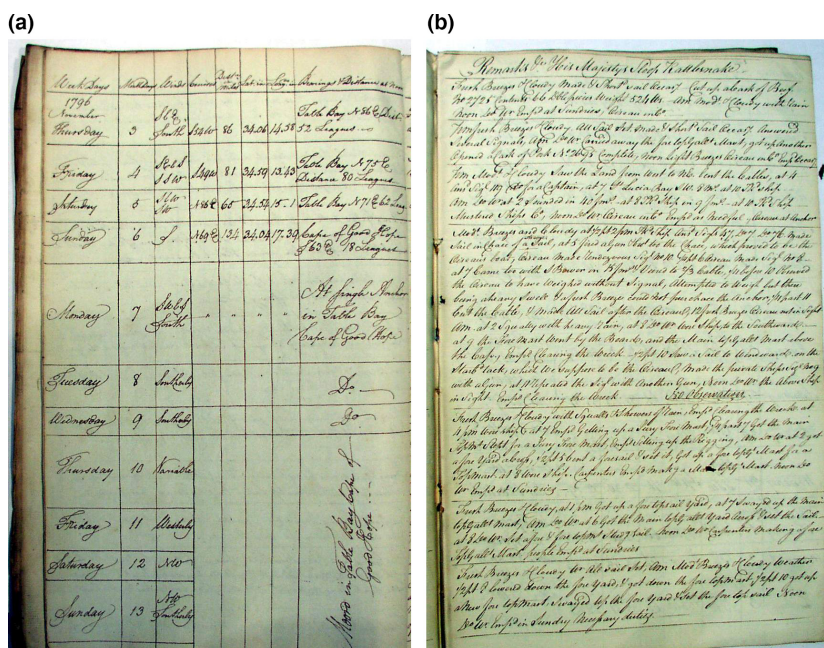


Figure 1. (a) Left-hand facing page of a Royal Navy logbooks (HMS *Rattlesnake* from 1797) showing dates and navigational information including wind direction. (b) Right-hand facing page of account of the day's activities including notes on wind force and weather. *Courtesy of the National Maritime Museum, Greenwich.*

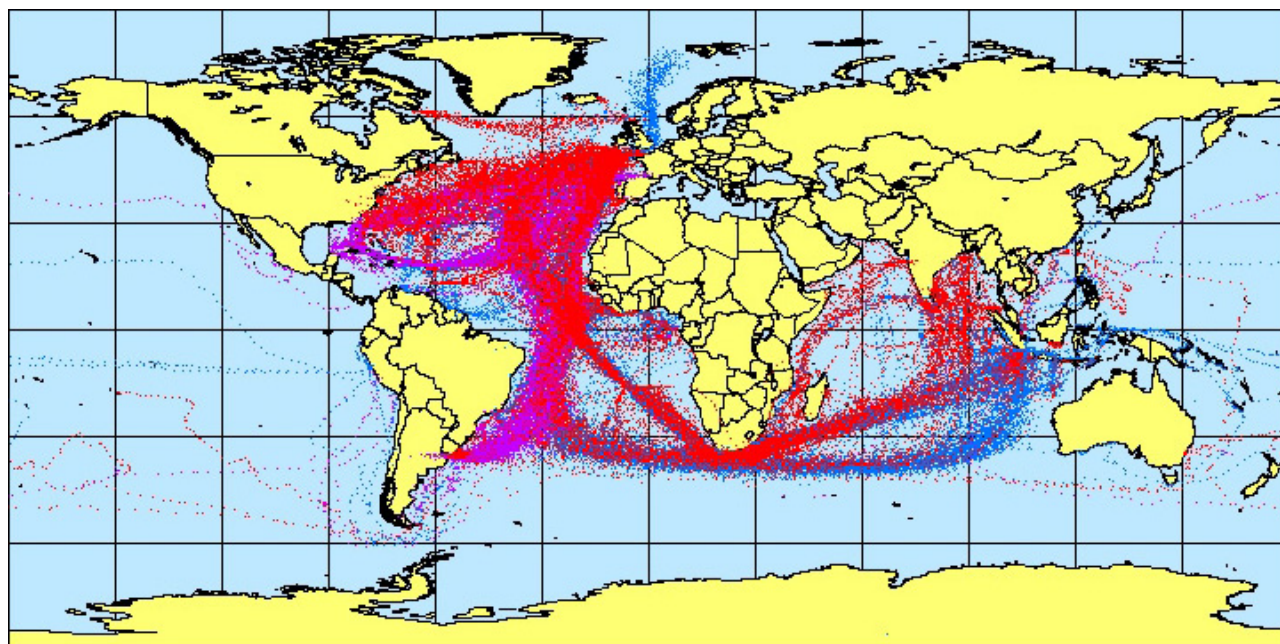


Figure 2. Data coverage from the CLIWOC project. Red: British vessels. Blue: Dutch and purple: Spanish.

