Gridded Monthly Sea Ice Extent and Concentration, 1850 Onwards, Version 1.1

Data set citation

Walsh, J. E., W. L. Chapman, and F. Fetterer. 2015, updated 2016. Gridded monthly sea ice extent and concentration, 1850 onwards, Version 1.1. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media. http://dx.doi.org/10.7265/N5833PZ5.

Table of Contents

1 Summary

Climate models and reanalyses often require gridded pan-Arctic sea ice data sets that cover a long period of time, on the order of a century. However, observation-based data sets that predate the satellite era are heterogeneous in information content and in format and are, therefore, difficult to use for a long-term data record. Here, observations from historical sources are the basis of a monthly gridded sea ice concentration product that begins in 1850. The historical observations come in many forms: ship observations, compilations by naval oceanographers, analyses by national ice services, and others. In 1979, these sources give way to a single source: concentration from satellite passive microwave data. In the 1990s, a sea ice concentration product was published that combined many of these sources and extended it with satellite data (Chapman and Walsh 1991). *Gridded Monthly Sea Ice Extent and Concentration, 1850 Onwards* builds on that earlier data product by adding historical sources, refining the technique used to merge data from different sources, and extending the record backward and forward in time. The data are provided in a NetCDF-4 file with monthly sea ice concentration in a variable given on a quarter-degree latitude by quarter-degree longitude grid. In addition to the concentration variable, a corresponding source variable indicates where each of 16 possible sources is used.

2 Modeling applications for these data

The data product described here represents a bridge between historical sea ice information and model applications in which lower boundary conditions, such as sea ice concentration and sea surface temperatures, are required. These model applications include atmospheric reanalyses and atmosphere-only simulations such as those used in the Atmospheric Model Intercomparison Project (AMIP), as well as validation for coupled atmosphere-ocean-ice simulations. For example, the project is linked to the HadISST data set used to drive the climate models of the Hadley Centre and other modeling groups. An earlier version of the data set described here provided input to HadISST1 in the 1990s and early 2000s (Chapman and Walsh 1991), and the present data set represents an enhancement of the sea ice database available for climate modeling projects requiring sea ice information from the pre-satellite period. Additional applications include the potential use of the present data set to drive historical atmospheric reconstructions (reanalyses) such as the 20th-Century Reanalysis at NOAA's Earth System Research Laboratory.

3 Product development background

As noted in the Summary, reanalysis and climate diagnostic applications often require gridded pan-Arctic sea ice data sets that span 100 years or more; yet, sea ice information is incomplete and heterogeneous prior to the satellite era. The data set described here, hereafter referred to as SIBT1850 for "sea ice back to 1850," starts with, extends, and improves upon one that begins in 1901 called *Arctic and Southern Ocean Sea Ice Concentrations* (Chapman and Walsh 1991), hereafter referred to as CW91. CW91 is available from NSIDC and described in Walsh and Chapman (2001). SIBT1850 improves upon CW91 by using newly available sea ice observations for years prior to 1953, while extending the record in time. Some of the newly added data have been published as separately available data products (e.g. Underhill, Fetterer, and Petersen 2014).

Sea ice area from SIBT1850 is plotted by month in Figure 1. In contrast to similar plots that can be made with the CW91 data, there are no "flat line" stretches indicative of the use of a constant climatology.

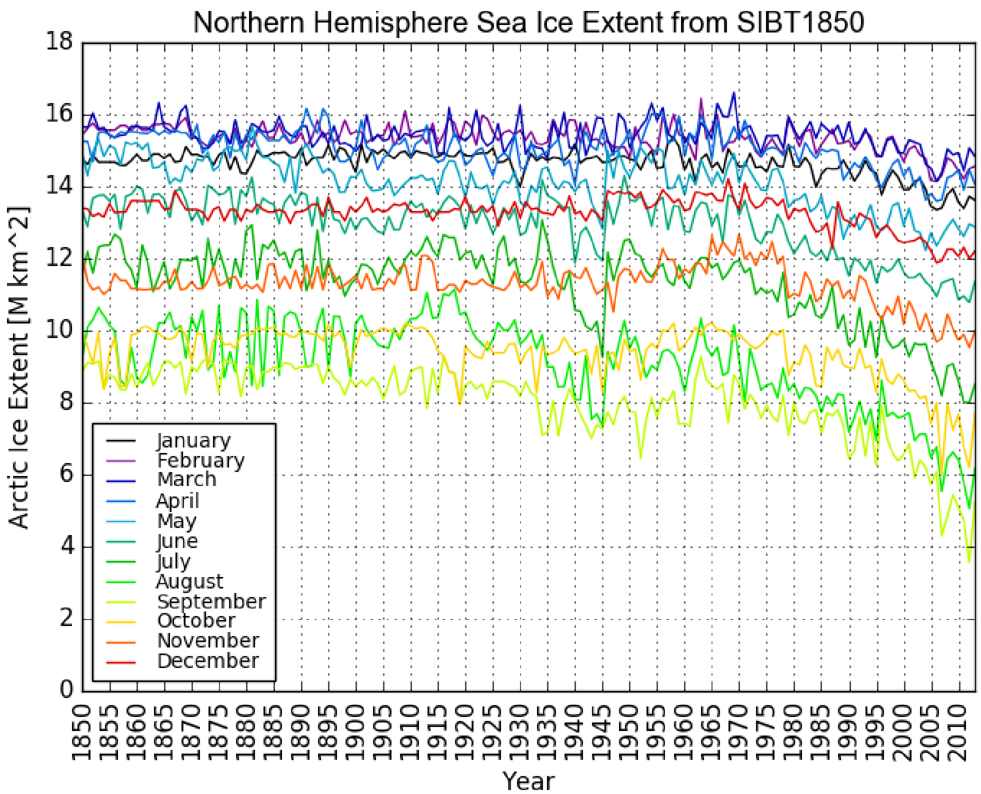

Figure 1. The extent, in millions of square km, covered by sea ice present at any concentration plotted for each calendar month, from 1850 through 2013.

CW91 spans 1901 through 1995. In constructing CW91, many of the data values were filled using an ice-concentration climatology. That monthly climatology, noted in Walsh and Chapman (2001), was derived using the concentration fields of the data described in Walsh and Johnson (1979).

SIBT1850 improves upon CW91 by using newly available sea ice observations for years prior to 1953. These add variability to the early part of the record that had in many instances been fixed at climatological values.

SIBT1850 spans 1850 through the most recent update. This span consists of four subsets that have differing data sources, as follows:

- **1850-1900:** Newly available data sources, with gaps filled by the analog method described in the Filling data gaps with an analog method section.
- **1901-1952:** Concentration values from CW91, with the addition of new data that overwrite CW91 values. Where new data are not available to overwrite a climatological grid cell value from CW91, the SIBT1850 grid cell is filled with a value that comes from gap filling using an analog method. See the Filling data gaps with an analog method section.
- **1953-1978:** Concentration values from CW91, with the addition of new data that overwrite CW91 values. This period is distinguished from 1901-1952, however, because beginning in 1953 there is an essentially continuous arctic-wide data record, thanks to sources cited in Walsh and Johnson (1979).
- **1979-latest update:** All concentration values are from satellite passive microwave data.

In these initial versions of SIBT1850, the focus is on improving the early part of the observation-based record for reanalysis and modeling. A later version is planned that will offer an alternate post-1978 record that incorporates newly available data sources for this part of the record as well.

John Walsh, University of Alaska, Fairbanks (UAF), led the development of this data set. Bill Chapman, University of Illinois Urbana-Champaign (UIUC), was responsible for the processing code. Florence Fetterer, NSIDC, facilitated the acquisition and use of source data and led the development of data derived from the Danish Meteorological Institute (DMI) charts. This work was a collaboration, in part, with work at UAF that produced the Historical Sea Ice Atlas for Alaska.

4 Format and data file description

Ice concentration data are in a single NetCDF-4 file, G10010_SIBT1850.nc, with a concentration variable and a source variable for each month beginning with January 1850. Data are on a quarter-degree latitude by quarter-degree longitude grid. The grid was chosen to be compatible with certain products developed at the Hadley Center, UK. Table 1 describes the variables in the NetCDF file.

Table 1. NetCDF Variable Description

<table>
<tr><th>Variable</th><th>Description</th></tr>
<tr><td>latitude</td><td>Latitude in degrees</td></tr>
<tr><td>longitude</td><td>Longitude in degrees (0 to 360)</td></tr>
<tr><td>seaice_conc</td><td>Sea ice concentration in percent with values from 0 to 100, inclusive. Land is indicated by -1.</td></tr>
<tr><td>seaice_source</td><td>Describes the source of the data for each month of data beginning with January 1850. The variable has values of 1 through 21 described in the table below. Numbers 6, 7, 14, 15, and 19 are not used but are retained as placeholders for the processing code.
<table>
<tr><td>Value</td><td>Data Source</td><td>Value</td><td>Data Source</td></tr>
<tr><td>1</td><td>DMI yearbook narrative</td><td>11</td><td>Hill</td></tr>
<tr><td>2</td><td>JMA charts</td><td>12</td><td>Dehn</td></tr>
<tr><td>3</td><td>NAVO yearbooks</td><td>13</td><td>Danish Meteorological Institute</td></tr>
<tr><td>4</td><td>Kelly ice extent grids</td><td>16</td><td>Whaling log books open water</td></tr>
<tr><td>5</td><td>Walsh and Johnson</td><td>17</td><td>Whaling log books partial sea ice</td></tr>
<tr><td>8</td><td>Satellite passive microwave</td><td>18</td><td>Whaling log books sea ice covered</td></tr>
<tr><td>9</td><td>ACSYS</td><td>20</td><td>Analog filling of spatial gaps</td></tr>
<tr><td>10</td><td>AARI</td><td>21</td><td>Analog filling of temporal gap</td></tr>
</table>
</td></tr>
<tr><td>time</td><td>Time of the data observation in days since 2001-01-01 00:00:00.</td></tr>
</table>

The figures in this documentation that illustrate source and concentration fields from the NetCDF file were created using the NASA Panoply data viewer with custom color tables that are provided as part of this data product. For information on obtaining Panoply and using it with this data product, see the section on Exploring the data with Panoply.

Image files with names like montage09.gif are included as well, where the number in the file name indicates the data source that is identified in the image. Data source values are listed in Table 1. Each image is comprised of smaller images taking up 12 columns, one for each month, and 164 rows, one for each year (Figure 2). In these, areas where the source is used appear in yellow. These provide a quick way to view the coverage of various sources in the data set. They are included for illustration purposes and for the convenience of the user; they are not meant for analysis.

