Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward, Version 2

Walsh, J. E., W. L. Chapman, F. Fetterer, and S. Stewart. 2019. *Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward, Version 2*. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: https://dx.doi.org/10.7265/jj4s-tq79.

Table of Contents

Gridded Monthly Sea Ice Extent and 1
Concentration, 1850 Onward, Version 2 1
1 Summary 3
2 Applications for these data 3
3 Product development background 4
4 Data files: Formats and descriptions 6
4.1 Gridded concentration and source netCDF file 6
4.2 Source coverage montage files (png format) 8
4.3 Sea ice extent and area files (csv format) 10
4.4 Ice extent browse files (png format) 10
5 The land mask and regional areas 10
6 Contributing data sources 11
7 Detailed source descriptions 14
7.1 Source 1: Sea ice concentration from satellite passive microwave data 14
7.2 Source 2: Data from the Danish Meteorological Institute (DMI) 14
7.3 Source 3: Data from the Dehn collection 15
7.4 Source 5: Data from the Russian Arctic and Antarctic Research Institute 16
7.5 Source 6: Data from NRC Canada, B.T. Hill 16
7.6 Sources 7, 8, and 9: Data from whaling ship log books 17
7.7 Source 11: Data from the Arctic Climate System Study (ACSYS) 21
7.8 Source 15: Land mask correction fill 21
7.9 Source 16: Land mask change fill 22
7.10 Source 17: Analog filling of spatial gaps 22
7.11 Source 18: Analog filling of temporal gaps 22
8 Method used to merge data sources 22
8.1 Ranking 22
8.2 Imposing temporal consistency 22
8.3 Estimating marginal ice zone concentration from ice-edge-only data 23
9 Filling data gaps with an analog method 24
9.1 Filling spatial gaps 24
9.2 Filling temporal gaps 25
9.3 Summary of gap field concentration value sources 26
9.4 How the DMI compilation is used: Estimating ice concentration from ice-edge-only data and filling spatial and temporal gaps in CW91 26
10 Errors and uncertainties in derived ice concentration from historical and satellite data sources 27
11 Product update history 30
12 Exploring the data with Panoply 30
13 Related data collections outside of NSIDC 32
14 Related NSIDC data collections 33
15 Acknowledgments 33
16 References 34
17 Document Information 36

17.1 Author 36
17.2 Publication Date 36
17.3 Document Revision Date 36
18 Appendix 1: Selected concentration and source field examples 37
19 Appendix 2: Known erroneous data 40
20 Appendix 3: Differences in sea ice extent and area between Version 1 and Version 2, and notes on the extent and area time series 43
20.1 V1/V2 differences 43
20.2 General characteristics of the time series 43

NOTICE
January 2020: A user has made us aware that the ice cover derived from the Danish Meteorological Institute Yearbooks may have been incorporated incorrectly into the Kelly grid source as it was used in this data set. The Kelly grids may have used the ice map from the subsequent month, rather than the intended month. The Kelly grids were used over April through August or September of 1901 -1956. We are investigating the problem, and will inform registered users of the outcome.

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.

Gridded Monthly Sea Ice Extent and Concentration, 1850 Onwards builds on an earlier data product (Chapman and Walsh 1991) 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 netCDF4 file as monthly sea ice concentration. The fields represent a single mid-month day, and are not monthly averages. Data are on a quarter-degree latitude by quarter-degree longitude grid, covering the northern hemisphere north of 30 degrees. In addition to the concentration variable, a corresponding source variable indicates where each of 18 possible sources is used. Regional and arctic-wide time series of ice extent and area in comma-separate-variable files, along with other ancillary files, are included.

2 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, Chapman and Walsh (1991), provided input to HadISST1 in the 1990s and early 2000s, and the present data set improves on earlier data 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.

Finally, the data product can be used in diagnostic studies of sea ice variability over time periods that pre-date the satellite era and for the placement of recent trends into a longer historical context (Walsh et al., 2016). The data product can also complement model reconstructions of Arctic sea ice spanning a century or longer (e.g., Schewiger et al., 2019).

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 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. Some of the newly added data have been published as separately available data products (e.g. Underhill, Fetterer, and Petersen 2014).

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).
- **November 1978-latest update:** All concentration values are from satellite passive microwave data.

In September 2019, Version 2 replaced Version 1. Version 2 updates the product through 2017 and makes other significant improvements detailed in the Product update history section.

Walsh et al. (2016) describe SIBT1850V1 and include a broad analysis of the changes in ice conditions that the time series reveals. Walsh et al. (2019) use SIBT1850V2 extent values in a regional analysis.

Figure 1 shows sea ice extent from SIBT1850V2 plotted by month. 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. Aspects of these timeseries are discussed in Errors and uncertainties and Appendix 3: Differences in sea ice extent and area between Version 1 and Version 2, and notes on the extent and area time series.

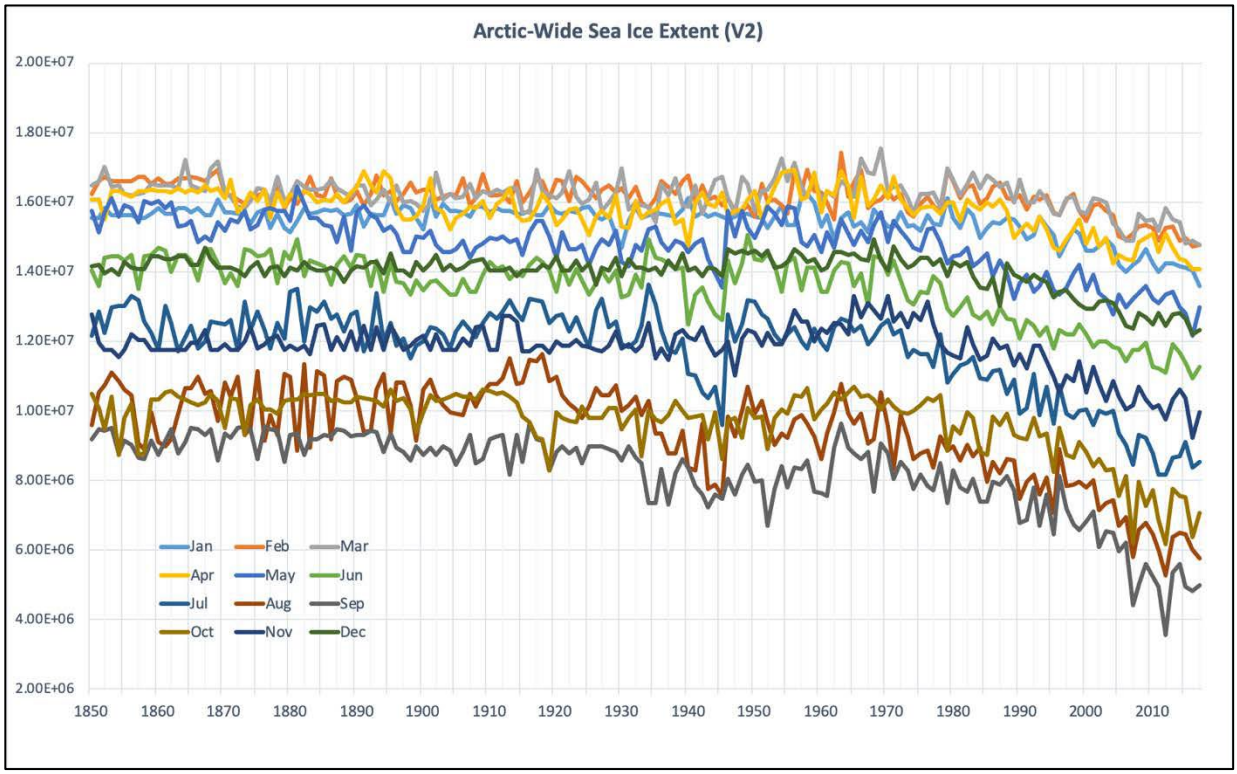


Figure 1. Sea ice extent in square km plotted for each calendar month, from 1850 through 2017. This plot was made using data in sibt_extents_v2.csv.

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 original processing code. Florence Fetterer, NSIDC, facilitated the acquisition and use of source data, led the development of data derived from the Danish Meteorological Institute (DMI) charts, and documented the data product. Scott Stewart led the processing behind the updates and improvements in Version 2, provided the custom Panoply color tables and instruction, and assisted with review and quality control of the data. This work was a collaboration, in part, with work at UAF that produced the Historical Sea Ice Atlas for Alaska.

4 Data files: Formats and descriptions

Table 1 lists the files that make up this data product with size and description. All the data files are provided in a single zip file at the following location:
https://noaadata.apps.nsidc.org/NOAA/G10010_V2/

Table 1. Data file and directory summary

File or directory name	Description
G10010_sibt1850_v2.0.nc	See Table 2
sibt1850_ conc_v2_to120.cpt	Color bar for sea ice concentrations (blue to white) for use with Panoply
sibt1850_source_v2_pano.cpt	Color table for the source of the data for use with Panoply
legend_for_sibt1850_source_v2_pano.cpt.png	Color-bar legend for concentration sources
source_montages	This directory holds 19 png files, one for each source plus a legend. These give a quick view of the spatial and temporal coverage of each source.
sibt_extents_v2.csv	Ice extent, one day per month, by region in km^2
sibt_areas_v2.csv	Ice area, one day per month, by region in km^2
sibt1850_browse	This directory holds 2,016 png files, one for each month's extent. These have names like sibt_ext_ease2_201712.png for December 2017 extent. Each file is 7 KB.