vocabulary. Although there are minor exceptions, the captains, masters and officers of the various services – those of the Royal Navy, the East India Company, the HBC and the merchant service in general – shared a common vocabulary and indeed a common method in logbook preparation. However, they were not scientific records and were instead navigational documents in which the wind and weather were recorded as an aid to navigation; the wind and weather materially influencing the progress and direction of the vessels (Norie, 1889). Furthermore, these observations were made daily, sometimes hourly. However, this abundance of information requires cautious treatment, and the entries, while appearing to be both familiar (written in English), and reliable (the wind directions are recorded on a 32-point compass) they cannot be taken at 'face value'. The wind force terms in particular are couched in a vocabulary that is archaic. The Beaufort Scale, while proposed in 1806, was not adopted by the Royal Navy until the 1830s (Wheeler and Wilkinson, 2004). Before that time, the wind force vocabulary had no such official standing. That vocabulary was, however, one that was understood and employed by all practicing mariners who were trained in a common maritime oral tradition. There remains, however, the need to express such terms in Beaufort Scale equivalents. The situation is not improved when attention is turned to wind direction. This was recorded on the magnetic compass but to be of scientific value the observations need to be expressed in terms of true north. The difference between the two is known as 'variation'. In the high latitudes, variation can assume extreme values, often in the order of 40° or more. It is important, therefore, that such corrections are made, but the issue is complicated by the

fact that variation differs spatially and temporally in response to the behaviour of the Earth's core where the terrestrial magnetic field is generated (Jackson *et al.*, 2000). The problem of determining exactly, sometimes even approximately, where a ship was located is reviewed below.

The weather data are recorded for every day, and are helpfully located in time by the date – the principal observations being made at midday – and place – the latitudes and longitudes are often, but by no means always, given. Yet the latter locational data need also to be considered carefully. Latitude can be generally relied upon, its value being estimated by the altitude of the midday sun, but longitude is another matter. Its estimation provided a challenge to mariners, and one that was to constitute, arguably, the greatest intellectual challenge of the 18th century (Hewson, 1951) and for this reason, those estimates recorded in the logbooks must be regarded as approximate. Finally, although the officers on board HBC vessels assiduously recorded the latitude, longitude and variation each day, the masters of whaling vessels were far less attentive to such details and longitudes go unrecorded for many days, even weeks and variation was rarely noted. Hence, voyage reconstruction for whaling vessels presents a significant challenge.

There remains however one feature of 'Arctic' logbooks that distinguishes them from all others, and makes them of unique value, and that is their reference to sea ice coverage and to iceberg sightings. However, and as with the wind and weather observations, those for sea ice also require re-expression in terms that render them of value for scientific purposes.

3. Managing Arctic logbook data

In this case two approaches were made to the basic task of data abstraction. In all cases the hand-written documents forbid the use of OCR methods and abstraction is necessarily manual and time-consuming. The HBC records were abstracted from the microfilm copies in the UK National Archives. The same procedure was used for the whaling logbooks, with the exception that the Hull collection was imaged beforehand by ARCdoc, and the images then used for abstraction. Transference was made directly into 'working' spread sheets that were then processed using MS Excel options and procedures such as look up tables, but based on the methods to be described below.

3.1. Wind force

The vocabulary for wind force in the pre-Beaufort era is known to have been conventional and consistent. Assuredly there were more terms in use than the current 13 that suffice for the Beaufort Scale, but almost all of them can be re-expressed as one of those 13. This re-expression can be executed using the wind force dictionary prepared by the CLIWOC project team (Garcia-Herrera *et al.*, 2003) and employing the look up tables of the latter for automatic conversion.

3.2. Wind direction

Wind directions were, as noted, recorded with regard to magnetic rather than true north. To render the wind directions suitable for modern climate analysis, these observations need to be converted to true north taking into account the ship's location. A number of historical magnetic variation databases exist but use was made in ARCdoc of data held by the British Geological Survey.¹ The known variations at specified locations and dates could then be applied to each of the wind direction records to reduce them to true north bearings. HBC logbooks tended to have variation recorded on a daily basis but for whaling logbook conversions such a database was vital as they generally failed to note variation. This operation was carried out separately, the revised directions then being entered into the spreadsheets.