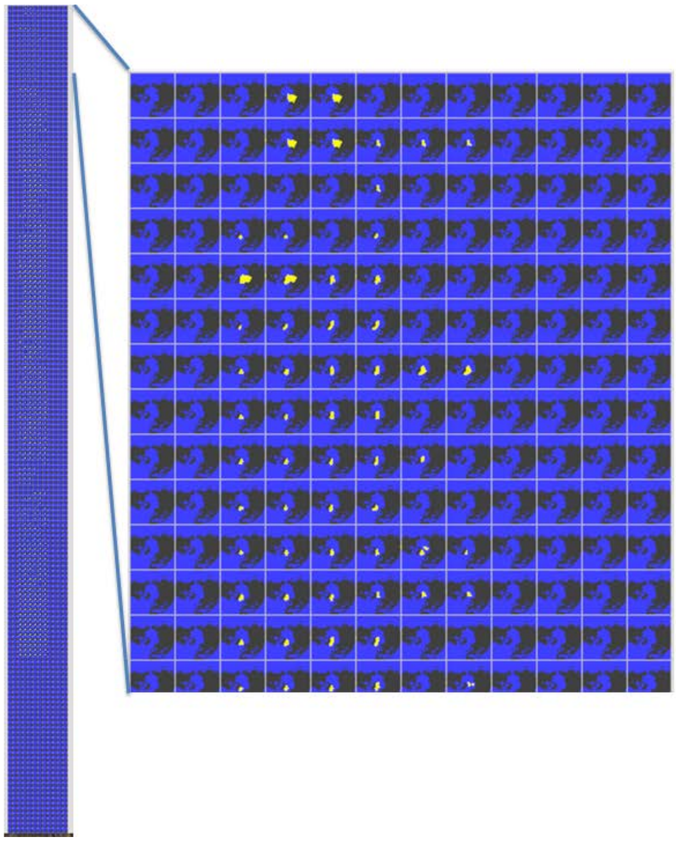

Figure 2. An example of a montage image file, with the first few years shown at right. The montage09.gif file is for source 9. Yellow appears in the first row, corresponding to 1850, and for most of the data set until the satellite era starts in 1978, but only in the central columns corresponding to the summer months.

Table 2 provides a description of the four files or file types that make up this data product, along with the naming convention, format, and size for each. All the data files are provided in a single zip file at the following location: ftp://sidads.colorado.edu/DATASETS/NOAA/G10010/G10010_SIBT1850.zip

Table 2. Files in this data set

Name	Description	Format	Size
G10010_SIBT1850.nc	Data file containing sea ice concentration and sea ice data source.	NetCDF-4 file	260 MB
montageXX.gif Where XX is the data source number (See Table 1).	An image file for quickly seeing the spatial and temporal coverage of a data source at low resolution.	GIF image File	242 - 307 KB
siindex.cpt	A custom color bar for use with ice concentration displayed using Panoply.	Panoply color table	1 KB
sisource.0.5.cpt	A custom color bar for use with ice source displayed using Panoply.	Panoply color table	1 KB

5 Contributing data sources

In Table 3, each row has a short description of a data source that is used for this data set. Data sources are numbered as they are identified in the NetCDF *seaice_source* variable. Note that numbers 6, 7, 14, 15, and 19 are not used in the current version but are retained as placeholders for the processing code and NetCDF file.

Sources 1, 2, 3, 4, and 5 (DMI yearbook narratives, JMA charts, NAVO yearbooks, Kelly ice extent grids, and Walsh and Johnson ice concentration grids) were, along with passive microwave satellite data, the sources for CW91. For more information on these data sources, see the documentation for CW91 as well as Walsh and Chapman (2001), Chapman and Walsh (1991), Kelly (1979), Walsh (1978), and Walsh and Johnson (1978). Because SIBT1850 builds on the methods as well as the sources used for CW91, we encourage users to read these references.

The Kelly grids that were converted to concentration have a distinct source number, 4, as do the Walsh and Johnson grids, 5, even though both sources include information that may have come from source 1, 2, or 3, and other sources named in the cited references. Sources 1, 2, and 3 are not replaced by source 4 (Kelly) or 5 (Walsh and Johnson) when they occur within the domain of the Kelly or Walsh and Johnson source grids.

Table 3 has a column that indicates whether concentrations from a given source include concentrations that are "not observed." These "not observed" concentrations are ones that are estimated when only the ice edge position was observed or ones that come from the analog gap filling method. We label concentrations from the DMI charts (source 13) and from whaling ship observations (sources 16, 17, and 18) as "observed" even though the value is imprecise. For these, we infer ice concentration from a description of observed ice conditions at a particular location.

Sources 8 to 21 are described in more detail in the Detailed descriptions for sources 8-21 section.

Table 3. Data Source Description

Source reference number and short name	Short description	Existence of "non-observed" data?	Year source occurs first, last
1. DMI yearbook narrative	The DMI yearbooks included summaries of ship reports. These sometimes noted ice conditions. When these observations were available for months in which the ice maps were not available, CW91 used them as ice edge observations (processed like source 4). The first example of this source can be seen in the North Pacific in April 1901. The DMI observations indicated "ice free" where source 1 appears in the NetCDF file. This is reflected in the corresponding concentration field. See Figure 11. Sources 1, 4, and 13 all use historical information from DMI.	Yes. Where an edge position is observed, concentration is estimated from edge position using the method of CW91.	1901,1938
2. JMA charts	Information derived by Walsh (1978) from analyses done by the JMA appear in some winter and spring months of years 1968-1978, with exception of 1970 and 1972. See Figure 12.	No. Analog concentration information was digitized for this source. Imprecise concentration classifications in the analog source were handled as described in Walsh (1978) and Walsh and Johnson (1979).	1968, 1978
3. NAVO yearbooks	The Naval Oceanographic Office (NAVOCEANO) compiled sea ice maps for Alaska and Greenland sectors into yearbooks for the period 1953–1971. These maps were digitized by Walsh and became part of the pan-Arctic data set described in Walsh (1978). See Figure 13. *In 1972, source 3 appears for Feb-Apr in the west Pacific. This is mislabeled. It should be source 5.	No. Analog concentration information was digitized for this source. Imprecise concentration classifications in the analog source were handled as described in Walsh (1978) and Walsh and Johnson (1979).	1953,1971*
4. Kelly ice extent grids	These are April through August or September fields of concentration estimated from Kelly's (1979) DMI map-based edge positions using the method of CW91. See Figure 14. Sources 1, 4, and 13 all use historical information from DMI.	Yes. Concentration is estimated from edge position using the method of CW91.	1901,1956
5. Walsh and Johnson	Analog maps of ice concentration estimates from the U.S. Fleet Weather Facility, the Navy-NOAA Joint Ice Center, and a number of other sources including the U.K. Met Office, the Canadian Ice Service, and a Norwegian source were digitized to make the pan-Arctic data set described in Walsh (1978) and in Walsh and Johnson (1979). The NAVOCEANO yearbook data (source 3) contributed to the pan-Arctic data set described in these publications. See Figure 15.	No. Analog concentration information was digitized for this source. Imprecise concentration classifications in the analog source were handled as described in Walsh (1978) and Walsh and Johnson (1979).	1953,1978

8. Satellite passive microwave	NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 2. The CDR product uses results from both the NASA Team and NASA Bootstrap algorithm for deriving ice concentration.	No. These are satellite passive microwave-derived observations.	1978,2013
9. ACSYS	North Atlantic ice edges from the Arctic Climate System Study. Ice edge information is irregular in time and space and is most frequent in the 20th century.	Yes. Concentration is estimated from edge position using the method of CW91.	1850,1978
10. AARI	Digitized ice charts from the Russian Arctic and Antarctic Research Institute. The grids include ice concentration information.	No. See the source product documentation for how these were compiled.	1933,1978
11. Hill	From a collection of historical ice edge position data covering Newfoundland and the Canadian Maritime region.	Yes. Concentration is estimated from edge position using the method of CW91.	1870,1962
12. Dehn	From a digitized subset of scanned analog charts converted to shapefiles with concentration information. *There are instances where 12 appears as the source after 1978 but these are limited to land grid cells only. See the section Source 12: Data from the Dehn collection.	No. Analog concentration information was digitized for this source. See the documentation for more information.	1954,1978*
13. Danish Meteorological Institute	These are concentration fields derived from an interpretation of the DMI maps.	No. Concentrations were inferred from direct observations.	1901,1956
16. Whaling log books: open water	From a mapping of whaling ship logbook entries that indicated open water.	No. Concentrations were inferred from direct observations, in a manner similar to but less precise than that used for Source 13. See the section Sources 16, 17, and 18: Data from whaling ship log books for more information.	1850,1919
17. Whaling log books: mixed	From a mapping of whaling ship logbook entries for areas where ice and open water observations were mixed.	No. As for Source 16.	1850,1918
18. Whaling log books: sea ice observed	From a mapping of whaling ship logbook entries that indicated ice at any concentration greater than about 1/8th.	No. As for Source 16.	1850,1919
20. Analog filling of spatial gaps	Analog filling of spatial gaps - spatial gaps in a given grid filled with best analog representations of the given month,	Yes.	1850,1978
21. Analog filling of temporal gaps	Analog filling of temporal gaps - temporal gaps of entire missing grids filled with analog representations of the missing month	Yes.	1850,1952

5.1 Uncertainties in source numbers

There may be cases where sources are assigned erroneously within the domain of the Kelly (source 4) or Walsh and Johnson (source 5) source grids. For example, it appears that April 1969 has information from JMA charts over most of the Sea of Okhotsk. In fact, this information came from the UK Met Office. In December 1971 and January 1972, the source for much of the Sea of Okhotsk is NAVO Yearbooks (source 3), when in fact these yearbooks did not have observations for this area. This source should be tagged as source 5 (Walsh and Johnson). The concentration values are not affected.

In addition, later sources may have been erroneously coded in some cases. For example, source 13 (DMI concentrations) is the apparent source over a domain mapped by the Kelly grid in the northern Pacific. This appears to be an error in processing, and in this case, concentrations are affected. These errors in concentration and source numbers will be corrected in a future version of the data set.

The errors in source number labeling that first occurred in CW91 and that have persisted through this update are believed to be inconsequential because concentration values were not affected. Errors that may have occurred in the processing to create SIBT1850 are believed to be anomalies with little effect on Arctic-wide concentration values, but a thorough assessment cannot be made at this time. Those who wish to subset the data for particular months are encouraged to view the individual months. This can be done using the Panoply application; see Exploring the data with Panoply.

6 Detailed descriptions for sources 8-21

This section describes sources 8 through 21. For a description of sources 1 through 5, see the Contributing data sources section.

6.1 Source 8: Sea ice concentration from satellite passive microwave data.

These data are from *NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 2*. The CDR uses results from both the NASA Team and NASA Bootstrap algorithm for deriving ice concentration. For more information, see the product documentation (Meier et al. 2013).

The parameter *goddard_merged_seaice_conc* is used for 1979 to 1987; and the parameter *seaice_conc_cdr* is used for values from 1987 on. The mid-month (15th or 16th) daily field is used. These satellite passive microwave-derived data are the only source used in SIBT1850 from 1979 onward.

6.2 Source 9: Data from the Arctic Climate System Study (ACSYS)

There are several related data collections that may be referred to as ACSYS historical ice data. The ACSYS Historical Ice Chart Archive (1553-2002) was published on CD-ROM in 2003 (ACSYS 2003). It includes the original ACSYS archive plus data from the Norwegian Meteorological Institute (NMI).

In 2007, NOAA@NSIDC published an extension to that collection titled *March through August Ice Edge Positions in the Nordic Seas, 1750-2002* (Divine and Dick 2007). Ice edge location (position points) is the only parameter in that data set. Contributing to its derivation, however, are ice concentration files from the NMI, from AARI, and from satellite passive microwave data. Figure 6 in that documentation illustrates the data density of sea ice edge positions from the ACSYS archive between 1850 and 1899.