4.1 Gridded concentration and source netCDF file

Ice concentration data are in a single netCDF4 file, G10010_sibt1850_v2.0.nc, with a concentration variable and a source variable targeted to the 15th day of each month beginning with January 1850. Data are on a quarter-degree latitude by quarter-degree longitude grid, covering the northern hemisphere north of 30 degrees. The grid was chosen to be compatible with certain products developed at the Hadley Center, UK. NetCDF file variables are listed with short descriptions in Table 2.

Table 2. NetCDF4 variable description

Variable and dimension	Description
Grid-cell area (1D)	Area in km^2, decreasing with increasing latitude. Area was calculated using an oblate spheroid representation of the Earth's surface.
LandRegion_mask (Geo2D)	00 = "Ocean, no region specified"; 01 = "Beaufort Sea"; 02 = "Chukchi Sea"; 03 = "East Siberian Sea"; 04 = "Laptev Sea"; 05 = "Kara Sea"; 06 = "Barents Sea"; 07 = "Greenland Sea"; 08 = "Baffin Bay and Gulf of St Lawrence, on polar stereo grid"; 09 = "Canadian Archipelago"; 10 = "Hudson Bay"; 11 = "Central Arctic"; 12 = "Bering Sea"; 13 = "Baltic Sea"; 14 = "Sea of Okhotsk"; 15 = "Yellow Sea, on polar stereo grid"; 16 = "Cook Inlet, on polar stereo grid"; 17 = "Yellow Sea, off polar stereo grid"; 18 = "Baffin Bay and Gulf of St Lawrence, off polar stereo grid"; 19 = "Cook Inlet, off polar stereo grid"; 20 = "Land"; 21 = "Unnamed ocean on polar stereo grid"; Note that the "Land" mask covers lakes as well
latitude (1D)	Latitude in degrees North (30 to 90)
longitude (1D)	Longitude in degrees East (0 to 360)
seaice_conc (Geo2D)	Sea ice concentration in integer percent with values from 0 to 100, inclusive. Values of 120 indicate land.

seaice_source (Geo2D)	Identifies the source for each grid-cell's ice concentration value. 00 = "Land"; 01 = "Satellite passive microwave"; 02 = "Danish Meteorological Institute"; 03 = "Dehn"; 04 = "NAVO yearbooks"; 05 = "AARI"; 06 = "Hill"; 07 = "Whaling Records - Complete Sea Ice"; 08 = "Whaling Records - Partial Sea Ice"; 09 = "Whaling Records - No Sea Ice"; 10 = "DMI yearbook narrative"; 11 = "ACSYS"; 12 = "Walsh and Johnson"; 13 = "JMA charts"; 14 = "Kelly ice extent grids"; 15 = "Land mask correction fill"; 16 = "Change-of-land mask ocean"; 17 = "Analog fill - spatial"; 18 = "Analog fill - temporal";
time (1D)	Time assigned to the concentration field in days since 1850-01-01

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.

4.2 Source coverage montage files (png format)

The directory, source_montage, holds 19 png files, one for each source plus a legend. These give a quick high-level view of the spatial and temporal coverage of each source. They are intended to assist the user in determining whether the information in SIBT1850V2 is sufficient to meet the user's needs for a particular time period.

Each image is comprised of a 4 by 5 array of sub-images, one sub-image for each decade, with each sub-image itself comprised of images taking up 10 rows, one for each year, and 12 columns, one for each month. These year/month images are very small representations of source spatial coverage, displayed at a much-reduced resolution. In these, areas where the source is used appear in red. That is, if a particular source is used anywhere within the area covered by a reduced resolution grid cell, the entire area of that grid cell, about 900 km by 900 km, appears red.

Figure 2 illustrates how this works for Source 16. Source 16 flags where a gap between the new land mask and the coastline was filled. Figure 3 shows the montage source file legend with the montage for Source 11, the ACSYS ice edge data.

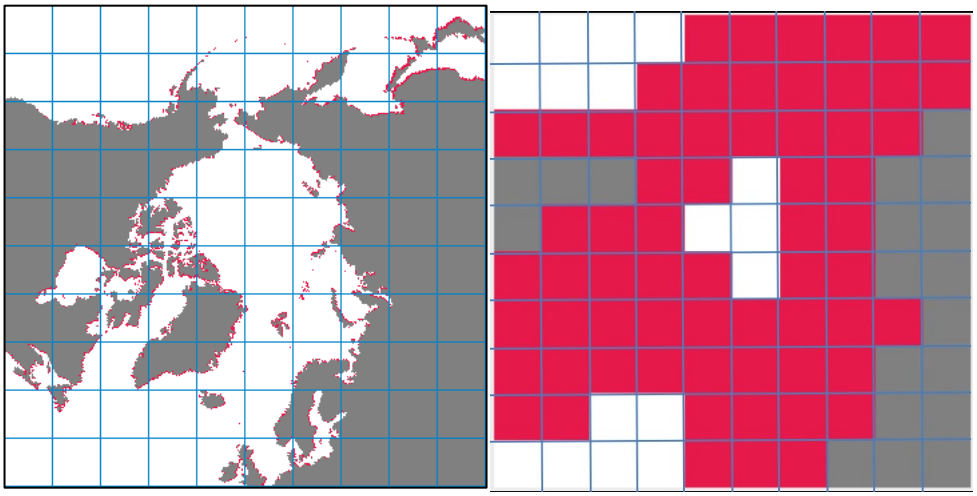


Figure 2. Construction of the montage year/month images starts with re-gridding the source fields to an equal area 360 by 360 grid. In the result for Source 16, (left), most coastlines are fringed in red showing where extrapolation to fill gaps was needed. This grid was further reduced to a 10 by 10 grid, (right), where if *any* point in the larger grid includes the source, the entire low-resolution grid cell will be positive, or red, for the source.

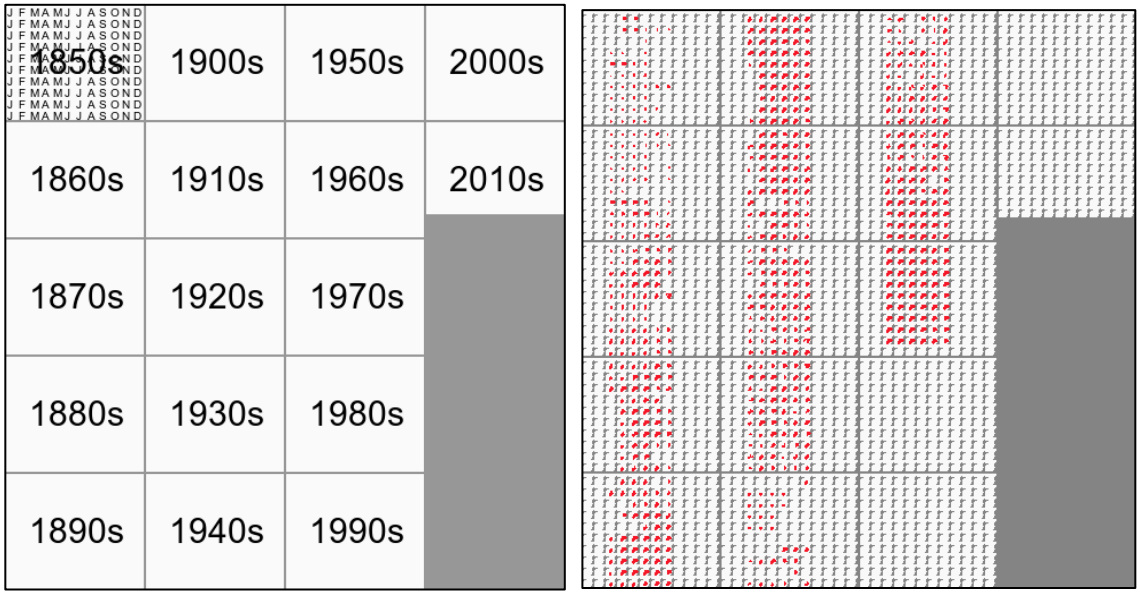


Figure 3. Montage image file legend (left) and source montage (right). This source montage, montage_11_ACSYS.png, shows at a glance that ACSYS data were used in the 1850s, and throughout most of the data set until the satellite era starts in 1978, but only in the central columns corresponding to the summer months.

The montages provide a quick way to view the temporal coverage of various sources in the data set. They also provide a high-level and compact view of the spatial coverage of each source; however, because of the way they are constructed, they may make it appear as if a source's spatial coverage is more extensive than it is. For a higher resolution view of spatial coverage, one can view the netCDF source field.

4.3 Sea ice extent and area files (csv format)

Files sibt_extents_v2.csv and sibt_areas_v2.csv hold ice extent and ice-covered area on the 15th of each month in comma-separated variable files. Extent is the summed area covered by all grid cells with an ice concentration of 15% or greater, while area is the sum of the area of each grid cell times the ice concentration at any value within that cell. (Ice concentrations in the netCDF file are in integer percent and range from 0 to 100). Arctic-wide and regional values are given. The total area encompassed by each region is provided in the third row of these csv files. The time series of regional and pan-Arctic ice extent are less subject to heterogeneities than are the time series of ice-covered area, as discussed further in Appendix 3.

4.4 Ice extent browse files (png format)

The directory sibt1850_browse holds 2,016 png files, one for each month's extent. Ice extent is mapped by all grid cells having an ice concentration value of 15 percent or greater. Files have names like sibt_ext_ease2_196203.png for March 1962 extent (Figure 4) where *ease2* refers to the NSIDC grid used in mapping the data.