3.3. Weather records

These are the various references to snow, rain, fog, cloud cover, etc. In contrast to the specifically maritime character of the wind force terms, the descriptions of the weather used 'standard' English of

¹The ARCdoc team express their collective and sincere thanks to Susan Macmillan of the BGS who with efficiency and dependability provided the variation data required for this project.

a form familiar to landsmen and seafarers alike. No need was evident for any translation and the terms could be entered directly.

3.4. Latitude and longitude

Observations of latitude can, for the most part, be relied upon and require no conversion. Observations of longitude present a challenge because of its unreliability or absence. Longitude estimates must all be regarded as approximate but absence is another matter. Linear interpolation between known points is a partial solution to the problem and can be achieved by carefully plotting the voyage using observations of the course, distances sailed in the day and landfall sightings onto marine charts. It should be noted, however, that this process cannot be automated and can, therefore, be a time-consuming task. Finally, any longitudes noted in the logbooks need to be adjusted from the local to the Greenwich zero meridian. This is a simple calculation of the difference in degrees between the adopted meridian for that leg of the voyage – which is either given or can be easily derived from the logbook records – and Greenwich zero meridian. Fortunately, the whaling captains preferred to use Greenwich as their prime meridian, and while the HBC officers rarely did so, their logbooks contain sufficient information to convert the chosen meridians into Greenwich equivalent values.

3.5. Date

Attention is also needed to be given to the dates. Before September 1752, England continued with the Julian calendar; only in that month was the more reliable Gregorian calendar adopted. The 11-day time difference between the two needed to be taken into account for all those items in which the Julian calendar was employed. It should also be noted that the nautical day, used exclusively while at sea, began at midday and 12 hours ahead of the civil day (Harries, 1928 and Cotter, 1978).

4. Index derivation: narrative into number

The above procedures allow for the preparation of reliable daily series of wind and weather observations. These are, however, non-instrumental and non-numerical in form and, additionally, require aggregation into monthly summaries, a task that provides opportunity for presenting the original data in 'index' form on an ordinal scale. The nature of logbook entries, in particular their inherent homogeneity, renders such tasks easier than in the case of the more idiosyncratic character of other documentary source material. It is, as a result of this characteristic, possible to go from one logbook to another over the study period confident that terms have the same meaning and the observations made to a consistent standard.

The monthly aggregated indices derived from the Arctic logbooks are essentially simple in form, but not lacking scientific value for that simplicity. Attention is focused on the HBC set as they allow for the construction of an almost unbroken century-long series, but the same methods apply to the whaling logbooks although the resulting series are more fragmentary.

4.1. Wind force

The re-expression of wind force terms in Beaufort Force equivalents allows for two indices of wind strength to be derived:

1. the mean wind force for each month
2. a gale index, based on the proportion (a scale of 0–1.0) of days each month when wind force equalled or exceeded gale force on the Beaufort Scale

4.2. Wind direction

Wind directions, originally on a 32-point compass, can be reduced to the four cardinal quadrants (N, E, W and S). This results in a 'loss of information' as, for example, all winds registered from between NWbN and NE are rendered as being N (northerly) sector. Such loss cannot be quantified but the conversion process provides a dataset of well-populated categories suitable for later analysis and in a form not wholly dissimilar to the Lamb weather types (Lamb, 1950). An index can be derived for each of the four categories based on the proportion of days each month that the wind is from that quarter. Most useful perhaps would be the W index as it might be expected to relate to similar circulation indices, such as the Arctic and North Atlantic Oscillations offering another, and independent, means of expressing such important climatic behaviour.

4.3. Weather indices

Weather indices can also be derived for those weather elements that found their way most often and reliably into the logbooks, of these, fog, snow and rain are the best examples. Each day can be seen to be a day of fog, snow or rain depending on their absence or presence in the written record. Again, the three indices are proportions of days each month when these phenomena were recorded. In the case of the ARCdoc project no attempt was made to qualify or quantify the phenomena further by reference to attached adjectives such as 'heavy', 'thick', 'strong' etc. Expressed otherwise, a 'rainy day' is classified as such irrespective of whether it be heavy, light or any other form of rain. Table 1 summarizes the monthly indices that were derived from the logbooks.