For SIBT1850, edge positions for 1850 to 1978 were obtained from *March through August Ice Edge Positions in the Nordic Seas, 1750-2002*. An estimate of ice concentration was derived from these using the method described in the Estimating marginal ice zone concentration from ice-edge-only data section.

6.3 Source 10: Data from the Russian Arctic and Antarctic Research Institute

These data were published as *Sea Ice Charts of the Russian Arctic in Gridded Format, 1933-2006* (Arctic and Antarctic Research Institute 2007) and are available with a description of how ice concentrations were derived. Figure 8 in that documentation illustrates the spatial and temporal coverage of these data.

6.4 Source 11: Data from NRC Canada, B.T. Hill

These data cover the Labrador Sea and Baffin Bay areas. They are from a collection assembled by Brian T. Hill, Institute for Ocean Technology, National Research Council (NRC), Canada. They were acquired for this project in 2012 as files downloaded from http://www.icedata.ca/; however, they are no longer available from that site. We derived an estimate of ice concentration from the ice edge positions using the method described in Estimating marginal ice zone concentration from ice-edge-only data. In these data, winter sea ice extent off Newfoundland between 1810 and 2010 (Figure 3) shows considerable variability. More information about these data may be found in Hill (1999) and Hill and Jones (1990).

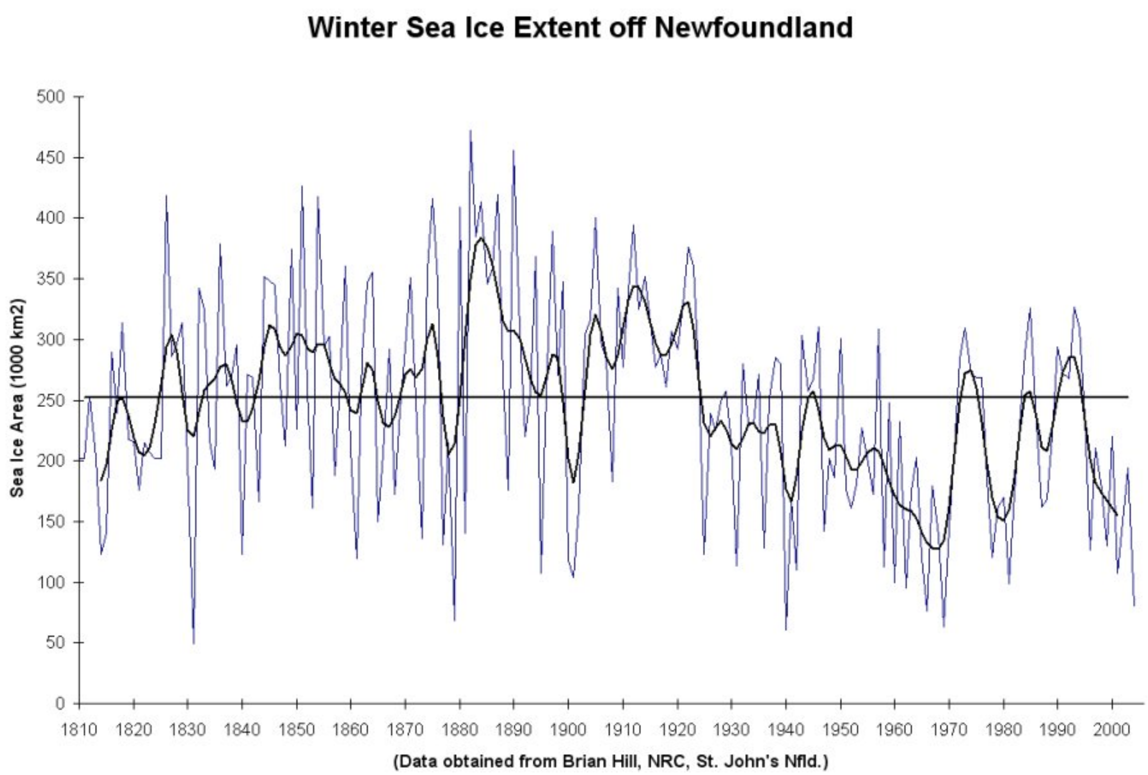

Figure 3. Yearly sea ice extent east of Newfoundland, Canada (data from Hill (1999) and Hill and Jones (1990).

6.5 Source 12: Data from the Dehn collection

These data are scanned charts published by NOAA@NSIDC as *The Dehn Collection of Arctic Sea Ice Charts, 1953-1986* (NSIDC 2005). An article by Dehn (1972) gives an overview of ice observing and forecasting in Alaska at the dawn of the digital age. It provides useful context for understanding airborne ice observation data.

A subset of the scanned charts was digitized by the Alaska Center for Climate Assessment and Policy (ACCAP) at UAF. There, the scans from NSIDC were georeferenced and converted to shapefiles with concentration information so that the data could be used with the GIS-driven Historical Sea Ice Atlas for Alaskan Waters as well as in SIBT1850.

Not all charts were used; selections were based on information content. Polygons containing ice of some concentration determined by the notations on the chart were traced by hand. The lines on the scanned charts were often faint or broken, and the meaning of the code written near them was not always clear. Notation was often sparse, especially for older charts in the series. Our partners, with UAF's Scenarios Network for Alaska Planning (SNAP), led by Lena Krutikov, found that in spite of these issues they were able to retrieve much of the unique data these charts hold.

The Dehn charts were converted to shapefile format with concentration information for input into the GIS-driven Historical Sea Ice Atlas for Alaskan Waters.

Figure 4 provides an overview of how the Dehn charts were processed at UAF. In summary, the charts went from a non-spatially-specific raster to a georeferenced and attributed shapefile. Each shapefile was then combined with other Dehn charts from that month before being provided to UIUC.

Note: Source 12 is not a valid source for ice concentration information after 1978. However, source 12 consistently appears in the source fields as the source for 21 grid cells near Alaska from 1979 through 2013. Twenty of these consistently have values of -1 in the corresponding concentration fields, meaning they are "land". This erroneous encoding of the source fields will be corrected in a future version of the data set.

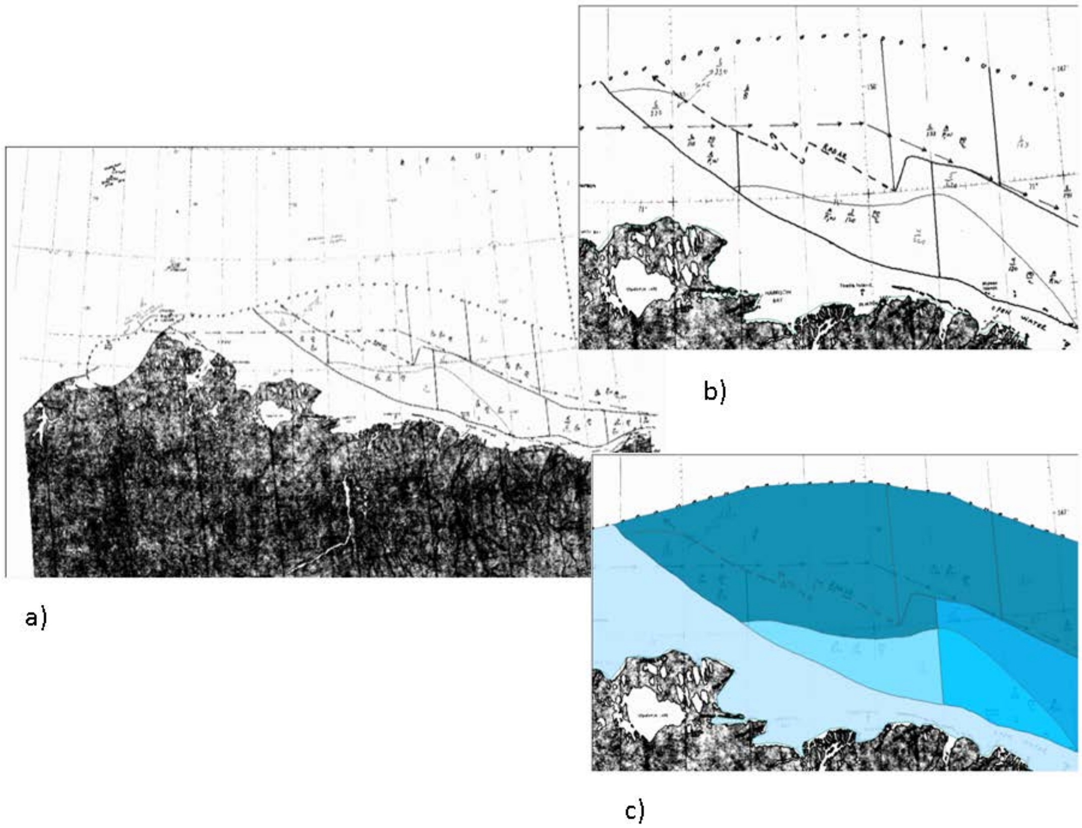

Figure 4. This example illustrates how the Dehn charts were processed at UAF. a) A Dehn chart scan for August 1954. These were provided to UAF-ACCAP as TIFF files. b) A close-up section of the chart. Note that some areas have a great deal of detail. Deciphering notation or handwriting could be challenging, but UAF-ACCAP noted that there was a good deal of consistency in notations. Most charts had "fraction-looking numbers" whose top digit corresponds to the sea ice concentration for that area in tenths. c) Once charts were georeferenced, ice edges were digitized in ArcGIS to create polygons of individual sea ice concentrations.

6.6 Source 13: Data from the Danish Meteorological Institute (DMI)

These data were published by NOAA@NSIDC as *Arctic Sea Ice Concentration and Extent from Danish Meteorological Institute Sea Ice Charts, 1901-1956* (Underhill, Fetterer, and Petersen 2014) and are available with a description of how ice concentrations were derived.

Source 1 and source 13 both originate with DMI, but concentrations from source 1 are as described in the documentation for Chapman and Walsh (1991).

6.7 Sources 16, 17, and 18: Data from whaling ship log books

From the middle of the 19th century into the beginning of the 20th, whaling vessels were active in the Bering, Chukchi, and Beaufort Seas. The logbooks for these ships have entries noting the ice coverage along with meteorological conditions and the date and geographical position of the ship. The information in the logbooks was collected by the New Bedford Whaling Museum and later analyzed by Mahoney et al.

(2011). We obtained a version of these data from Dr. Mahoney. Each record includes a ship identifier; the year, month, and day of observation; the latitude and longitude of the ship; and a binary indicator of whether ice was present at any concentration greater than about 1/8th.

In all, there are 52,717 daily records starting in June of 1849 and ending in October of 1914. For each month beginning with January 1850, we created image files with a yellow dot for every location where ice was noted and a red dot for those locations where sea ice was not observed to be present. For many months in the record, the observations were too sparse to be useful for defining the ice cover. We therefore included observations from the same month in the five years preceding and following, as a kind of running sum. These points are white where ice was observed and grey where it was not.