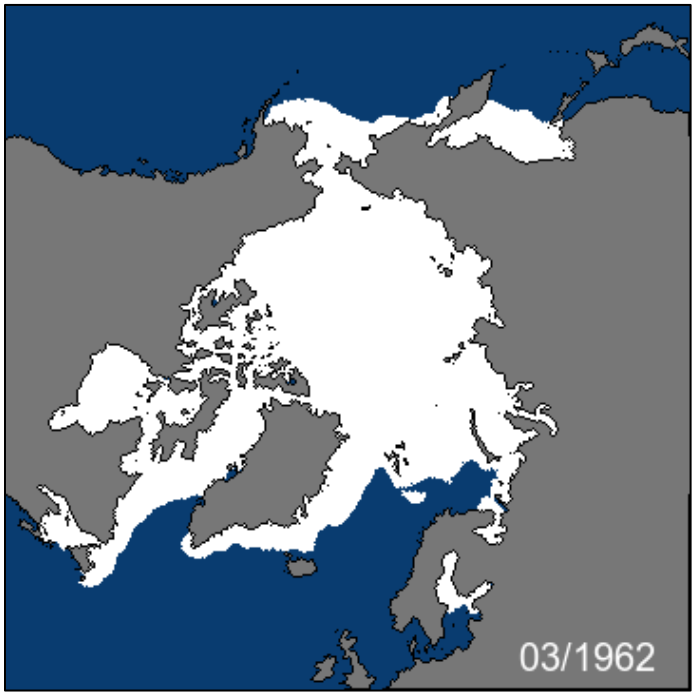


Figure 4. An ice extent browse file (sibt_ext_ease2_196203.png)

5 The land mask and regional areas

In Version 1.1, the land mask was made to be consistent throughout as follows: if any grid cell in any contributing data source was marked *land* at any point in time, that cell was masked as *land* throughout. This had the effect of artificially expanding the area covered by land and reducing the area that could have sea ice concentration values. In order to avoid unnecessarily omitting information for cells that are marked *land* by this method but that

do have an ice concentration value, a new land mask was created for V2 that is based on the Multisensor Analyzed Sea Ice Extent (MASIE) product land mask.

The 4 km resolution MASIE mask was mapped to the quarter-degree grid used for this product. In cases where a quarter-degree grid cell lay over a MASIE ocean/land boundary, one would naturally decide whether the large quarter-degree cell would be *ocean* or *not ocean* based on how much of each it covered, as determined by MASIE.

Rather than having the threshold for contributing surface be 50%, however, the threshold was chosen so that the resulting areas of the SIBT1850 ocean regions in G10010 are close to equal to those ocean region areas in MASIE. This was done in order to facilitate comparisons with other data products that use the MASIE regions and because MASIE regions are receiving increased use as Arctic sub-regions. The area of each region is given in sq km in the third row of sea ice extent and area comma-separated-variable files.

In Version 1, the land mask did not mask all lakes. The Version 2 mask does mask out lakes.

6 Contributing data sources

All data sources are listed in Table 3, along with a short description. Data sources are numbered as they are identified in the netCDF *seaice_source* variable.

Data sources 15, 16, 17, and 18 are essentially flags, showing where gaps caused by re-aligning sources or imposing a higher resolution land mask in the process of creating Version 2 were filled (sources 15 and 16 respectively), or where cells were filled using the spatial or temporal fill method (sources 17 and 18 respectively).

Table 3 has a column that indicates whether concentrations from a given source include concentrations that are "not observed." These "not observed" concentrations are values that are estimated when only the ice edge position was observed or values that come from one of the gap-filling methods. We label concentrations from the DMI charts (Source 2) and from whaling ship observations (Sources 7, 8, and 9) 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 10, 13, 4, 14, and 12 (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 (Source 14), as do the Walsh and Johnson grids (Source 12), even though both sources include information that may have come from Source 10, 13, or 4, and other sources named in the cited references. Sources 10, 13, and 4 are not replaced by Source 14 (Kelly) or 12 (Walsh and Johnson) when 10, 13, and 4 occur within the domain of the Kelly or Walsh and Johnson source grids.

There may be cases where sources were identified erroneously when CW91 was constructed, and this has carried through into the present product. For example, in December 1971 and January 1972, the source for much of the Sea of Okhotsk is shown as Source 4 (NAVO Yearbooks), when in fact these yearbooks did not have observations for this area. Source 12 (Walsh and Johnson) supplied the concentrations in this area. (John Walsh, personal communication, 2015). The concentration values are not affected. With the exception of Sources 10, 13, 4, 14, and 12 (DMI yearbook narratives, JMA charts, NAVO yearbooks, Kelly ice extent grids, and Walsh and Johnson ice concentration grids) sources are described in more detail in the Detailed source descriptions section.

Table 3. Data Source Description

Source reference number and short name	Short description	Existence of "non-observed" data?	First year, last year source occurs
1. Satellite passive microwave	NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration "merged" values 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 (although the "pole hole" is filled).	1978,2017
2. Danish Meteorological Institute derived ice concentration	These are concentration fields derived from an interpretation of the DMI maps. Sources 10 (DMI yearbook narratives), 14 (Kelly ice extent grids), and 2 (DMI ice concentration) all include or are based on historical information from DMI.	No. Concentrations were inferred from direct observations.	1901,1956
3. Dehn	From a digitized subset of scanned analog charts converted to shapefiles with concentration information.	No. Analog concentration information was digitized for this source. See the documentation for more information.	1954,1978
4. NAVO yearbooks	The Naval Oceanographic Office (NAVOCEANO or NAVO) 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 17.	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
5. AARI	Digitized ice charts from the Russian Arctic and Antarctic Research Institute. The gridded charts include ice concentration information.	No. See the source product documentation for how these were compiled.	1933,1978
6. 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

7. 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. Concentrations were inferred from direct observations, in a manner similar to but less precise than that used for Source 2 (DMI derived concentrations). See the section Sources 7, 8, and 9: Data from whaling ship log books for more information.	1850,1919
8. 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 7.	1850,1918
9. Whaling log books: open water	From a mapping of whaling ship logbook entries that indicated open water.	No. As for Source 7.	1850,1919
10. DMI yearbook narrative	The DMI yearbooks included summaries of ship reports. These summaries 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 14, Kelly ice extent grids). The first example of this source can be seen in the North Pacific in April 1901. The DMI observations indicated "ice free" where Source 10 appears in the netCDF file. This is reflected in the corresponding concentration field. See Figure 15. Sources 10 (DMI yearbook narratives), 14 (Kelly ice extent grids), and 2 (DMI ice concentration) all include or are based on historical information from DMI.	Yes. Where an edge position is observed, concentration is estimated from edge position using the method of CW91.	1901,1938
11. 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
12. 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 4) contributed to the pan-Arctic data set described in these publications. See Figure 19.	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
13. JMA charts	Information derived by Walsh (1978) from analyses done by the Japan Meteorological Agency appear in some winter and spring months of years 1968-1978, with exception of 1970 and 1972. See Figure 16.	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

14. 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 18. Sources 10 (DMI yearbook narratives), 14 (Kelly ice extent grids), and 2 (DMI ice concentration) all include or are based on historical information from DMI.	Yes. Concentration is estimated from edge position using the method of CW91.	1901,1956
15. Land mask correction fill	Misaligned V1 source fields were aligned; any resulting gaps were filled by extrapolation and labeled as Source 15	Yes.	1850,1978
16. Land mask change fill	Application of new V2 land mask left gaps between ice and coastline; grid cells in areas where this occurred were filled by extrapolation and labeled as Source 16.	Yes.	1850,1978
17. Analog filling of spatial gaps	Spatial gaps in a given grid were filled with best analog representations of the given month.	Yes.	1850,1978
18. Analog filling of temporal gaps	Temporal gaps of entire missing grids were filled with analog representations of the missing month; analogs years are from nearest month with at least some data.	Yes.	1850,1952

7 Detailed source descriptions

Sources 10, 13, 14, and 12 (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.

7.1 Source 1: Sea ice concentration from satellite passive microwave data.

These data are from *NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 3*. The Climate Data Record (CDR) blends output from both the NASA Team and NASA Bootstrap algorithms for deriving ice concentration. For more information, see the product documentation (Meier et al. 2013). These satellite passive microwave-derived data are the only source used in SIBT1850 from November 1978 onward.

The mid-month (15th or 16th) daily field of the parameter *goddard_merged_seaice_conc* is taken from the CDR. The "pole hole", or area around the North Pole that is missed by the satellite instrument swath, is filled using the average of orthogonally-neighboring grid cells on the NH polar stereo grid before the data are re-projected onto the SIBT1850 quarter-degree grid. Note that pole hole grid cells that have been filled with a concentration value in this way are not marked by a special source flag. They are marked as Source 1.

7.2 Source 2: 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 2 and Source 10 both originate with DMI, but concentrations from Source 10 are as described in the documentation for Chapman and Walsh (1991).

7.3 Source 3: 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 5 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.