One of the virtues of a daily record spanning, as the HBC datasets do, a century, is that it allows the climate record to be reconstructed casting a fresh light on the changing conditions of a time. Figures 3–5

Table 1. Monthly indices derived from logbooks.

Index	Definition
N	Proportion of observed days with wind from that quarter
S	Proportion of observed days with wind from that quarter
E	Proportion of observed days with wind from that quarter
W	Proportion of observed days with wind from that quarter
Gale	Proportion of observed days with wind of force nine or greater
Snow	Proportion of observed days with falling snow
Rain	Proportion of observed days with some form of rain
Fog	Proportion of observed days with fog
Wind force	Mean Beaufort force

These monthly indices (weighted by number of days of record) can be additionally aggregated to produce the annual indices used in figures 3 to 5.

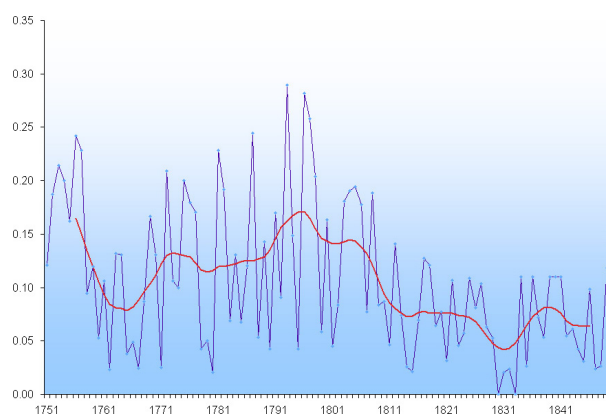


Figure 3. The gales index derived from HBC logbooks and based on annually aggregated figures. The running mean is a Gaussian filtered 12-year window.

illustrate the benefits from this exercise using samples selected from the full range of indices described in Table 1. The figures use annually aggregated data but, as noted above, the data are for the summer season only; the vessels not being active in this area during the winter. They reveal, and possibly for the first time, the detailed changing climate – albeit expressed through the medium of selected phenomena – of the Arctic and far North Atlantic region at this time. Their value is that they prompt a number of important questions to direct future studies, for example:

1. Why does the gales index show the steady decline through the first half of the 19th century?
2. What might account for the abrupt increase in snow fall in the first decades of the 19th century?
3. Why, in light of the above changes, does the W index remain more constant?

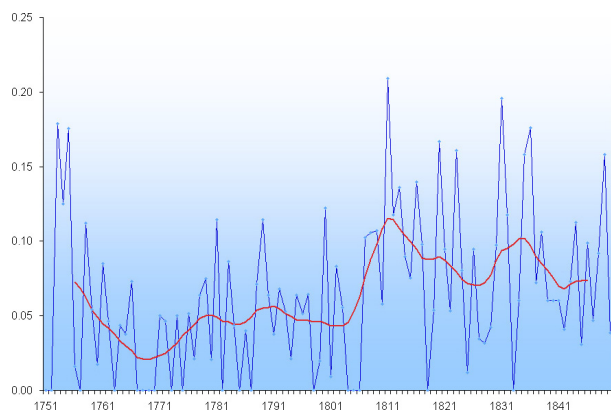


Figure 4. The snow index derived from HBC logbooks and based on annually aggregated figures. The running mean is a Gaussian filtered 12-year window.

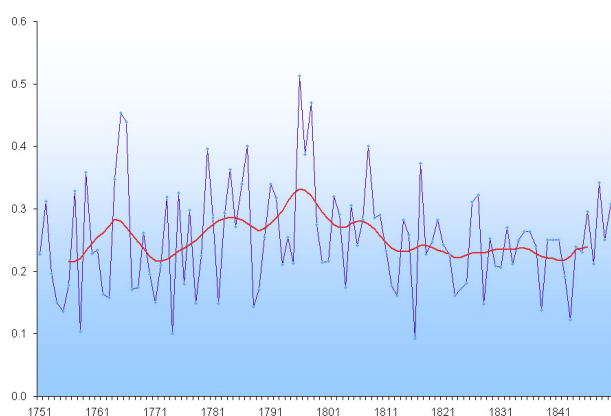


Figure 5. The W index derived from HBC logbooks and based on annually aggregated figures. The running mean is a Gaussian filtered 12-year window.

This is not the arena in which to answer such profound questions, but the illumination yielded by the data offers an indication of the potential of such logbook collections when the data are properly treated.

5. Sea ice and iceberg data methods

Arctic logbooks are very much distinguished by their valuable records of sea ice cover and iceberg incidence. Again, however, these observations cannot be used without some careful treatment.