Figure 5 gives examples from April of 1882 and 1908. In these examples, the observations of *ice* and *no ice* are fairly well separated, even when the entire 10-year window of surrounding April observations is taken into account. In examples from June of 1852 and 1864, *ice* and *no ice* observations are more mixed (Figure 6). To use these data in SIBT1850, we manually mapped the point observation images into polygons of *ice*, *no ice*, and *mixed* with the aid of student assistant Vivian Underhill. The *ice* polygons are yellow, *no ice* or open water polygons are blue, and *mixed* are green. Figure 7 illustrates mapping the June 1850 compilation at an intermediate stage. In the resulting product (Figure 8), some areas of ice (yellow) are included. In SIBT1850, ice polygons are labeled as source 18, mixed ice polygons are source 17, and open water polygons are source 16.

It is important to note that the source 18 polygons do not indicate 100 percent ice cover; rather, they indicate ice at any concentration greater than about 1/8th. See Mahoney et al. (2011) for an explanation of how 1/8th was determined. Similarly, source 17 does not explicitly map areas of partial concentration for the month in question.

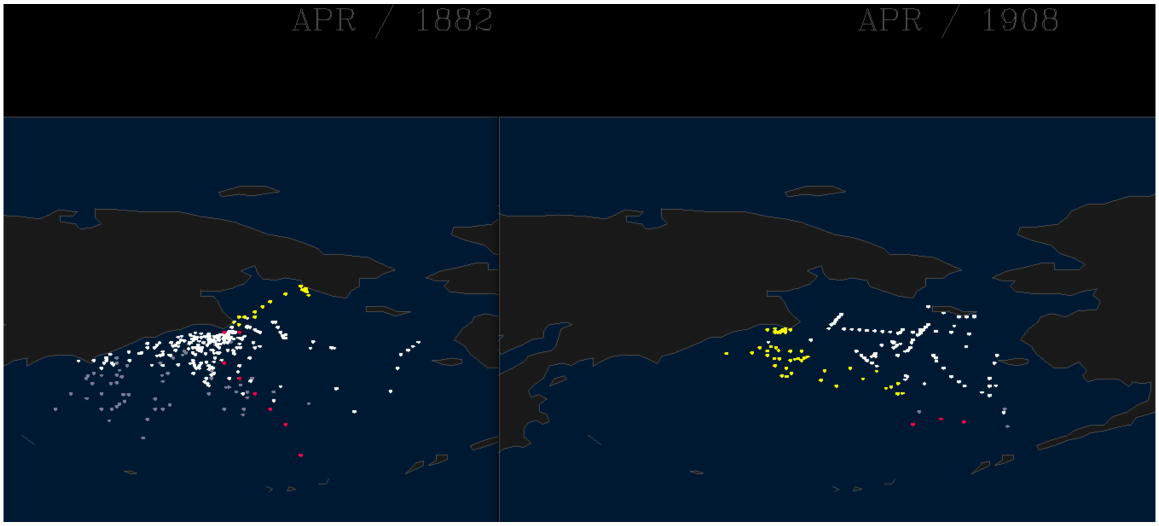

Figure 5. April 1882 (left) and 1908 (right). Observations of ice are yellow or, if from the 10-year window of Aprils around the month in question, white. Open water observations are red or grey.

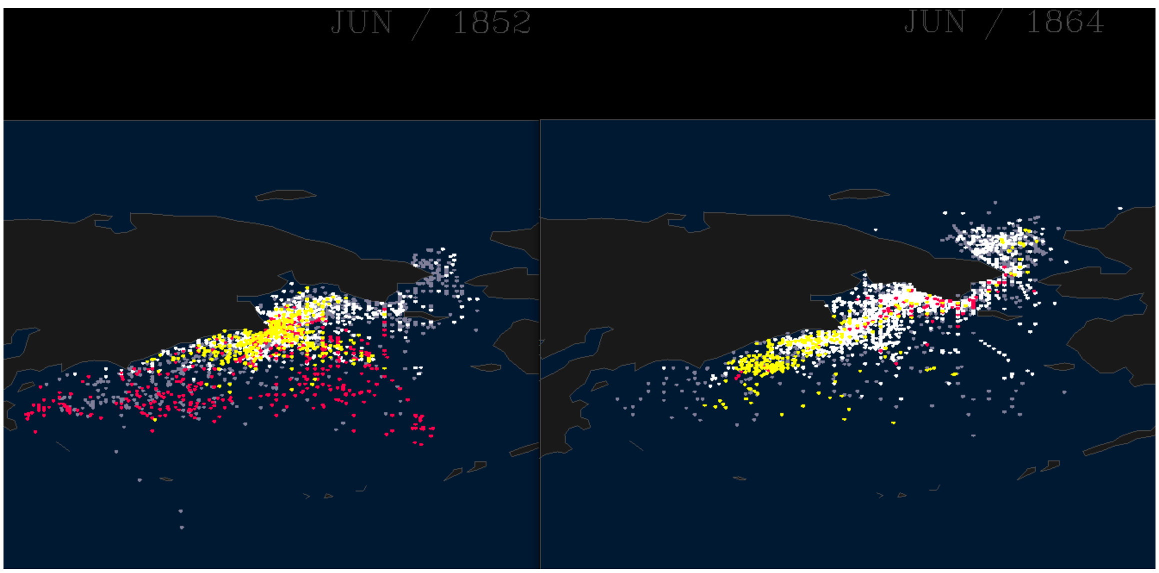

Figure 6. June 1852 (left) and 1864 (right). Observations of ice are yellow or, if from the 10-year window of Junes around the month in question, white. Open water observations are red or grey.

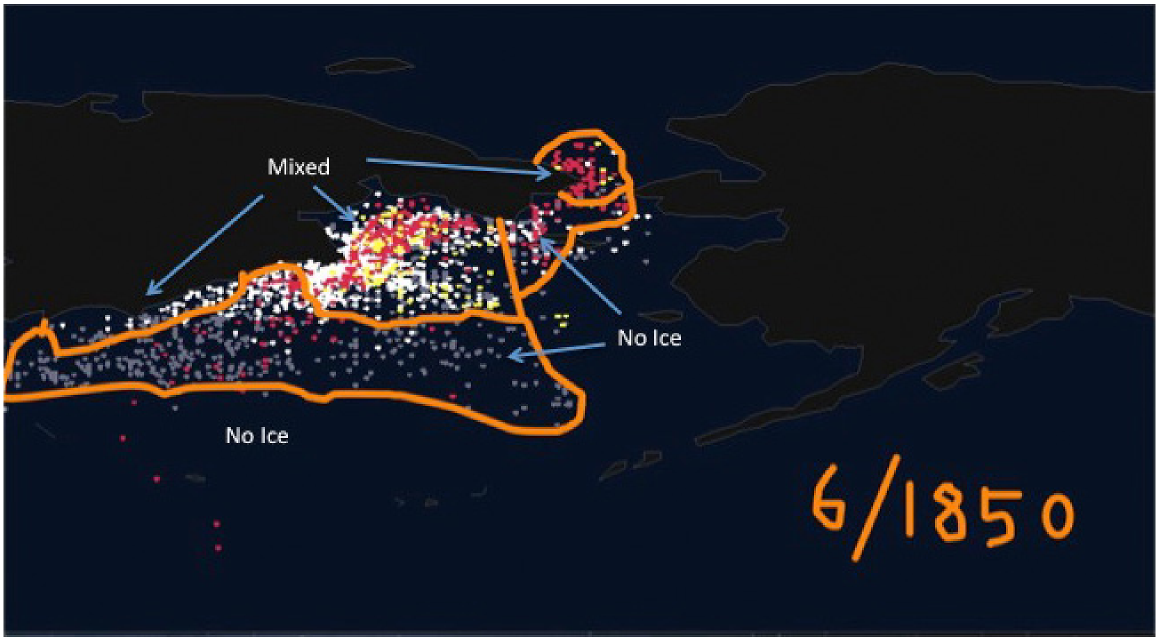

Figure 7. The GIMP image manipulation application was used to draw around areas of ice, open water, or mixed ice and open water observations for June 185-.

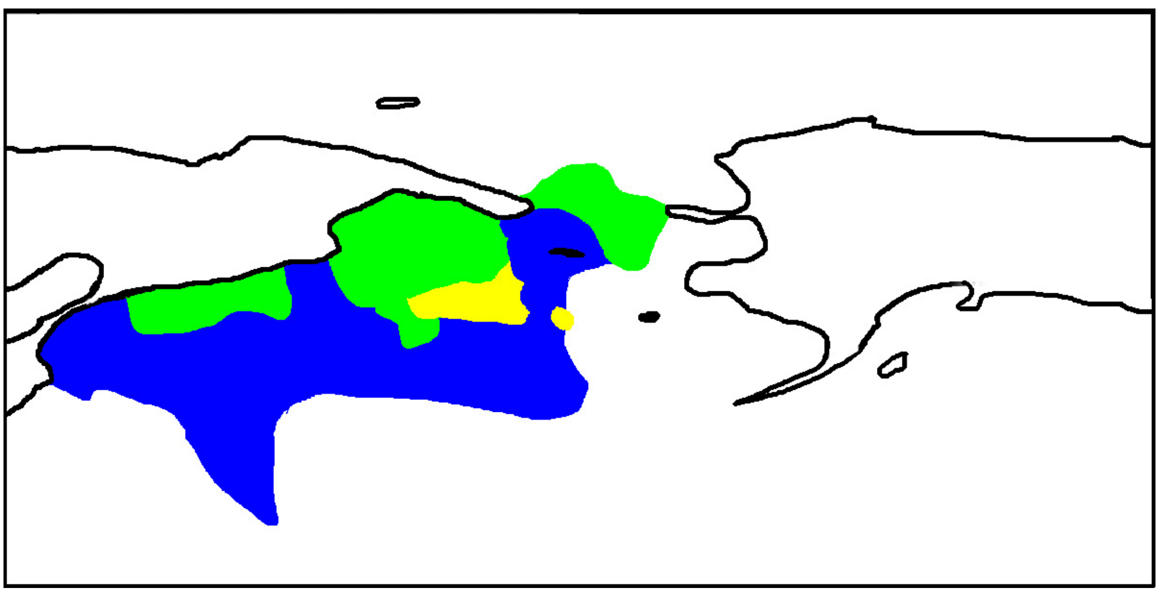

Figure 8. This image shows the results of the June 1850 mapping where these are made into color-coded polygons marking *ice* in yellow, *no ice* in blue, and *mixed* in green.

In cases of conflicting observations, observations from the year in question were given higher priority than observations from the surrounding ten years. Observations at a point were considered to apply over an area with about a 10 km to 20 km radius around that point. This varied depending on surrounding observations and location; for instance, open water observations in lower latitudes were extrapolated to cover more distance than open water observations at high latitudes.

Mapping the ice cover from whaleship observations in this way clearly requires subjective judgment. Should we wish to review how a particular month was analyzed, it would be possible to review the analysis at intermediate steps. Three types of files were saved: layered GIMP files (in .xcf format), showing both the plotted observations and the polygon tracings at 50 percent transparency; final polygons (in .tiff and .png format); and the original plotted observations (in .gif format). These products are not distributed but are archived at NSIDC.