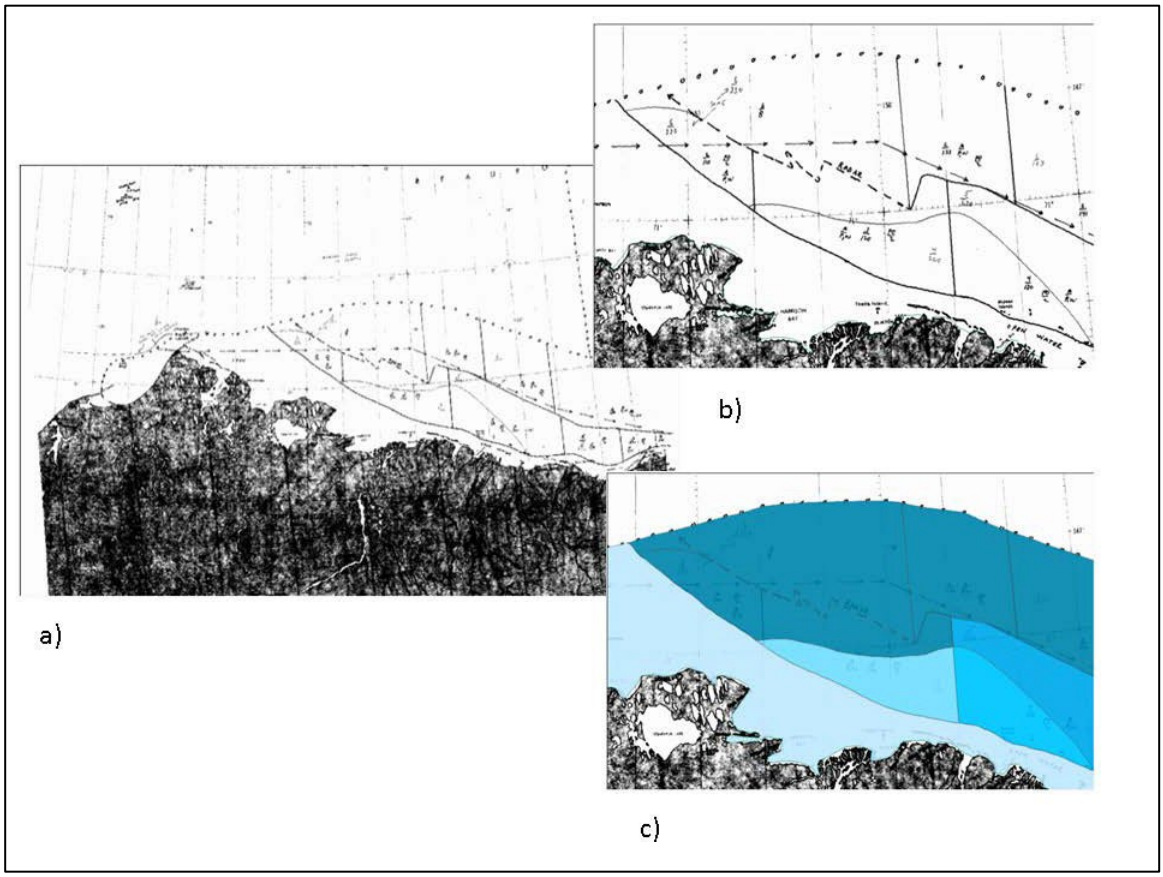


Figure 5. This example illustrates how the Dehn charts were processed at UAF. a) A Dehn chart scan for August 1954. These were provided to UAF-ACCAP as TIFF files. b) A close-up section of the chart. Note that some areas have a great deal of detail. Deciphering notation or handwriting could be challenging, 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.

7.4 Source 5: 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.

7.5 Source 6: 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 6) shows considerable variability. More information about these data may be found in Hill (1999) and Hill and Jones (1990).

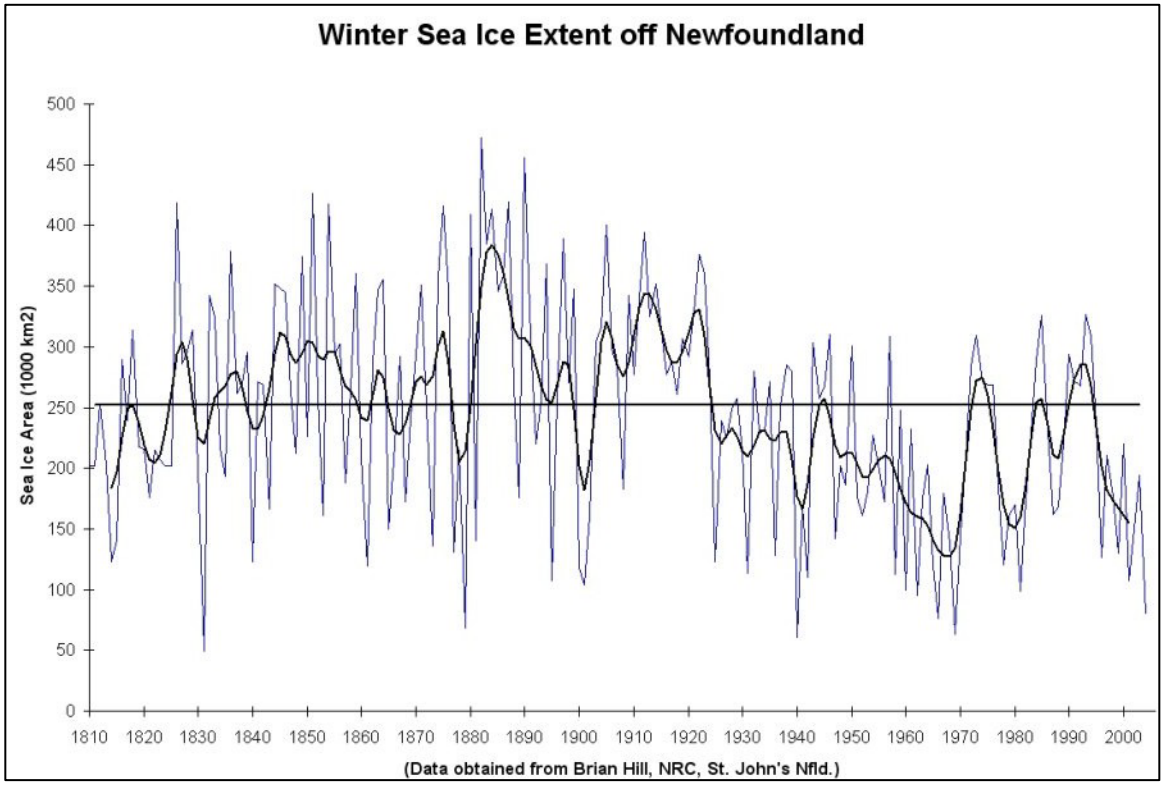


Figure 6. Yearly sea ice extent east of Newfoundland, Canada (data from Hill (1999) and Hill and Jones (1990).

7.6 Sources 7, 8, and 9: 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 7 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 considered. In examples from June of 1852 and 1864, *ice* and *no ice* observations are more mixed (Figure 8). 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 9 illustrates mapping the June 1850 compilation at an intermediate stage. In the resulting product (Figure 10), some areas of ice (yellow) are included. In SIBT1850, ice polygons are labeled as Source 7, mixed ice polygons are Source 8, and open water polygons are Source 9.

Note that the Source 7 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 8 does not explicitly map areas of partial concentration for the month in question.

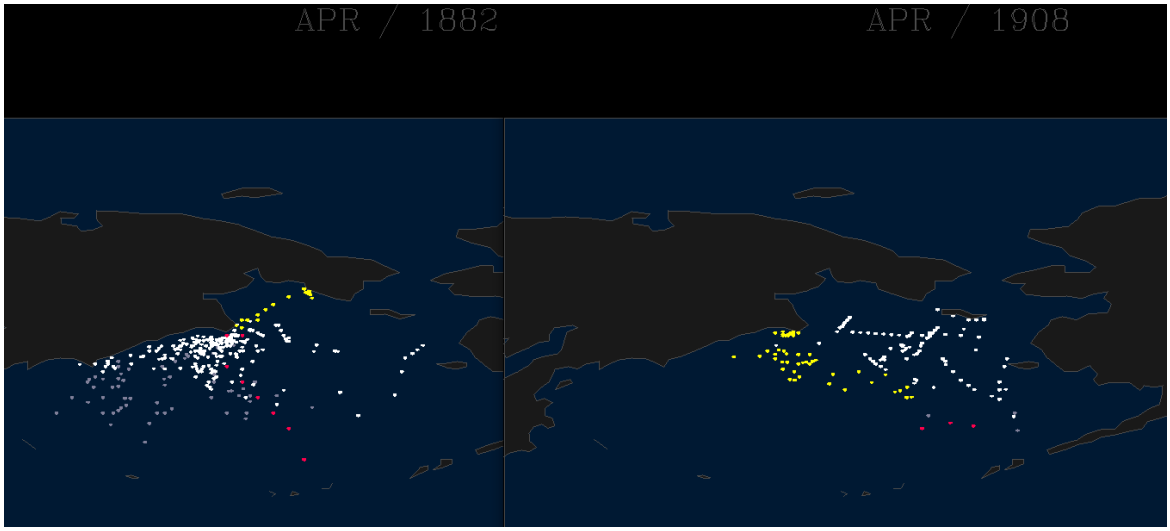


Figure 7. April 1882 (left) and 1908 (right). Observations of ice are yellow or, if from the 10-year window of Aprils around the month in question, white. Open water observations are red or grey.