5.1. Sea ice observations

It is this category that distinguishes the Arctic logbooks, and a number of sea ice indices can be derived. First, a simple proportion of days each month where sea ice is recorded can be used to measure the time vessels spent in proximity to the sea ice and, more importantly, the latter's position. Second, the vocabulary by which ice types are described is comparable to that of wind force, with the sailors distinguishing between different coverage levels and thicknesses

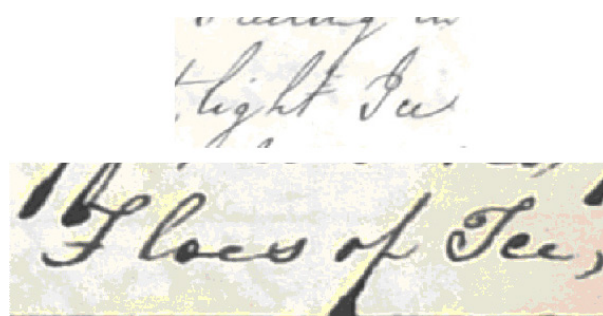


Figure 6. An example of a thickness (age) term 'light ice', index value of 3, representing ice of up to 1m in thickness (first year ice) and an example of a coverage term 'Floes', also with an index value of 3, representing ice coverage of 50–74%.

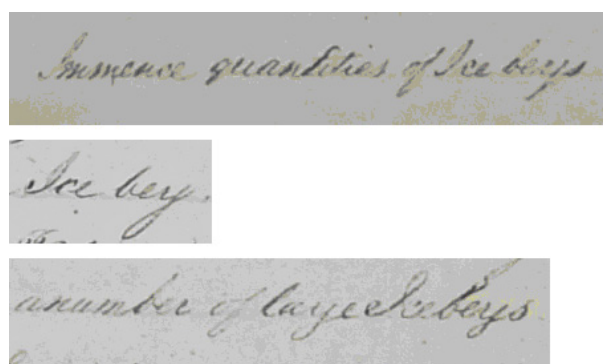


Figure 7. An example of the progression in the narrative description of iceberg numbers taken from the logbook of the Cumbrian of Hull 1827. Ranging from a single 'iceberg' (indexed as '1') to 'a number of large icebergs' (indexed as '10') to 'immense quantities of icebergs' (indexed as '100').

(or age) of sea ice (Figure 6). Work by ARCdoc members provided the basis for both a coverage and thickness index based on this, again consistent, vocabulary. Coverage was categorized into a six-point index with 1–5 denoting increasing coverage through 20% stages. Age contains a similar five-point index with 1–4 denoting increasing thickness and 5 brash (decaying) ice.

5.2. Iceberg observations

The observations also include iceberg sightings and the mariners were careful to differentiate these from sea ice. Descriptions of numbers of bergs observed are consistent with the narrative nature of these logbooks in the sense they rarely include numerical counts and the authors instead resort to terms such as 'one', 'many', 'an immense quantity', etc. (Figure 7). It is, however, again important to express these descriptions on an index scale.

Indexing on the basis of iceberg records required, however, a degree of subjectivity not encountered in the other areas. Most importantly, a simple count of

the number of days when icebergs are observed is not representative of the number sighted each year. Expressed in simple terms, 1 year may contain ten observations of a 'single iceberg' whereas another may contain a single observation of 'thousands of icebergs'. On this basis there would be a tendency to over-estimate the frequency or number of icebergs. In order to secure a more representative estimation of the relative number of icebergs, the terms employed by the whalers had to be first categorized. It was decided that a fourfold categorization would offer an optimum system in which indexing could take place, not only allowing on the one hand for numerical expression of narrative accounts but also taking into consideration the unavoidable vagueness of those accounts. The iceberg records have here been indexed on a logarithmic scale as summarized in Table 2. Such a system was preferred in order not to misrepresent the inferred monthly and annual counts on the basis of the concerns expressed earlier. Despite this large number of entries, the descriptions seem not to have been random and the terms used are embraced by just 22 such descriptors. Table 2 lists these descriptors and indicates how they were decided to equate to the

Table 2. The 22 individual descriptions encountered in the logbooks and their respective indices.