In many cases, data for the month in question can be in contradiction because we are including ten years' worth of data in each image. This can lead to maps that show sea ice in patterns contrary to normally observed ice coverage. For instance, in September 1910 (Figure 9), there are areas of blue (open water) north of the dominant area of yellow (ice). It is unlikely that there were open-water polynyas of that size north of an ice band of such size in September, especially in 1910, when ice conditions were considerably heavier than they are today. The small coastal open-water areas north of Alaska, also, were not likely to have had that configuration in September 1910. These areas of open water were mapped based on grey dots, meaning they are from September in a year between 1905 and 1915, but not 1910. Though they contradict the probable ice extent configuration for that year, we leave them in the final images. We judge it better to be true to our basic tracing algorithm than to include some varying element of subjective judgment in mixing plotted observations with personal knowledge about sea ice movement and monthly conditions.

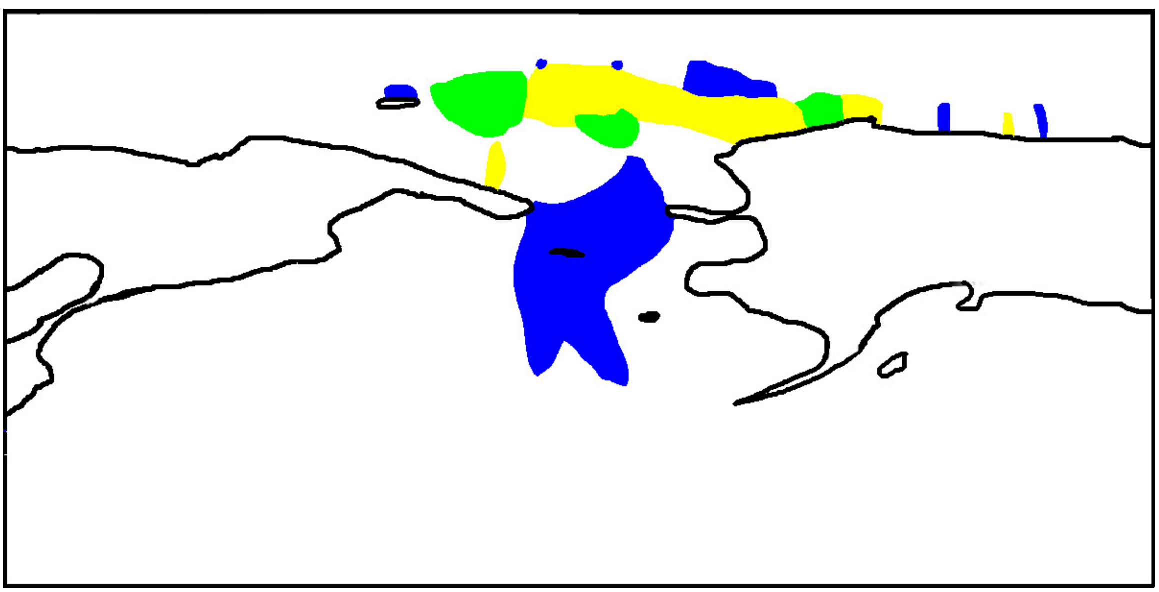

Figure 9. Polygons for ice cover in September 1910. While ice (yellow) north of open water (blue) in this example is unrealistic, altering the product would introduce a higher order of subjectivity into the resulting analysis.

6.8 Source 20: Analog filling of spatial gaps

See the section on Filling spatial gaps.

6.9 Source 21: Analog filling of temporal gaps

See the section on Filling temporal gaps.

7 Method used to merge data sources

The data were merged with code that was developed and run at UIUC using a ranking hierarchy to determine what source to use when more than one source was available for a location. Each source was made temporally consistent before the merging began.

7.1 Ranking

Each of the possible sources for a concentration value at a particular location was given a rank imposed by processing order as indicated by the row number in the table of sources (Table 3), where higher numbers outrank lower ones. For example, if AARI data (source 9) were available, those data took the place of any data from 1-7 for a given month. Data sources 9 to 18 rarely overlap in space and time.

Note that source 8 (satellite passive microwave concentration) is the exception to this ranking rule. Source 8 is used exclusively after 1978, even though data from higher-ranking sources, such as AARI data, may be available after 1978.

7.2 Imposing temporal consistency

The monthly files are intended to represent ice at the 15th or 16th of each month. This mid-month target can be met with the daily satellite passive microwave grids, but other sources have data at the end of the month or for arbitrary dates within the month. In most cases, values in the available data file were used without adjustment. The whaling ship observations are a special case; see the section on Sources 16, 17, and 18: Data from whaling ship log books.

For the Dehn charts (source 12) coverage was frequent enough that liner interpolation between values for concentrations before and after the day in question was a reasonable means of approximating mid-month concentration. The Dehn charts often cover different areas within the same month. At UAF, the series was reviewed and charts judged to be best for the purpose were selected for interpolation.

The mid-month daily value was selected instead of the monthly average value because the historical sources are too sparse to allow a realistic monthly average to be constructed. Each instance of a historical source's concentration information is more like a snapshot than like an average of variable ice conditions. To keep variability realistically consistent through the entire series, and to increase the probability of capturing conditions around the time of transitioning to freeze-up in the fall or melt season in the spring, the mid-month day was used even when constructing a field of monthly averages was possible. For example, the annual minimum extent typically occurs around the middle of September. An average ice concentration field for September would likely show ice extent to be greater than one constructed from a single mid-month day.

7.3 Estimating marginal ice zone concentration from ice-edge-only data

Two sources, 9 and 11, are *ice-edge-only* data sets. The source data are in the form of a series of ice edge positions. One source, 13, has an inferred ice edge position where the ice concentration was not observed.

To make the most of these data, an estimated concentration was assigned for points within the edge. The estimate was arrived at using a series of functions (gradients) derived from preexisting monthly fields for 1953 to 2002. These consisted of CW91 fields extended with passive microwave derived concentrations.

Ice concentration as a function of distance from the ice edge, where the edge is defined by cells having 10 percent or greater concentration, was determined for every grid point as well as each month and longitude in the 49-year series. This provided a sample size of 49 with which to derive the gradient function for a given month. With the resulting 12 monthly functions or look-up tables for each longitude, a concentration could be assigned to each grid cell within the ice edge for a particular ice edge source file, using the month and that grid cell's distance to nearest open water. This method of estimating concentration within a marginal ice zone based on month of year and distance from the ice edge takes into account seasonal, as well as regional, variability. Figure 10 illustrates this with an example from the ACSYS ice edge data collection (source 9).

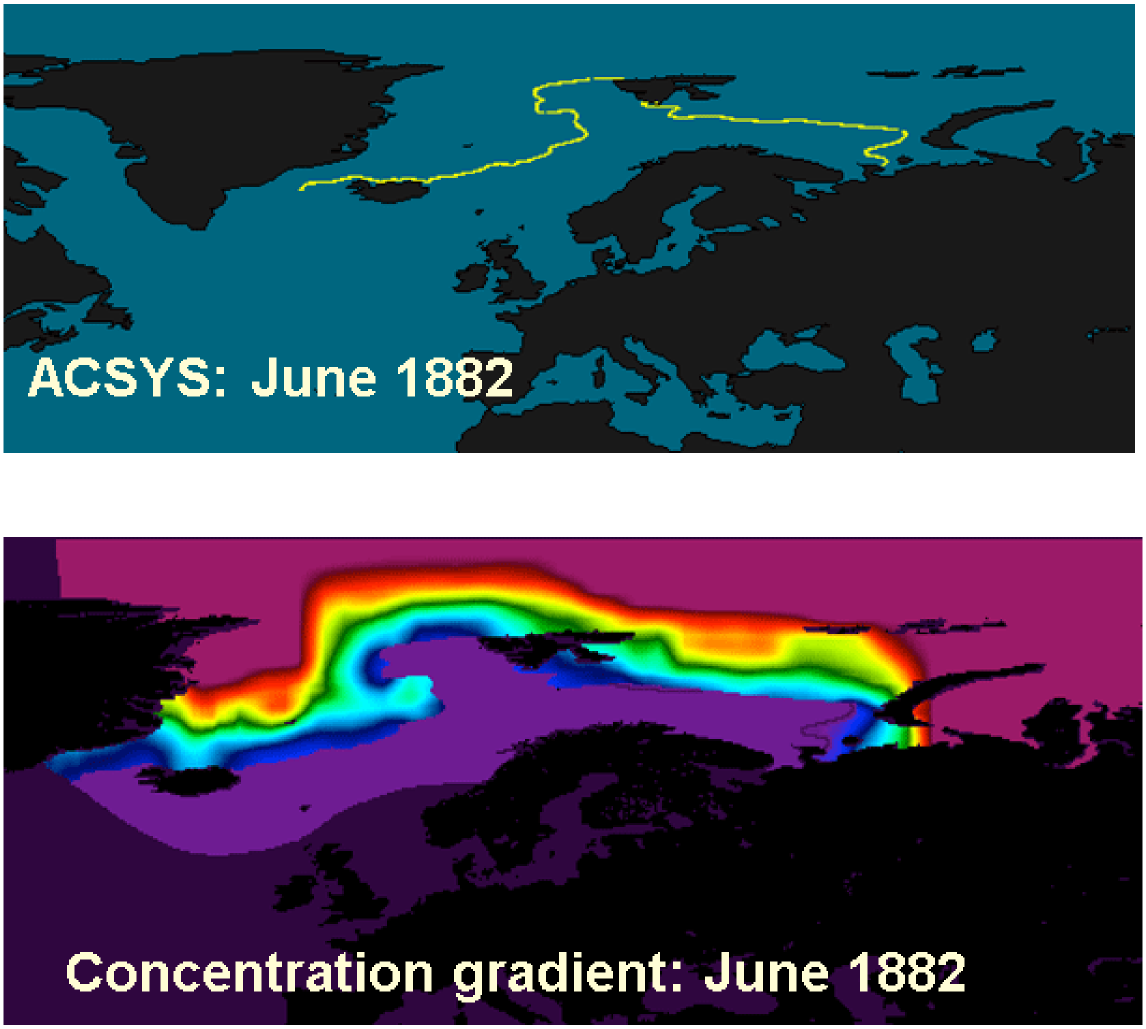

Figure 10. Historical data products that give only the ice edge position are used to infer a concentration within the edge by applying a gradient function. In this example, the information in the ACSYS ice edge product for June 1882 (top) is used to infer a concentration gradient within the edge (bottom).

8 Filling data gaps with an analog method

8.1 Filling spatial gaps

In 1953, the U.S. Navy began to produce regional ice analyses on a regular basis; the analysis domain became the entire Arctic in 1972, when the predecessor of the National Ice Center performed the charting. Continuous coverage began with the satellite era in 1978. Prior to 1953, ice concentrations were sparsely observed. To fill gaps in a field for a particular year and month, the month before and the month after are checked; and if these both have concentrations for the missing points in the field with gaps, hereafter the “gap field,” those are interpolated to fill concentrations in grid cells where they are missing. Such grid cells will be labeled as source 20 (analog filling of spatial gaps).

If the surrounding months do not have data with which to fill the missing value for a given point or points, the field is compared to the ice concentration fields between 1900 and 2000 for the same month. How well a potential analog field matches the gap field is measured by spatial correlation for the sub-domain mapped out by the points that have data in the gap field. The best three matches that have data for points where data are missing in the gap field are selected as analogs. They are then averaged together and used to fill in the non-observed part of the field with gaps. Such grid cells will also be labeled as source 20 (analog filling of spatial gaps).