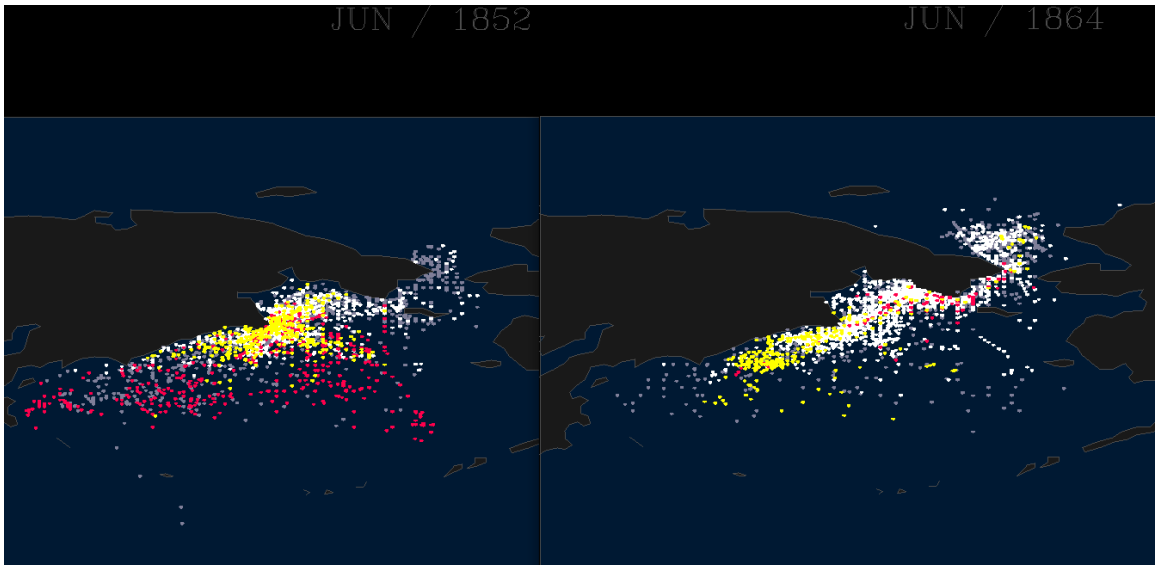


Figure 8. June 1852 (left) and 1864 (right). Observations of ice are yellow or, if from the 10-year window of Junes around the month in question, white. Open water observations are red or grey.

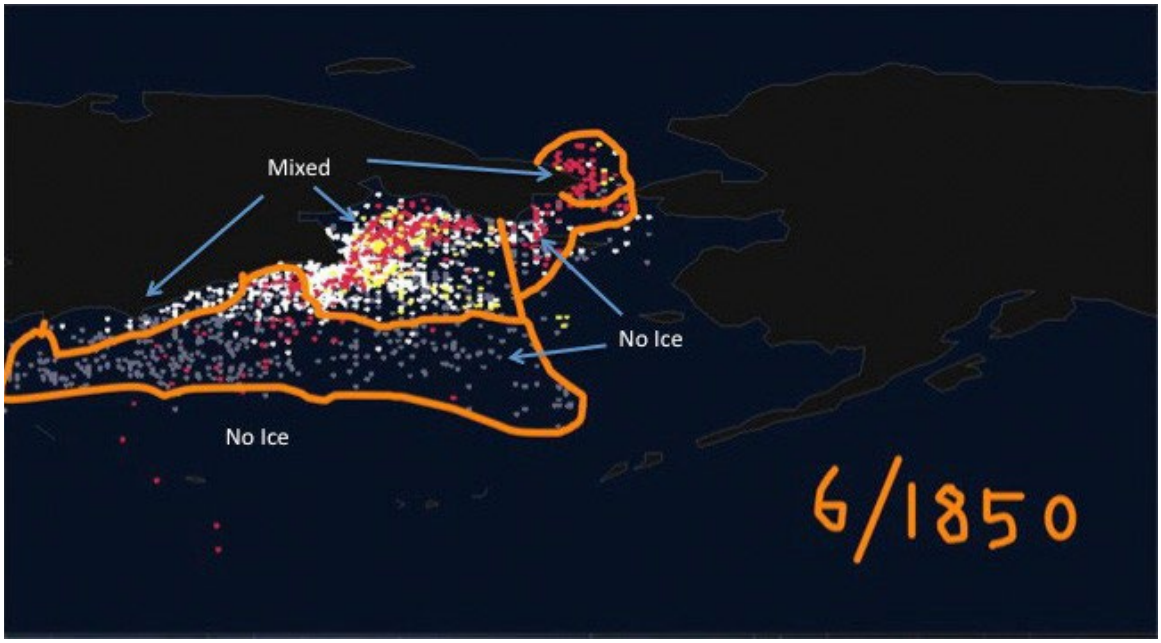


Figure 9. The GIMP image manipulation application was used to draw around areas of ice, open water, or mixed ice and open water observations for June 1850.

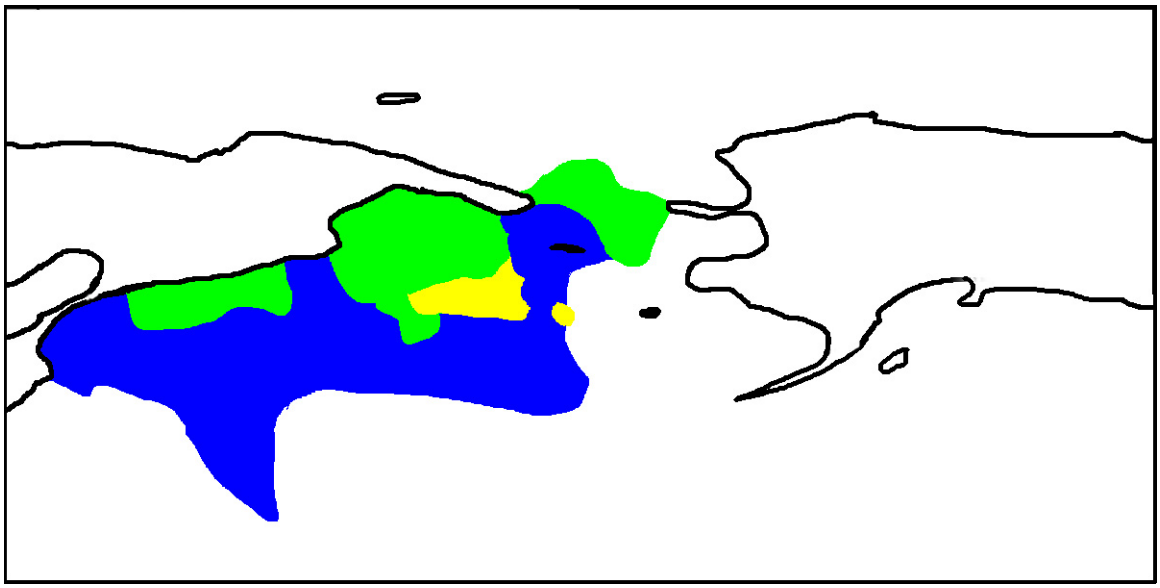


Figure 10. This image shows the results of the June 1850 mapping where these are made into color-coded polygons marking *ice* in yellow, *no ice* in blue, and *mixed* in green.

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 11), 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.

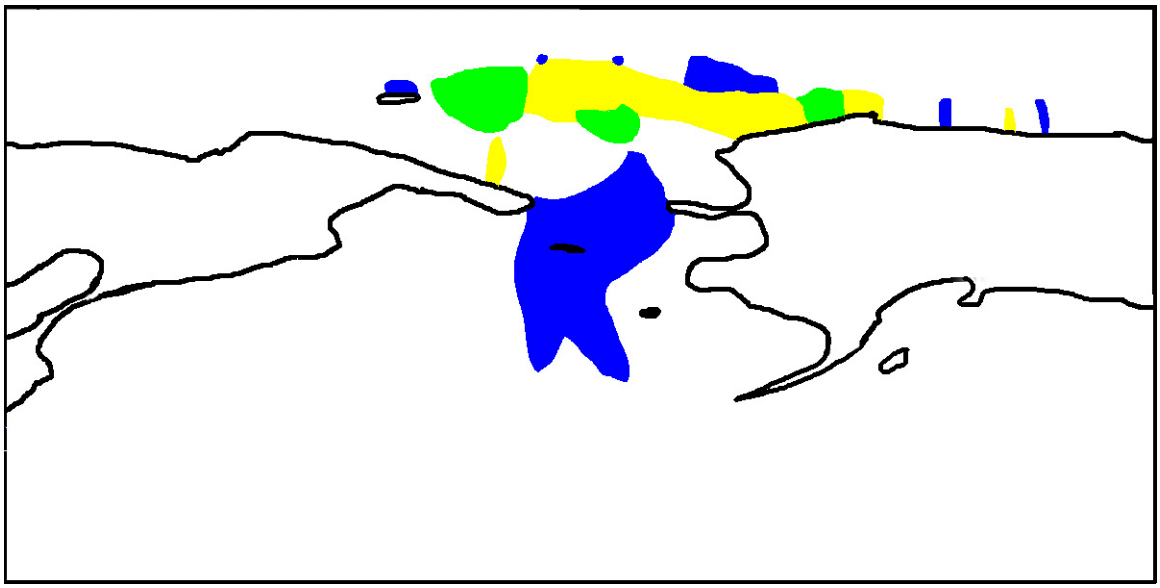


Figure 11. Polygons for ice cover in September 1910. While ice (yellow) north of open water (blue) in this example is unrealistic, altering the product would introduce a higher order of subjectivity into the resulting analysis.

Rather than use a single source designation for all whaling data derived concentration estimate categories, we retained the three as separate sources, as it may be helpful to interpreting some features in ice concentration fields.

7.7 Source 11: 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 Estimating marginal ice zone concentration from ice-edge-only data.