	Term	Representation
1	Berg	Individual iceberg
10	A number of icebergs Good few icebergs A few bergs Many bergs Good number of bergs Quantity of icebergs Cluster of bergs Numerous bergs Several bergs	Descriptions representing 10s of icebergs (>1, <100)
100	Great many icebergs Great quantity of bergs Great number of icebergs Vast number of bergs Numerous bergs Hundreds of icebergs Several hundred icebergs Very great number of bergs Very large number of bergs Immense amount of bergs	Descriptions representing 100s of icebergs (>100, <1000)
1000	Bergs in thousands Bergs innumerable	Descriptions representing 1000s of icebergs (>1000)

These descriptions were abstracted from the logbooks of whaling vessels in the Davis Strait and Baffin Bay 1809–1898.

number scale. Unlike wind force or sea ice, which are both highly dynamic and variable phenomena, icebergs are persistent, slow-moving and slowly evolving entities. In consequence of this and of the poor quality of locational data, observations with the same index value, occurring on consecutive days were counted as one observation. Moreover, if such consecutive observations span a change in month, the index will be equally split proportionally over the two monthly totals.

6. Data storage and retrieval: making data available

All of the above data, both in its raw and processed form are now available online at <http://www.hull.ac.uk/mhsc/ARCDoc/>. In this regard, not atypically, the ARCDoc project has been faced with searching data management issues that have been recognized, but are often perceived as being outside of the standard remit of academic research. Taking a proactive approach to data management, the team has engaged with the growing recognition of the need to share climate data and information. This recognition, while requiring the assembling of reliable and validated data, is one step along the path to developing, implementing and maintaining a robust data management strategy (Kuhn, 1962). The process of developing a data storage and retrieval solution which encompasses provable standards of robustness, integrity and future proofing requires specialist skills. These specialist skills are increasingly falling within the remit of the academic research arena. From this perspective, the ARCDoc project has adopted a clear data management strategy based on a data management plan, informed by a research lifecycle model developed along gestalt principles (Perls *et al.*, 1951) and, in practical terms, accepted as an industry standard FEDORA base repository: HYDRA.²

The submission of data has been followed by a rigorous process of normalization, formatting and standardizing to provide a series of datasets that are reliable and preservable (Data Curation Centre, 2013) in the HYDRA repository (<https://hydra.hull.ac.uk/>). This process included a basic analysis of the data structures, removal of repeated fields, standardization of formats and the vital inclusion of accompanying documentation to inform the data. The result – a

²The Hydra digital data repository has been developed in conjunction with the University of Hull (UK), Stanford University (USA) and the University of Virginia (USA), together with partner institutions from around the world (<http://projecthydra.org/>). It adheres to internationally accepted standards of data preservation and integrity, and incorporates high level searchability and confidentiality standards. The Hull Hydra repository undertakes to provide a minimum 10 year after last use data preservation policy to ensure continuity and preservation.

series of datasets – is published, accessible within the public domain and permanently maintained.

Among other considerations, the prospect of versioning has been broached. Original data submissions, typically in a standard spreadsheet, flat file tabular mode, are preserved in their original formats. Subsequent aggregated data based on the methods reviewed above reflect clearer, presentation-friendly and searchable datasets, but do not obscure the literal content of the original data. The final outcome is *information* (data informed by accompanying documentation and formatted for general use) that is publically available via a robust, online repository.

The ARCdoc project has recognized the importance of data management and adopted a strategy for data storage and retrieval that not only presents the final aggregated dataset, but also includes the stages in data processing from the raw, unadjusted daily data, through the second stage of the managed daily data, leading to the aggregated monthly figures described earlier. By this means, current and future workers can trace the development of the database from its inception to its conclusion.

7. Conclusion

Ships' logbook data are now a well-established part of the evidence base for climate change, but much remains to be accomplished; only a small fraction of those that are known to exist worldwide have been examined, and many more, doubtless, remain to be discovered in archives and collections around the world and attention has here be confined exclusively to the UK sources. The principle activity must be directed towards the onerous task of data abstraction. This can only be achieved by manual methods, as no OCR system exists to allow for automated abstraction. However, excellent work in this direction is being undertaken by the Old Weather project (<http://oldweather.org>)³ based on methods of 'crowd sourcing'. Yet, the mere abstraction of the beguilingly vast quantity of data is not of itself sufficient. As this article has made clear, much needs to be done to prepare the raw data for scientific use. While some of the issues introduced here are unique to UK Arctic logbooks, others have a more general application that embraces all such pre-instrumental sources. The uses to which these data can be put will do much to shed a fresh light on past climates, especially of the oceanic regions and applications such as those described by Küttel *et al.* (2009) may become more frequent and more informative.

³The Old Weather team are always pleased to enlist the participation of new members and enquires should be directed to the website.