If the three best analog years from 1900 through 2000 do not have data for points in question, the search is repeated by limiting the analog candidates to the 48 possibilities between 1953 and 2000. The resulting best three matches may not rank as highly as the best three from the 101 candidates between 1900 through 2000, but these later years have much better coverage and so are more likely to have data for the gap grid cell points.

If the search between 1953 and 2000 does not produce three analog fields that have data at gap points, then fewer than three analogs are used.

Missing values in the gap field are then filled with the average concentration values for those points from three, two, or one analog.

Stated succinctly, the process is as follows:

1. For each grid point *p* with no data in calendar month *m* and year *y*, and no data in the surrounding 2 months, areas with existing data in *m*-*y* are compared with calendar month *m* of all years 1900-2000 to select the best analogs.
2. If the three best analog years do not have data at point *p*, the search is repeated by limiting the analog candidates to 1953-2000.
3. If Step 2 does not produce three analogs with data at point *p*, then fewer than three analogs are used.
4. Point *p* is "filled in" with the average concentration of the (up to) 3 analog fields.

8.2 Filling temporal gaps

For some of the early part of the data record, there are absolutely no direct observations available in particular months. In these cases, we find the best three analogs from the historical record for the nearest month that has some direct observations. For example, if there are no observations for February 1853 but there are some observations for March 1853, we would find the best three analogs to March 1853. These might be March 1932, 1964, and 1982. The concentration values for the grid for February of 1853 would then be the average of February 1932, 1964, and 1982. The source would be set to 21 (analog filling of temporal gaps) for each grid cell for February of 1853 in this hypothetical example.

8.3 Summary of gap field concentration value sources

The final ice concentration field for a gap field might then consist of estimates from several sources:

1. An interpolation of observations coming from the month before and the month after, if both of those have values for the point in question. These are flagged with source number 20 (analog filling of spatial gaps).
2. An average of observed values from analogs of the same calendar month in the 20th century. These are flagged with source number 20 (analog filling of spatial gaps)) or 21 (analog filling of temporal gaps).
3. As above, but using analogs from only the post-1978 (satellite) era. These are flagged with source number 20 (analog filling of spatial gaps)) or 21 (analog filling of temporal gaps).

The period searched for analogs ends in 2005. This period was selected in order to have as long a period as possible offering fields with no gaps, thus improving the chances of finding higher spatial correlations. At the same time, it does not extend far into the 21st century, when ice cover is rapidly declining.

One artifact of this method is that, often, the same three analog months are chosen. This is because, overall, there was more ice in the early part of the record, where gaps need to be filled. The analogs for the pre-1953 fields are sometimes the same three post-1953 fields from the early part of the series because those are the ones with maximum ice cover.

8.4 How the DMI compilation is used: Estimating ice concentration from ice-edge-only data and filling spatial and temporal gaps in CW91

The early part of the CW91 record consisted largely of information drawn from the DMI compilation of maps and yearbook narratives. SIBT1850 includes information from DMI charts and yearbooks as sources 1, 4, and 13. This section describes how source 4 (Kelly ice extent grids) and a source used in CW91 called "Temporal extension of Kelly data" are related. In SIBT1850, there is no need to use the "Temporal extension of Kelly data." See Table 3 and the section on Source 13: Data from the Danish Meteorological Institute for information on how sources 1 and 13 use the DMI compilation.

Scanned versions of the DMI analog maps are available from NSIDC as DMI and NSIDC (2012). From the analog maps, Kelly (1979) produced a data set of the "circumpolar ice limit." Kelly digitized only the ice edge and only to a spatial resolution of approximately 100 km, depending on the distance of the ice boundary to the pole. CW91 uses these Kelly ice edge data in the following way:

> We add a marginal sea ice zone to the Kelly ice extent data by computing average ice concentration drop-off rates for the period during which there are satellite observations. These drop-off rates indicate the rate at which ice concentrations increase as a function of distance (from 10/10 ice

concentrations toward) open water. The drop off rates vary with season; the summer melt season drop-off rate is about 0.5 that of the freeze-up season. We apply these drop-off rates to the Kelly ice extent data to create a marginal sea ice zone (Chapman and Walsh 1991).

In CW91 and SIBT1850, the concentration field described above is source 4 (Kelly ice extent grids). This method of estimating the ice concentration from ice edge only data is slightly different than the method used in SIBT1850 for sources 9 and 11.

8.5 Discussion

Relatively few cases arise in which sources conflict. A ranking method was chosen over a more complicated method such as weighting sources depending on various factors. In the present version of the data set, a binary weighting of 1 or 0 was used, effectively basing the assigned concentration on a single, "judged best," source.

It is not always possible to trace the provenance of the pre-1979 sources. This, and the need to rely on subjective interpretation of long-ago written observations, as well as the fact that point observations are extrapolated to grid cell areas, means that it is not possible to assign a quantitative uncertainty to the data. A sense of where one may have high or low confidence in the data may be gained by viewing the coverage of various sources in the data file source layers and by reading the sections of the documentation pertaining to those sources.

In merging these sources, we have put together data of vastly different spatial and temporal resolution and precision. Each data source is unique. For example, there are data values that may have started with a subjective observation like "close pack ice" that we then interpreted as a concentration and drew an area around in order to use these observations in a grid.

9 Exploring the data with Panoply

The figures in this documentation used to illustrate sources and concentrations in the Appendix (figures 11-15) were created using NASA's Panoply data viewer. Custom color tables for the source and concentration arrays are available in the files sisource.0.5.cpt and siindex.cpt.

Users may find the following guidelines helpful if they wish to explore the data with Panoply. Note that these guidelines were developed for Windows running PanoplyWin-4.3.0, but Mac and Linux procedures are likely to be similar.

9.1 Display images using Panoply

Follow these guidelines to display images like those in this documentation.

1. Download the latest version of Panoply from http://www.giss.nasa.gov/tools/panoply/ and double-click to install the executable.
2. Double-click to run Panoply
3. The program starts and immediately goes to the Open screen. Find and select the SIBT1850 NetCDF data file. Note: This file should be approximately 350mb in size.
4. To create a 2D plot, choose either seaice_conc or seaice_source and clicking "Create Plot".
   - Allow it to create a georeferenced Longitude-Latitude plot
   - Unclick the "interpolate" box so that each color corresponds exactly to the underlying data.
   - If plotting the source fields, choose the custom sea_ice_source color table (sisouce.cpt) by opening it.
     - Choose File->Open from the main Panoply window and find the file "sisource.0.5.cpt"
     - Choose "Okay" to import the color table to the support library
     - Activate the color table in the current seaice_source view by selecting the lower tab "Scale" and choosing "sisource" from the dropdown menu
     - For the best appearance, manually set:
       - Scale Range of Min: -1 and Max: 21 (Do not choose "Center on 0" or "Fit to Data")
       - Bar width: 80%
       - Divisions: Major: 11 and Minor: 2
       - Tick Label Format: %.0f and Size 11

   - If plotting the concentration field, adapt the steps above accordingly:
     - Choose the "Scale" tab
     - Import the sea ice concentration color table, by going to File->Open and choosing "siindex.cpt" library.
     - Change the color bar to "siindex.cpt"
     - Change the Scale Range to Min: -1 Max: 100
5. From the "Map" tab, choose the Projection: "Stereographic" and set Center on "-90"E and "90"N with a Radius of perhaps 40 degrees to start.
6. Go to the "Array(s)" tab, near the bottom, to select a date to display.

9.2 Setting up Panoply defaults

Users may wish to set Panoply defaults as follows: Go to Edit->Preferences at the top of the Panoply window.

- General:
  - set Plot Size to Large (or larger)
  - set Interpolation to off = uncheck the box
- Scale:
  - Set Color Table to sisource.0.5.cpt (for example)

    - Set Bar Width to 80
    - Set Divisions: Major 11 Minor 2
    - Set Tick Label Format to %.0f
- Lon-Lat Plots:
    - Projection: Stereographic; Centered on -90E 90N
    - Radius: 40N

10 Version History

Table 4. Product Version History

Version	Date	Description
V1.1	Jan. 2016	The data were updated with a consistent land mask that fixed a problem in the consistency of the land area that occurred primarily at the 1979 pre-/post-satellite era boundary. Some grid cells that were marked as *land* prior to the satellite era that begins in 1979 were treated as ocean and marked with sea ice concentration values in the post-1979 period. This inconsistency was also present to a lesser extent along coastlines for some other source data transitions. Now, any grid cell that was flagged as *land* at any time in the record is now flagged as *land* throughout the entire record. In addition, all grid cells that are *land* were set to a value of -1.
V1.0	Dec. 2015	Initial release of data set.

11 Errors and Quality Assessment

NSIDC affiliate J. Scott Stewart provided the custom Panoply color tables and instruction, and assisted with formatting; review; and quality control of the data.

The Table 5 describes a list of errors found in version 1.1 of this data set. These errors will be addressed in the next version of this data set.

Table 5. Errors Found in Version 1.1

Error	Description
Mid-Atlantic Sea Ice present in June and July 1962	There are two anomalies in the data set for Jun-Jul '62. There is a patch of ice shown in both June and July of 1962 in the mid-Atlantic around latitude 35.375 degrees North and longitude 304.625 degrees East that is false.
Kelly fields are displaced by a month in June, July, and August	During 1935-1939 and 1946-1952 for June, July, and August, the Kelly fields are displaced by a month where the August data is actually July data, July data is actually June data and so forth. This issue is present every year of the Kelley data is available.

12 Related data sets

The Historical Sea Ice Atlas for Alaskan Waters is a data product that uses many of the same sources as does SIBT1850. We worked with the team at the University of Alaska to prepare the Dehn collection and the DMI collection for use in both data sets. Beyond differing geographical ranges, the data products differ in the following ways:

- There is an interface (http://seaiceatlas.snap.uaf.edu/explore) for the Alaskan atlas that makes it easy for a wide range of users to explore the data.
- The Alaskan atlas products are available weekly as well as monthly after 1952.

The Old Weather: Whaling project permits online volunteers to assist in transcribing ship's logs with a focus on sea ice information. It includes a discussion section that aids in understanding the nature of these observations and their use in scientific research.

13 Related NSIDC data collections

These are the data products that contributed to SIBT1850. A citation for each is listed in the reference section.

- Arctic Sea Ice Charts from Danish Meteorological Institute, 1893 – 1956
- Arctic Sea Ice Concentration and Extent from Danish Meteorological Institute Sea Ice Charts, 1901-1956
- Arctic and Southern Ocean Sea Ice Concentrations
- The Dehn Collection of Arctic Sea Ice Charts, 1953-1986
- National Ice Center Arctic Sea Ice Charts and Climatologies in Gridded Format
- Sea Ice Charts of the Russian Arctic in Gridded Format, 1933-2006
- NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 2

14 Acknowledgments

NOAA's Climate Program Office provided support for the development of this data product through Grant # NA11OAR4310172 (J. Walsh, PI; F. Fetterer, Co-I). Maintenance and distribution of these data is made possible through the support of NOAA to NOAA@NSIDC.