7.8 Source 15: Land mask correction fill

The occasional mis-registration of three sources (ACSYS, AARI, and Hill) in V1.1 became apparent when fields were inspected as part of the process of applying the new land mask. Shifting fields to properly align them with the land mask and with each other left gaps; grid

cells flagged with Source 15 map gaps that have been filled by extrapolating the concentration of adjoining ice-filled cells: the concentration field is dilated to fill gaps, and values are averaged if more than one grid cell is adjacent to a grid cell without data. Figure 23 in Appendix 2 illustrates a case of misaligned source fields.

7.9 Source 16: Land mask change fill

When the new land mask described in Section 5 replaced the V1.1. land mask, gaps between sea ice and land emerged. Grid cells flagged with Source 16 map gaps that have been filled by extrapolating the concentration of adjoining ice-filled cells. These gaps fringe coasts throughout the data set, as Figure 2 illustrates.

7.10 Source 17: Analog filling of spatial gaps

See the section on Filling spatial gaps.

7.11 Source 18: Analog filling of temporal gaps

See the section on Filling temporal gaps.

8 Method used to merge data sources

Concentration fields from the various sources were blended or merged using processing code that was developed and run at UIUC. A ranking hierarchy determined what source to use when more than one source was available for a given time and location. Each source was made temporally consistent before the merging began.

8.1 Ranking

The rank of each source was assigned based on a subjective assessment of the perceived reliability of the source in question. Criteria for the ranking included the level of detail, temporal continuity, supporting information (e.g., narratives), and types of information used (e.g., aircraft surveys vs. coastal or ship reports). Source numbers correspond inversely with source rank. For example, if AARI data (Source 5) were available, those data took the place of any other data that might be available within a given month from sources 6 and greater. Data sources 5 through 11 rarely overlap in space and time.

Source number 1, sea ice concentration from satellite passive microwave on the 15th or 16th of each month, is used exclusively from November 1978 on. October 1978 has three sources: AARI (Source 5), Dehn (Source 3) and passive microwave (Source 1). For this year/month only, the passive microwave sea ice concentration values are from 26 October, when the passive microwave sea ice concentration record begins, rather than the 15th of the month, and are from the NASA Team algorithm rather than the CDR.

8.2 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 7, 8, and 9: Data from whaling ship log books.

For the Dehn charts (Source 3) coverage was frequent enough that linear 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.

We use a mid-month daily value instead of a 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.

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

Two sources, Hill (Source 6) and ACSYS (Source 11), are *ice-edge-only* data sets. The source data are in the form of a series of ice edge positions. DMI concentration (Source 2) has an inferred ice edge position where the ice concentration was not observed.

To make the most of these data, we assigned an estimated concentration to 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. 2002 was chosen as the ending year in order to avoid a bias from the diminished ice cover of the past 10-15 years.

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 considers seasonal, as well as regional, variability. Figure 12 illustrates a resulting ice concentration gradient with an example from the ACSYS ice edge data collection (Source 11).

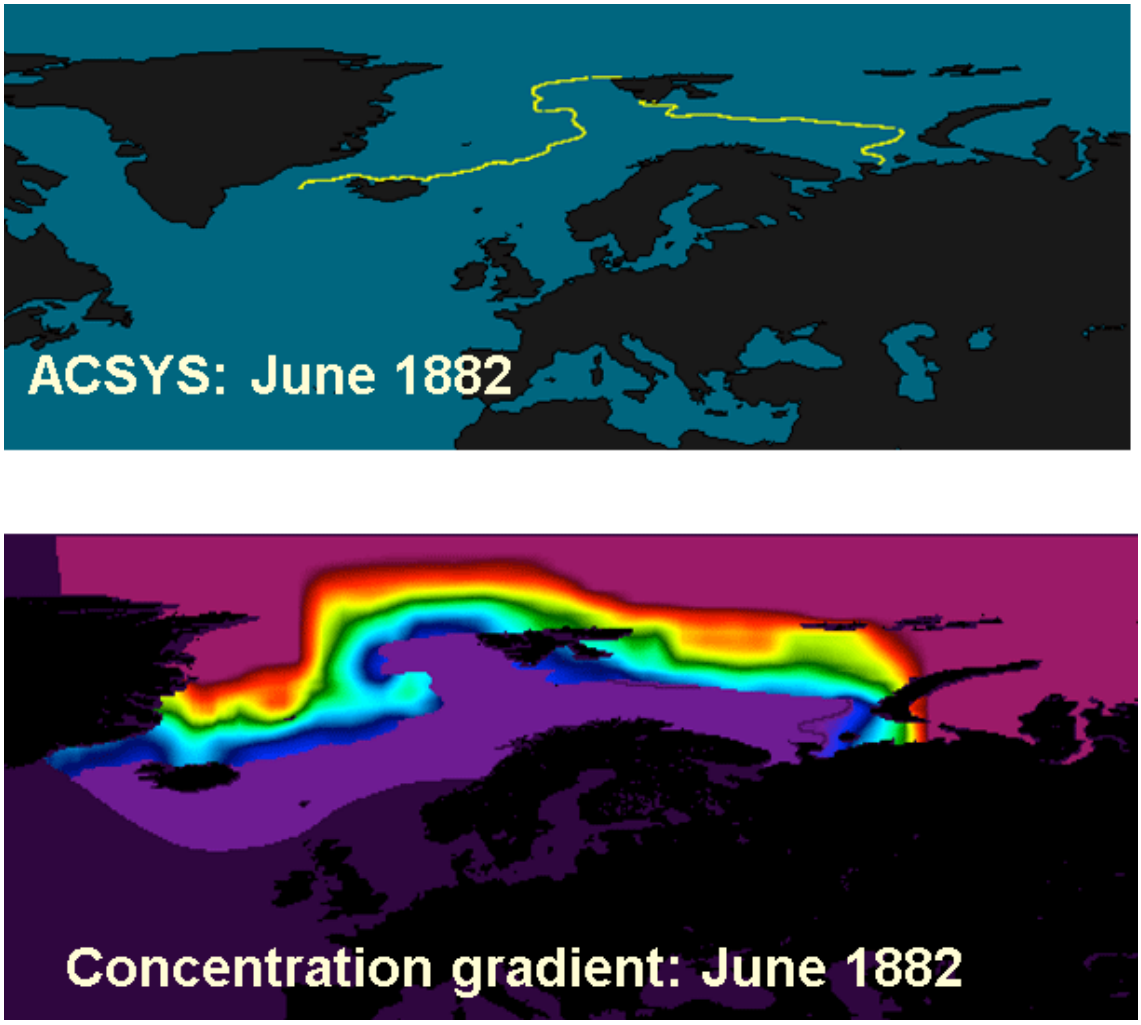


Figure 12. Historical data products that give only the ice edge position are used to infer a concentration within the edge by applying a gradient function. In this example, the information in the ACSYS ice edge product for June 1882 (top) is used to infer a concentration gradient within the edge (bottom).

9 Filling data gaps with an analog method

9.1 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 U.S. 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 were 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 are labeled as Source 17 (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 the non-observed part of the gap field. Such grid cells are also labeled as Source 17 (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.

To summarize, missing values in the gap field are filled with the average concentration values for those grid cell points from three, two, or one analog field.

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.

9.2 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 Source 18 (analog filling of temporal gaps) for each grid cell for February of 1853 in this hypothetical example. In most instances of temporal gap-filling, the temporal gap was small (1, 2 or 3 months). The 1920s were an exception: in the 1920s and into the 1930s, Septembers, Octobers, Novembers, and Decembers lacked data. In the 1850s and 1860s, November, December, January and February lacked data. Montage 18 illustrates the decades and months that required filling temporal gaps, and Figure 24 includes shading to indicate those months and years.

9.3 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 17 (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 17 (analog filling of spatial gaps) or 18 (analog filling of temporal gaps).
3. As above, but using analogs from only the post-1978 `(satellite) era. These are flagged with source number 17 (analog filling of spatial gaps)) or 18 (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.

9.4 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 2 (concentrations derived from DMI charts), 10 (DMI yearbooks) and 14 (Kelly ice extent grids). This section describes how Source 14 (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 2: Data from the Danish Meteorological Institute for information on how sources 2 and 10 use the DMI compilations.

Scanned versions of the DMI analog maps are available from NSIDC as *Arctic Sea Ice Charts from Danish Meteorological Institute, 1893 - 1956*. 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 14 (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 6 (Hill) and 11 (ACSYS).

10 Errors and uncertainties in derived ice concentration from historical and satellite data sources

Uncertainty may result when historical sources cover the same area and time period. Fortunately, relatively few cases arise in which sources conflict. We chose a ranking method for merging such source data 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 layer and by reading the sections of the documentation pertaining to those sources.

In merging 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.

Those who wish to subset the data for particular months in the historical period are strongly encouraged to view data sources for the individual months. This can be done using the Panoply application. See Exploring the data with Panoply.

Some of the variability and features that result from particular characteristics of source data and from the analog fill method are evident in the images of concentration and source from 1874 (Figure 13a). April, May, June, and July show a wedge of low concentration ice extending from the Barents Sea up to the North Pole. The shape is reflected in the ACSYS source data for these months. The August ACSYS source data are misaligned, and because of this, the August concentration field has odd and unrealistic features. This is described in Appendix 2: Known erroneous data.

In September of 1874, the overall central arctic concentration looks lower than in either August or October. The source for these concentrations is analog spatial fill (Source 17), and it may be that the abruptly lower central Arctic concentration is a result of the analog that was selected.

Despite the variability and features in the concentration fields, plotting area and extent over the course of 1874 produces fairly smooth curves (Figure 13b).