Acknowledgements

The authors thank the Leverhulme Trust who funded the ARCdoc project, upon which much of this article is based. We also acknowledge the support of the Durham University Alumni Fund who sponsored a project for the abstraction of iceberg records from UK whaling logbooks. The authors also thank the reviewers for their helpful guidance.

References

- Ball TF. 1985. A dramatic change in the general circulation on the west coast of Hudson Bay in 1760 AD: synoptic evidence based on historic evidence. In *Syllogeus*, vol. **55**, Harington CR (ed). National Museum of Canada: Toronto; 219–228.
- Ball TF. 1992. Historical and instrumental evidence of climate: western Hudson Bay, Canada. In *Climate since AD 1500*, Jones PD, Bradley RS (eds). Routledge: London; 40–73.
- Brázdil R, Pfister C, Wanner H, von Storch H, Luterbacher J. 2005. Historical climatology in Europe – the state of the art. *Climatic Change* **70**: 363–430.
- Brázdil R, Wheeler D, Pfister C. 2010. European climate of the past 500 years based on documentary and instrumental data – Editorial. *Climatic Change* **101**: 1–6.
- Brázdil R, Kotyza O, Dobrovolny P, Reznicková L, Valášek H. 2014. *Climate of the sixteenth century in the Czech lands*. Masaryk University: Brno, the Czech Republic.
- Brohan P, Allan R, Freeman E, Wheeler D, Wilkinson C, Williamson F. 2012. Constraining the temperature history of the past millennium using early instrumental observations. *Climate of the Past* **8**: 1551–1563.
- Brown S, Credland A, Savours A, Stonehouse B. 2008. British Arctic whaling logbooks and journals: a provisional listing. *Polar Record* **44**: 1–10.
- Camuffo D, Bertolin C, Barriendos M, Dominguez-Castro F, Cocheo C, Enzi S, Sghedoni M, Garnier E, Alcoforado MJ, Xoplaki E, Luterbacher J, Diadato N, Maugeri M, Nunes MF, and Rodriguez R. 2010. 500-year temperature reconstructions in the Mediterranean Basin by means of documentary data and instrumental observations. *Change Climate* **101**: 169–199.
- Catchpole AJW. 1985. Evidence from Hudson Bay region of severe cold in the summer of 1816. In *Syllogeus*, vol. **55**, Harington CR (ed). National Museum of Canada: Toronto; 121–146.
- Catchpole AJW. 1992. Hudson's Bay Company ships' logbooks as sources of sea ice data. In *Climate since AD 1500*. Jones PD, Bradley RS (ed). Routledge: London; 1751–1870.
- Catchpole AJW and Halpin J. 1987. Measuring summer sea ice severity in eastern Hudson Bay, 1751 to 1870. *Canadian Geographer* **31**: 233–244.
- Climatic Change. 2005. Special Issue – CLIWOC: climatology of the world's oceans 1750–1850. *Climatic Change* **73** Nos 1 and 2.
- Cotter CH. 1978. Early dead reckoning navigation. *Journal of Institute of Navigation* **31**: 20–26.