Lena Krutikov and Corey Peterson of the University of Alaska's SNAP performed the digitization of the Dehn ice charts for the Alaskan region and also participated in the preparation of the DMI chart data for incorporation into the pan-Arctic database.

University of Colorado student Vivian Underhill interpreted the whaling records and the DMI charts so that these sources could be used by the processing code. She did this with guidance from the investigators.

NSIDC affiliate J. Scott Stewart provided the custom Panoply color tables and instruction, and assisted with formatting; review; and quality control of the data.

Also linked to this work indirectly are Kevin Wood, University of Washington, who advised on DMI and whaling data; Andy Mahoney, UAF, who provided the whaling data and advised on its use; and Allaina Wallace, formerly NSIDC's archivist, who documented and had the Dehn analog charts scanned.

We appreciate reviews we received from early users of the data set. In particular, the testing provided by Holly Titchner, UK Met Office, was helpful. Cathy Smith, with NOAA Earth Systems Research Laboratory, greatly assisted us by improving the structure of the NetCDF file.

15 Document Information

15.1 Document Revisions

Table 6. Document Revision History

Date	Revision Description
Jan. 2016	A. Windnagel updated the document to reflect the new v1.1 data file by updating Table 1, Figure 1, and adding Version History and Document Information sections.
Dec. 2015	Document publication date.

15.2 Document URL

https://nsidc.org/data/docs/noaa/g10010-sea-ice-1850-onward/G10010_SIBT1850.pdf

16 References

ACSYS. 2003. ACSYS Historical Ice Chart Archive (1553-2002). *IACPO Informal Report* No. 8. Tromsø, Norway: Arctic Climate System Study.

Arctic and Antarctic Research Institute. 2007. *Sea ice charts of the Russian Arctic in gridded format, 1933-2006*. Edited and compiled by V. Smolyanitsky, V. Borodachev, A. Mahoney, F. Fetterer, and R. Barry. Boulder, Colorado USA: National Snow and Ice Data Center. http://dx.doi.org/10.7265/N5D21VHJ.

Chapman, W. L. and J. E. Walsh. 1991, updated 1996. *Arctic and Southern Ocean Sea Ice Concentrations.* Boulder, Colorado USA: National Snow and Ice Data Center. http://dx.doi.org/10.7265/N5057CVT.

Chapman, W. L. and J. E. Walsh. 1991. Long-Range Prediction of Regional Sea Ice Anomalies in the Arctic. *Weather and Forecasting* 6(2): 271-288.

Danish Meteorological Institute (DMI) and NSIDC. 2012. *Arctic Sea Ice Charts from the Danish Meteorological Institute, 1893 - 1956.* Compiled by V. Underhill and F. Fetterer. Boulder, Colorado USA: National Snow and Ice Data Center. http://dx.doi.org/10.7265/N56D5QXC.

Dehn, W. S. 1972. Alaskan Sea Ice. In *Sea Ice: Proceedings of an International Conference, Reykjavik, Iceland, May 10-13, 1971*, T. Karlsson, ed. National Research Council of Iceland: 125-129.

Divine, D. V. and C. Dick. 2007. *March through August ice edge positions in the Nordic Seas, 1750-2002.* Boulder, Colorado USA: National Snow and Ice Data Center. http://dx.doi.org/10.7265/N59884X1.

Hill, B. T. 1999. Historical Record of sea ice and iceberg distribution around Nfld and Labrador, 1810-1958. WCRP No. 108 or WMO/TD No. 949, April/99, ACSYS, *Proc. of the Workshop on Sea Ice Charts of the Arctic*, Seattle, WA AU 5-7, 1998.

Hill, B. T. and S. J. Jones. 1990. The Newfoundland Ice Extent and the Solar Cycle from 1860 to 1988. *J. Geophys. Res.* 95(C4): 5385-5394.

Kelly, P. M. 1979. An arctic sea ice data set, 1901-1956. *Glaciological Data*, Report GD-5: Workshop on Snow Cover and Sea Ice Data. National Snow and Ice Data Center: 101-106.

Meier, W., F. Fetterer, M. Savoie, S. Mallory, R. Duerr, and J. Stroeve. 2013. *NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration*, *Version 2*. Boulder, Colorado USA: National Snow and Ice Data Center. http://dx.doi.org/10.7265/N55M63M1.

National Snow and Ice Data Center (comp.). 2005. *The Dehn Collection of Arctic Sea Ice Charts, 1953-1986.* Boulder, Colorado USA: National Snow and Ice Data Center. http://dx.doi.org/10.7265/N5F769GD.

Underhill, V., F. Fetterer, and C. Petersen. 2014. *Arctic Sea Ice Concentration and Extent from Danish Meteorological Institute Sea Ice Charts, 1901-1956*. Boulder, Colorado USA: National Snow and Ice Data Center. http://dx.doi.org/10.7265/N5MP517M.

Walsh, John E., Florence Fetterer, J. Scott Stewart, and William L. Chapman. 2016. A database for depicting Arctic sea ice variations back to 1850. *Geographical Review*. doi: 10.1111/j.1931-0846.2016.12195.x.

Walsh, J. E. and W. L. Chapman. 2001. 20th-century sea-ice variations from observational data. *Annals of Glaciology* 33. http://dx.doi.org/10.3189/172756401781818671.

Walsh, J. E. 1978. A data set on Northern Hemisphere sea ice extent. *Glaciological Data*, Report GD-2, part 1, National Snow and Ice Data Center: 49 - 51.

Walsh, J. E. and C. M. Johnson. 1978. Analysis of Arctic sea ice fluctuations 1953-77. *Journal of Physical Oceanography* 9(3): 580-591.

17 Appendix: Selected concentration and source field examples

Each figure in this section shows the source field and corresponding concentration field from a selected month in the series. Examples were chosen to illustrate some characteristics of the data product and to suggest how varied source coverage can be. These images were created using the Panoply viewer.

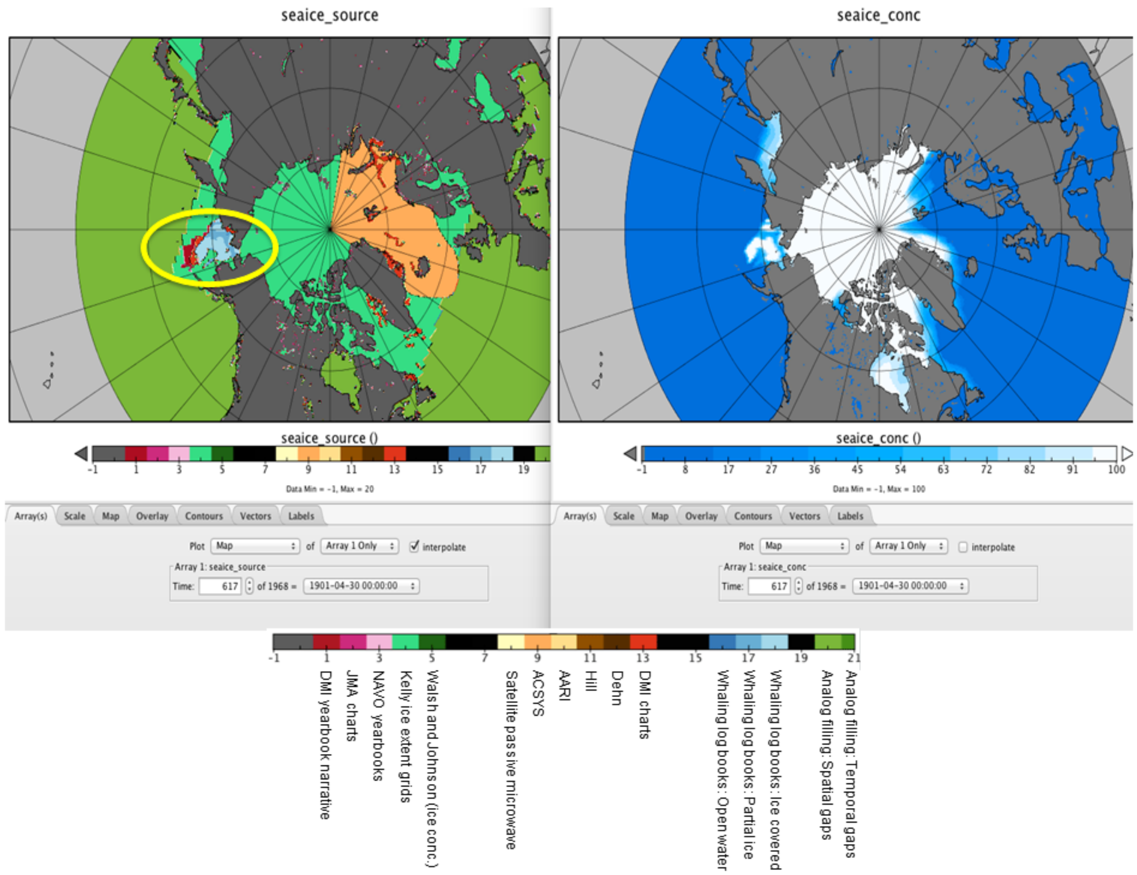

Figure 11. The first example of source 1, the DMI yearbook narrative source, can be seen in the North Pacific in April 1901 (left). The DMI observations indicated "ice free" where Source 1 appears. This is reflected in the corresponding concentration field (right).

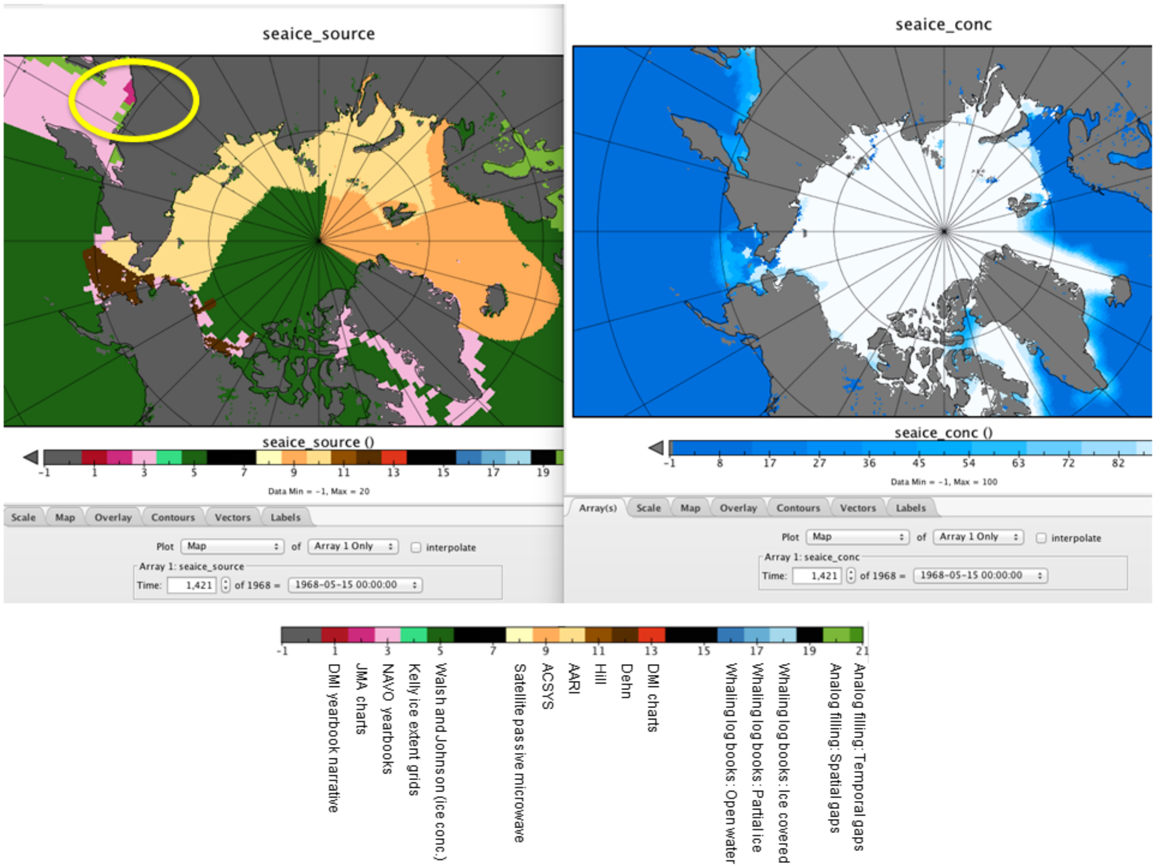

Figure 12. The first example of source 2, data from the Japan Meteorological Agency (JMA).