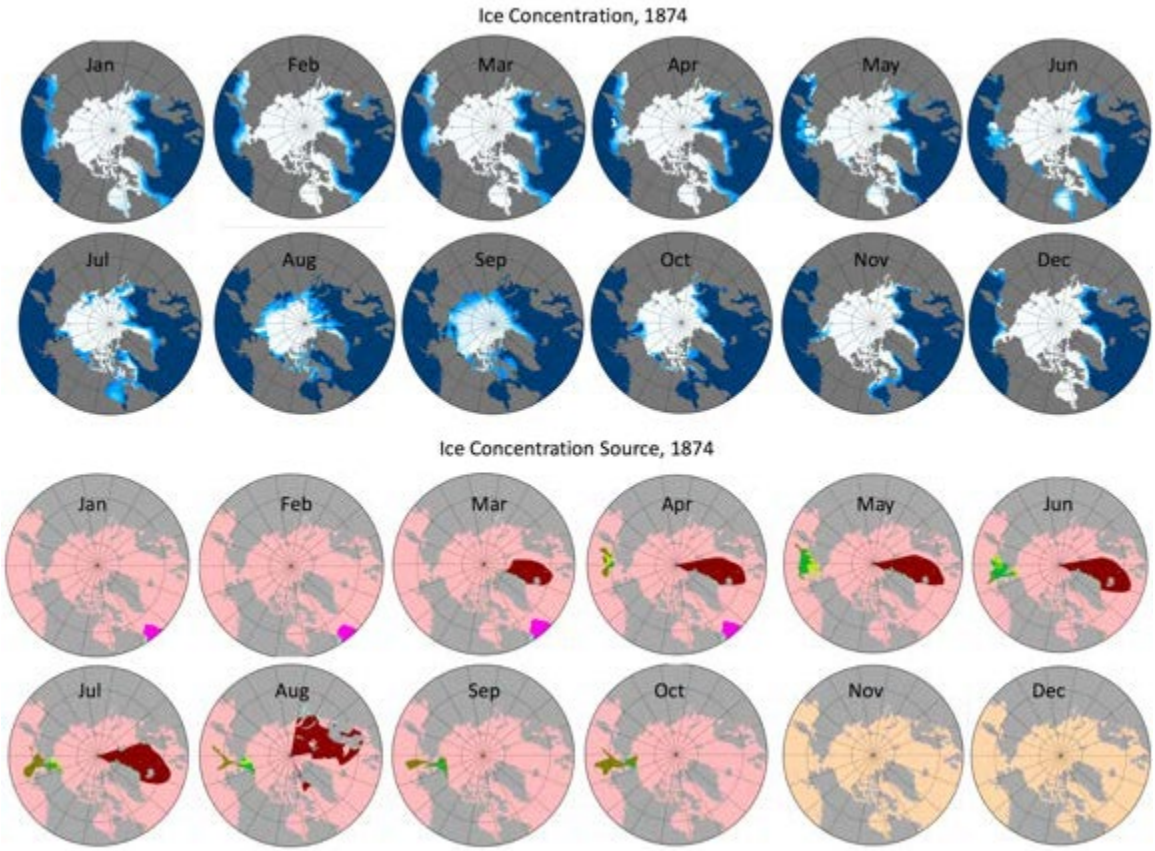


Figure 13a. Ice concentration fields from each month in 1874 (top) and corresponding source fields (bottom). Figure 14 gives the source color legend. Analog fill (light pink) covers most of each field until November and December, when temporal fill is the only source.

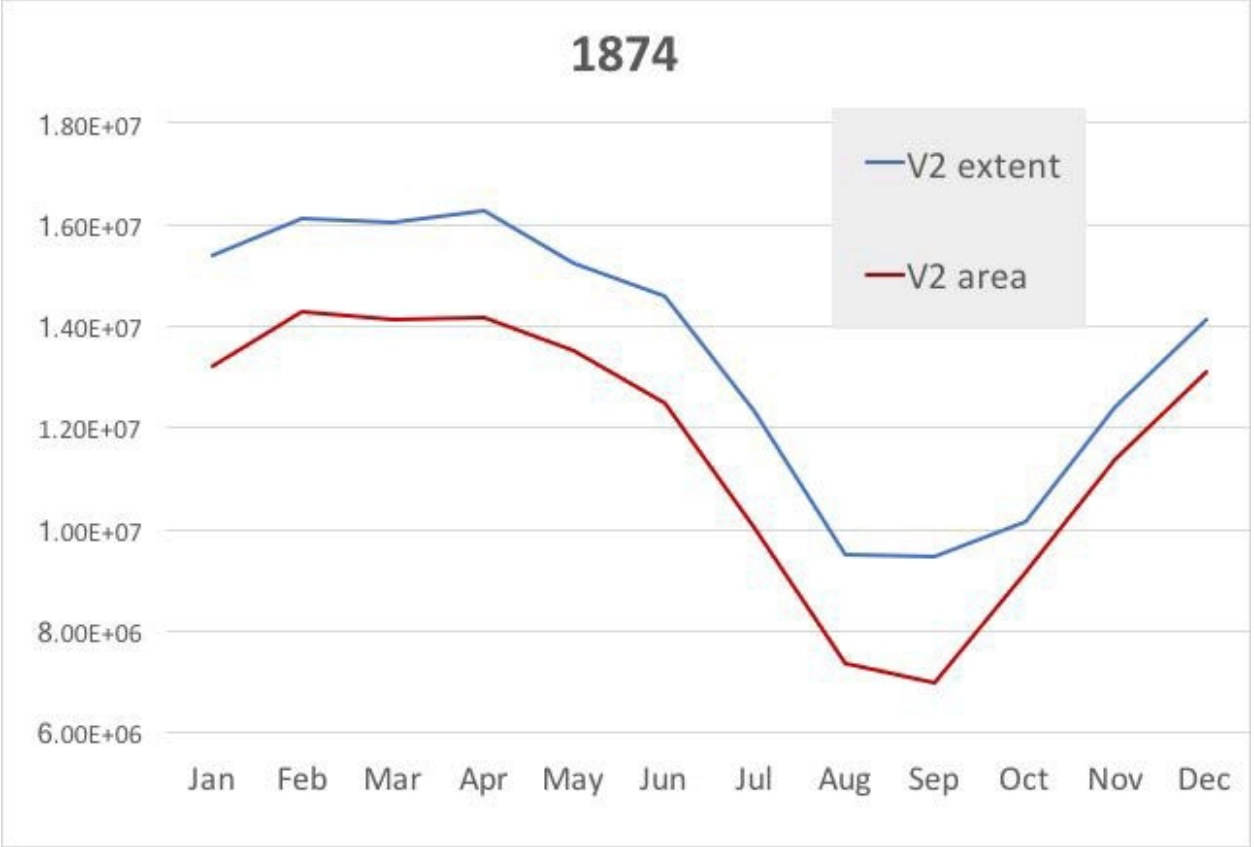


Figure 13b. Sea ice extent (blue) and area (red) in sq km for year 1874.

For the ice concentrations in the satellite data era, we use the *goddard_merged_seaice_conc* variable of the NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 3, or CDR. Ice concentration values from this product show relatively good agreement with ice concentrations derived from operational ice service charts. The User Guide documentation has a number of links to assessments of the product, including Sea Ice Concentration: NOAA/NSIDC Climate Data Record (Meier and NCAR 2014). As with most passive microwave derived ice concentration products, land spillover, or false detection of ice along coasts, can be a problem. The problem is reduced for the CDR through the use of Polar Stereographic Valid Ice Masks Derived from National Ice Center Monthly Sea Ice Climatologies. There are 12 masks, one for each month. In addition, the *goddard_merged_seaice_conc* fields use NASA Team and Bootstrap "final" data that have been quality controlled manually by inspection at Goddard Space Flight Center before being delivered to NSIDC and incorporated into CDR processing.

As shown in Appendix 3, there are heterogeneities in the ice area/ice extent ratio as one moves from the pre-satellite period to the passive microwave period. More specifically, the area is a larger fraction of extent in the satellite data. While the reasons for it are not apparent, the extents are considered much more reliable than the concentration-dependent areas because much of the information from earlier decades was from ships or coastal observations.

It is important to keep in mind that the monthly fields are not monthly averages. They are for the 15th of the month when satellite data is the only source, and for an approximation of the 15th of the month in the historical period.

11 Product update history

With Version 2, the following changes have been made. See Appendix 3: Differences in extent and area between Version 1 and Version 2 for more information.

Improvements to data product quality:

- A more accurate land mask is used. This land mask is consistent throughout the data product.
- Misaligned source fields have been aligned.
- Additional source numbers identify data cells that have been filled where gaps resulted from registering misaligned sources or imposing a new land mask.
- Source numbers have changed. In V1, source number loosely corresponded to rank order. Now, in V2, source number reflects correct inverse rank order throughout.
- Revised data source and concentration color tables compatible with Panoply are provided.
- Revised "montage" images are provided. These show an at-a-glance overview of coverage by data source.
- The satellite era portion is more accurate:
    - SIBT1850 V2 uses Version 3 of the CDR; this version of the passive microwave data employs different "valid ice" filters, resulting in fewer erroneous "false positive" sea ice detections.
    - The CDR *goddard_merged_seaice_conc* fields are used. These are based on data that have received more stringent quality control than the *seaice_conc_cdr* fields used in V1. (The User Guide for V1 stated that the merged CDR product was used, but this was found to be incorrect).