- Credland AG. 1995. *The Hull Whaling Trade: an Arctic enterprise*. Hutton Press: Beverley.
- Data Curation Centre. 2013. *DCC Because good research needs good data*. <http://www.dcc.ac.uk/> (accessed 10 July 2013).
- Dobrovolný P, Moberg A, Brázdil R, Pfister C, Glaser R, Wilson R, van Engelen A, Limanówka D, Kiss A, Halíčková M, Macková J, Riemann D, Luterbacher J, Böhm R. 2010. Monthly, seasonal and annual temperature reconstructions for Central Europe derived from documentary evidence and instrumental records since AD 1500. *Climatic Change* **101**: 69–107.
- Garcia-Herrera R, Wheeler D, Koek F, Wilkinson C, Können G, Rosario Prieto M. 2003. *CLIWOC multi-lingual meteorological dictionary*. KNMI Publication 205: De Bilt, The Netherlands.
- Garcia-Herrera R, Diaz HF, Garcia RR, Rosario Prieto M, Barriopedro D, Moyano R, Hernández E. 2008. A chronology of El Niño events from primary documentary sources in northern Peru. *Journal of Climate* **21**: 1948–1962.
- Harries H. 1928. Nautical time. *Mariner's Mirror* 364–370.
- Hewson JB. 1951. *A History of the Practice of Navigation*. Brown Son and Ferguson: Glasgow.
- Intergovernmental Panel on Climate Change. 2014. *Assessment Report 5, Working Group II. Chapter 28: Polar Regions*. IPCC: Geneva, Switzerland.
- Jackson A, Jonkers AR, Walker MR. 2000. Four centuries of geomagnetic secular variation from historical records. *Philosophical Transactions of the Royal Society of London (A)* **358**: 957–990.
- Jones PD, Bradley RS. 1992. Climate since AD 1500: Introduction. In *Climate since AD 1500*, Jones PD and Bradley RS (eds). Routledge, London: 1–16.
- Jones PD, Ogilvie AEJ, Davies TD, Briffa KR. 2001. Unlocking the doors to the past; recent developments in climate and climate-impact research. In *History and Climate: memories of the future?*, Jones PD, Ogilvie AEJ, Davies TD, Briffa K (eds). Kluwer: New York; 1–8.
- Kuhn TS. 1962. *The structure of scientific revolutions*. University of Chicago Press: London and Chicago.
- Küttel M, Xoplaki E, Gallego D, Luterbacher J, Garcia-Herrera R, Allan R, Barriendos M, Jones PD, Wheeler D, Wanner H. 2009. The importance of ship log data: reconstructing North Atlantic, European and Mediterranean sea level pressure fields back to 1750. *Climate Dynamics* **34**: 1115–1128.
- Lamb HH. 1950. Types and spells of weather around the year in the British Isles. *Quarterly Journal of the Royal Meteorological Society* **76**: 393–438.
- Lamb HH. 1982. *Climate, History and the Modern World*. Methuen: London.
- Lamb HH. 1991. *Historic Storms of the North Sea*. Cambridge University Press, Cambridge: British Isles and Northwest Europe.
- Leijonhufvud L, Wilson R, Moberg A. 2008. Documentary data provide evidence of Stockholm average winter to spring temperatures in the eighteenth and nineteenth centuries. *Holocene* **18**: 333–343.
- Leijonhufvud L, Wilson R, Moberg A, Söderburg J, Retsö D, Söderland U. 2010. Five centuries of Stockholm winter/spring temperatures reconstructed from documentary sources and instrumental observations. *Change Climate* **101**: 109–141.
- Moodie DW, Catchpole AJW. 1975. Environmental data from historical documents by content analysis. Manitoba Geographical Studies, 5, Department of Geography, University of Manitoba.
- Norie JW. 1889. *A Complete Epitome of Practical Navigation*. Norie and Wilson: London.
- Oliver J, Kington JA. 1970. The usefulness of ships' logbooks in the synoptic analysis of past climates. *Weather* **25**: 520–527.
- Perls FS, Hefferline RF, Goodman P. 1951. *Gestalt therapy: excitement and growth in the human personality*. N.Y, Julian Pr.
- Pfister C, Weingartner R, and Luterbacher J. 2006. Hydrological winter droughts over the last 450 years in the upper Rhine Basin: a methodological approach. *Hydrological Sciences Journal* **51**: 996–985.
- Ward C, Wheeler D. 2011. Hudson's Bay Company ship's logbooks: a source of far North Atlantic weather data. *Polar Record* **48**: 165–176.
- Wheeler D. 1985. The Weather at the Battle of Trafalgar. *Weather* **40**: 338–346.
- Wheeler D. 1987. The Trafalgar Storm: 22–29 October 1805. *Meteorological Magazine* **116**: 197–205.
- Wheeler D. 1991. The influence of the weather during the Camperdown Campaign of 1797. *Mariner's Mirror* **77**: 47–54.
- Wheeler D. 1995. A climatic reconstruction of the Battle of Quiberon Bay, 20 November 1759. *Weather* **50**: 230–238.
- Wheeler D. 2014. Hubert Lamb's 'treasure trove': ships' logbooks in climate research. *Weather* **69**: 133–139.
- Wheeler R, Garcia-Herrera. 2008. Ships' logbooks in climatological research: reflections and prospects. *Annals of the New York Academy of Science special publication "Trends and Directions in Climate Research"* **1146**: 1–15.
- Wheeler D, Wilkinson C. 2004. From calm to storm: the origins of the Beaufort Wind Scale. *Mariner's Mirror* **90**: 187–201.
- Wheeler D, Garcia-Herrera R, Koek F, Wilkinson C, Können G, Rosario Prieto M, Jones PD, Casale R. 2006. CLIWOC: climatological database for the world's oceans 1750 to 1850. European Commission: Brussels.