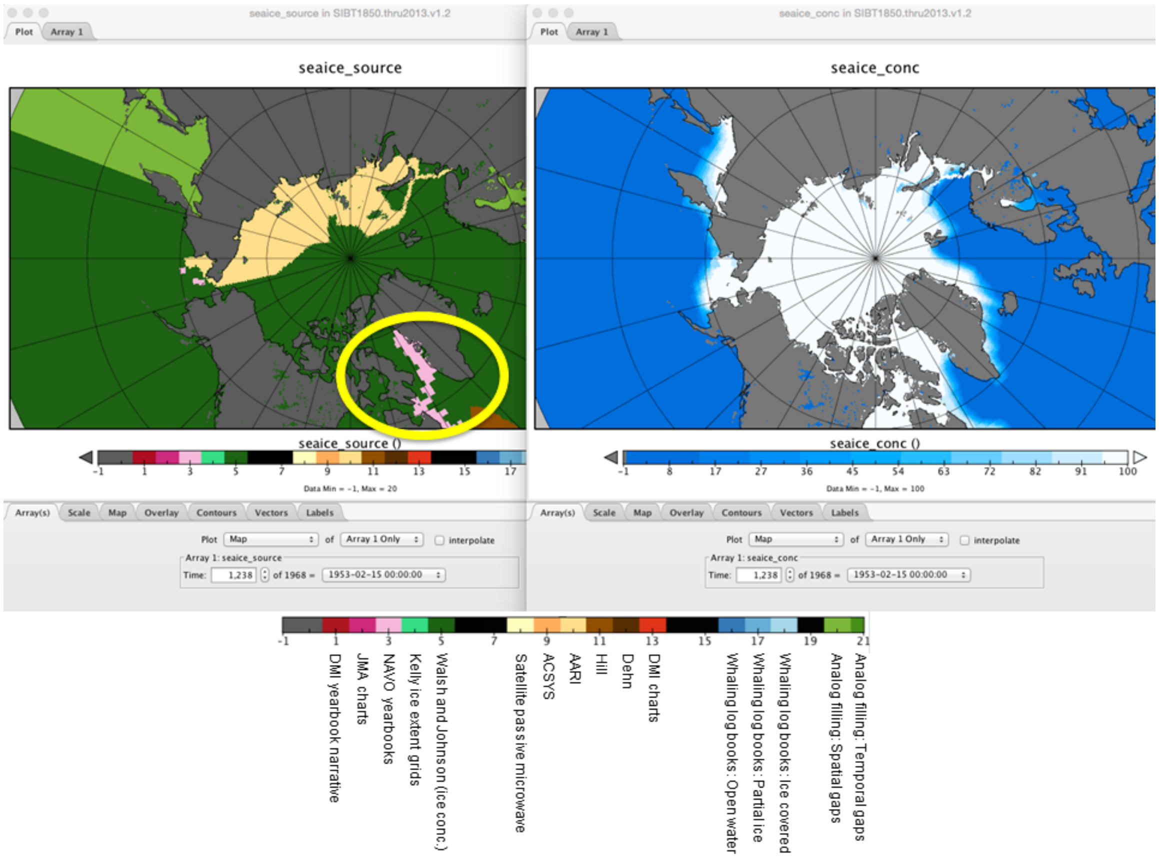

Figure 13. The first example of source 3, Naval Oceanographic Office (NAVO) yearbook data, occurs in February 1953, where it aids in defining the ice edge in Baffin Bay.

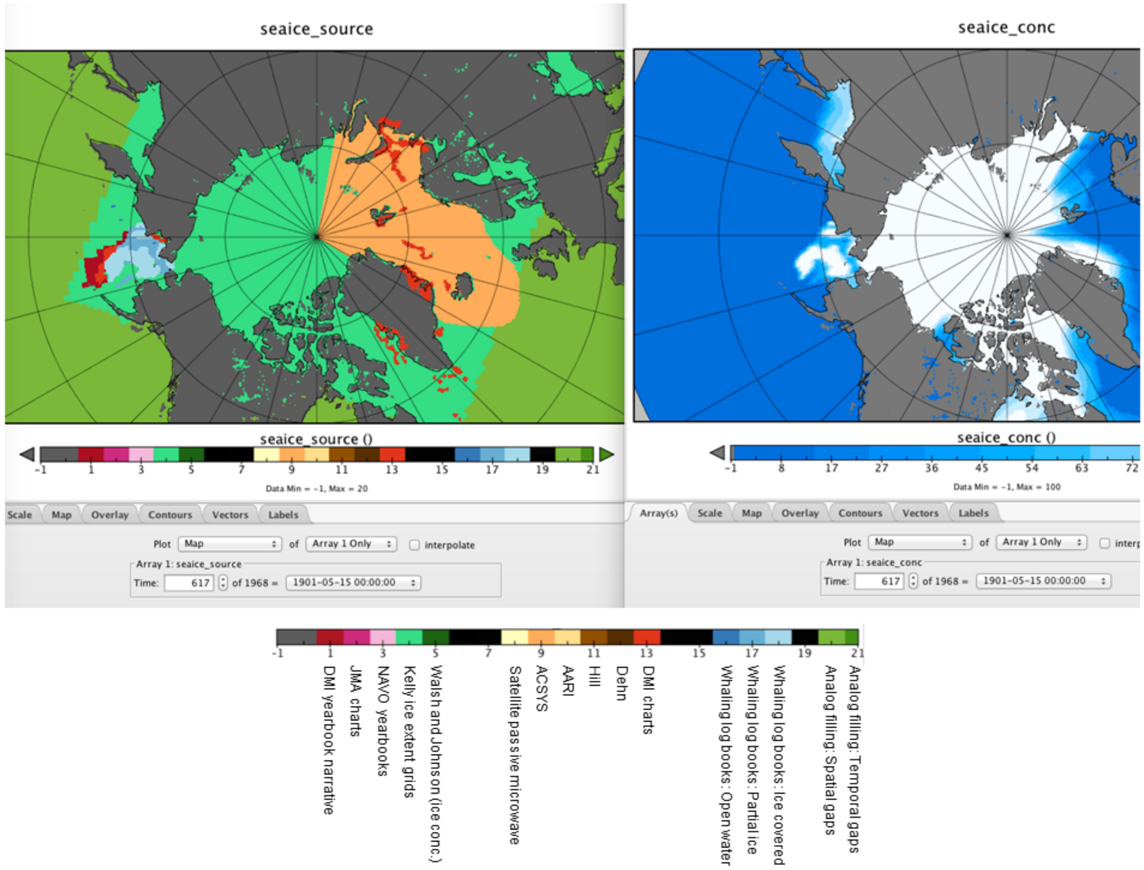

Figure 14. The first of source 4, the Kelly ice extent grids, occurs in May1901. This coverage is typical through 1952.

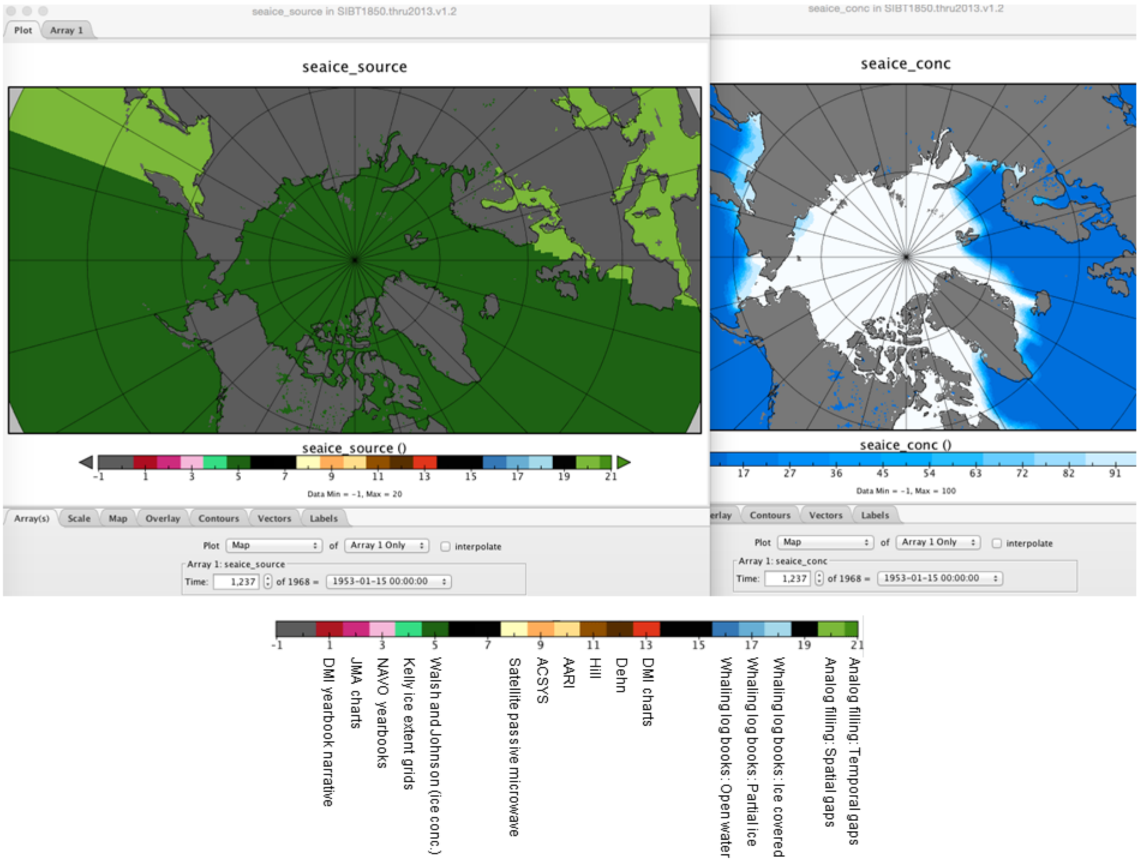

Figure 15. In January of 1953, the only source is Walsh and Johnson (source 5). Analog filling of spatial gaps is used elsewhere.