Additions to the data product:

- Time series of monthly sea ice area and extent are provided in csv files
- Regional area and extent are included in the csv files along with arctic-wide sea ice area and extent
- Grid cell areas are included as a variable (layer) in the netCDF file
- Region designators are included with land mask as a variable (layer) in the netCDF file
- Browse images of ice extent are provided in .png format

12 Exploring the data with Panoply

The figures in this documentation that illustrate source and concentration fields from the netCDF file were created using the NASA Panoply data viewer Version 4.10.9 and custom color tables sibt1850_conc_v2_to120.cpt and sibt1850_source_v2_pano.cpt.

When displaying source and ice concentration data with Panoply the following guidelines may be helpful. These were tested on a Mac but Windows and Linux procedures are likely to be similar.

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. From the toolbar, choose "Open…". Find and select the SIBT1850 netCDF data file.
3. To create a 2D plot, choose either seaice_conc or seaice_source and click "Create Plot". Allow it to create a georeferenced Longitude-Latitude plot.
4. From the "Map" tab, choose Projection: "Stereographic" and set Center on "-90"E and "90"N with an Edge Angle (radius) of perhaps 40 degrees to start. Uncheck "Fill corners". (This will make a map of all the data north of 50 degrees. The data grid extends south to 30 degrees, so increase the Edge Angle if you want to see ice that may be further south.)
5. From the "Array(s)" tab, unclick the "interpolate" box, then select a date to display. (Go to the drop-down and type the year to jump to the desired year, and then select a month within the year using the drop-down menu.)
6. To apply the provided color table (color bar) to a sea ice source graph, choose the custom source color table by opening it.
    - Choose File->Open from the main Panoply toolbar and find the file "sibt1850_source_v2_pano.cpt"
    - Choose "Okay" to import the color table to the support library
    - Activate the color table by selecting the lower tab "Scale" and choose "sibt1850_source_v2_pano.cpt" from the dropdown menu
        - Change the Scale Range to Min: -1 Max: 19 (Do not choose "Fit to Data")
        - Divisions: Major: 20 and Minor: 1
        - *Source numbers will not be matched correctly with scale bar colors unless the above steps are taken.*
    - Choose the Overlays tab and change Overlay 1 to None. This step removes the Panoply default overlay. The default overlay does not match the SIBT1850 land mask well.
7. To apply the provided color table (color bar) to a sea ice concentration graph, adapt the steps above accordingly:
    - Choose File->Open from the main Panoply toolbar and find the file "sibt1850_ conc_v2_to120.cpt"
    - Choose "Okay" to import the color table to the support library
    - Activate the color table by selecting the lower tab "Scale" and choose "sibt1850__ conc_v2_to120.cpt" from the dropdown menu. With this color bar, 0% - 15% are the same dark blue color, 15% - 100% are in 5% increments, 101% - 120% are grey (like land)
        - Change the Scale Range to Min: 0 Max: 120 (Do not choose "Fit to Data")
        - Divisions: Major: 12 and Minor: 1
    - Choose the Overlays tab and change Overlay 1 to None. This step removes the Panoply default overlay. The default overlay does not match the SIBT1850 land mask well.

Users may wish to set the Panoply defaults for displaying either source or concentration data. To this, go to Preferences under Panoply in the top bar. Then, modify the settings under General, Scale, and Lon-Lat Plots so that these defaults are in accordance with the settings above. Or, choose "Save plot settings to Preferences" under Plot in the toolbar.

An image file (legend_for_sibt1850_source_v2_pano.cpt.png) that serves as a legend for the color bar is provided (Figure 14).

Figure 14. The legend for the source color bar (legend_for_sibt1850_source_v2_pano.cpt.png).

13 Related data collections outside of NSIDC

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 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.

14 Related NSIDC data collections

These are the data products that contributed to or are closely related 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 3

15 Acknowledgments

NOAA's Climate Program Office provided support for the development of this data product through Grants # NA11OAR4310172 and NA16OAR4310162. 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.

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 had the Dehn analog charts scanned and documented them.

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. Axel Schweiger, University of Washington, explored the data and found examples of erroneous fields that assisted us in planning Version 2. Cathy Smith, with NOAA Earth Systems Research Laboratory, greatly assisted us by improving the structure of the netCDF file.

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. World Data Center A: 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 Ice Center. Compiled by F. Fetterer and C. Fowler. 2006, updated 2009. *National Ice Center Arctic Sea Ice Charts and Climatologies in Gridded Format, Version 1*. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. https://doi.org/10.7265/N5X34VDB.

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.

Schweiger, A.J., K.R. Wood and J. Zhang. 2019. Arctic sea ice volume variability over 1901-2010: A model-based reconstruction. J. Climate, 32, 4731-4752, https://doi.org/10.1175/JCLI-D-19-0008.s1.

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, J. E. 1978. A data set on Northern Hemisphere sea ice extent. *Glaciological Data*, Report GD-2, Arctic Sea Ice Part 1. World Data Center A: 49 - 51.

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., F. Fetterer**,** J. Scott Stewart, and W. L. Chapman. 2016. A database for depicting Arctic sea ice variations back to 1850. *Geographical Review* 107(1): 89-107. doi:10.1111/j.1931-0846.2016.12195.x

Walsh, J. E. and C. M. Johnson. 1979. Analysis of Arctic sea ice fluctuations 1953-77. *Journal of Physical Oceanography* 9(3): 580-591.

Walsh, J.E., J.S. Stewart, and F. Fetterer**.** 2019. Benchmark seasonal prediction skill estimates based on regional indices. *The Cryosphere* 13: 1073-1088. doi:10.5194/tc-13-1073-2019.

Walsh, J. E**.,** F. Fetterer, J. Scott Stewart, and W. L. Chapman. 2016. A database for depicting Arctic sea ice variations back to 1850. *Geographical Review* 107(1): 89-107. doi:10.1111/j.1931-0846.2016.12195.x.

Walsh, J. E., Stewart, J. S., and Fetterer, F. 2019. Benchmark seasonal prediction skill estimates based on regional indices, *The Cryosphere* 13: 1073-1088, https://doi.org/10.5194/tc-13-1073-2019.

16.1 Related References

Walsh, J. E., Stewart, J. S., & Fetterer, F. (2019). Benchmark seasonal prediction skill estimates based on regional indices. *The Cryosphere*, *13*(4), 1073-1088. https://doi.org/10.5194/tc-13-1073-2019.

Walsh, J.E., Fetterer, F., Scott Stewart, J. and Chapman, W.L. 2017. A database for depicting Arctic sea ice variations back to 1850. *Geogr Rev* 107: 89-107. https://doi.org/10.1111/j.1931-0846.2016.12195.x

17 Document Information

17.1 Author

F. Fetterer and A. Windnagel

17.2 Publication Date

July 2016

17.3 Document Revision Date

January 2020: A. Windnagel added notice about the incorrect Kelly grids and updated section 4.4 to note that the browse images show concentration at 15% or greater.

September 2019: A. Windnagel and F. Fetterer updated the documentation to reflect the new Version 2 data.

18 Appendix 1: 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. Note that Figures 14 through 18 only show data north of 50 degrees. Figure 20 shows the entire data grid area of the Northern Hemisphere north of 30 degrees.

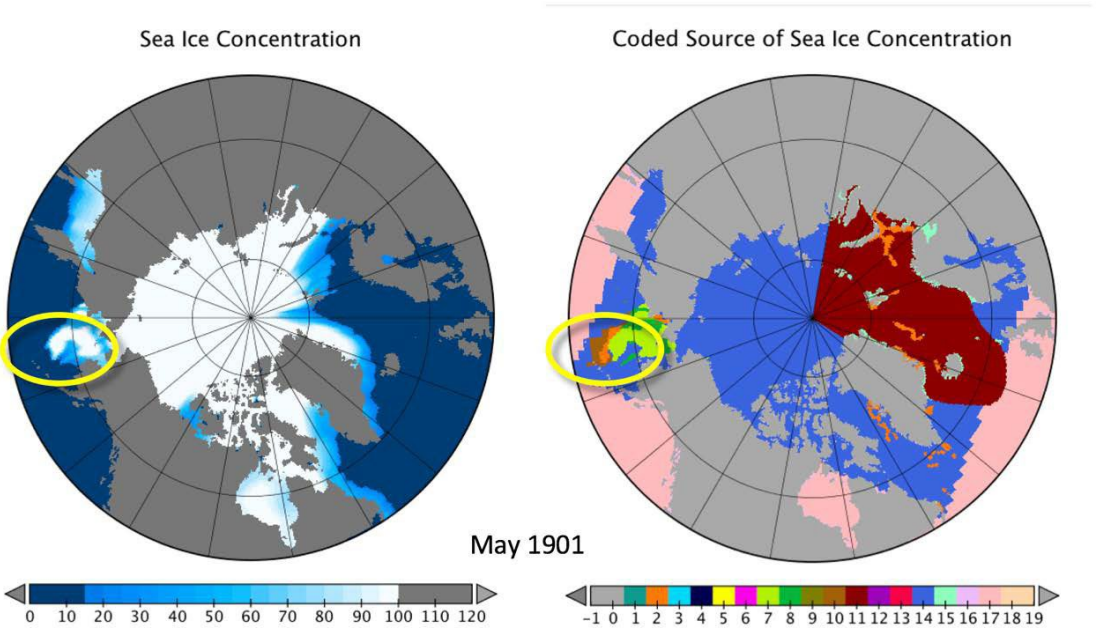


Figure 15. The first example of Source 10, the DMI yearbook narrative source, can be seen in the North Pacific in May 1901 (left). The DMI observations indicated "ice free" where Source 10 and Source 2 appear. Whaling log books (Source 7) mark where ice is present in this region, as reflected in the corresponding concentration field (right).

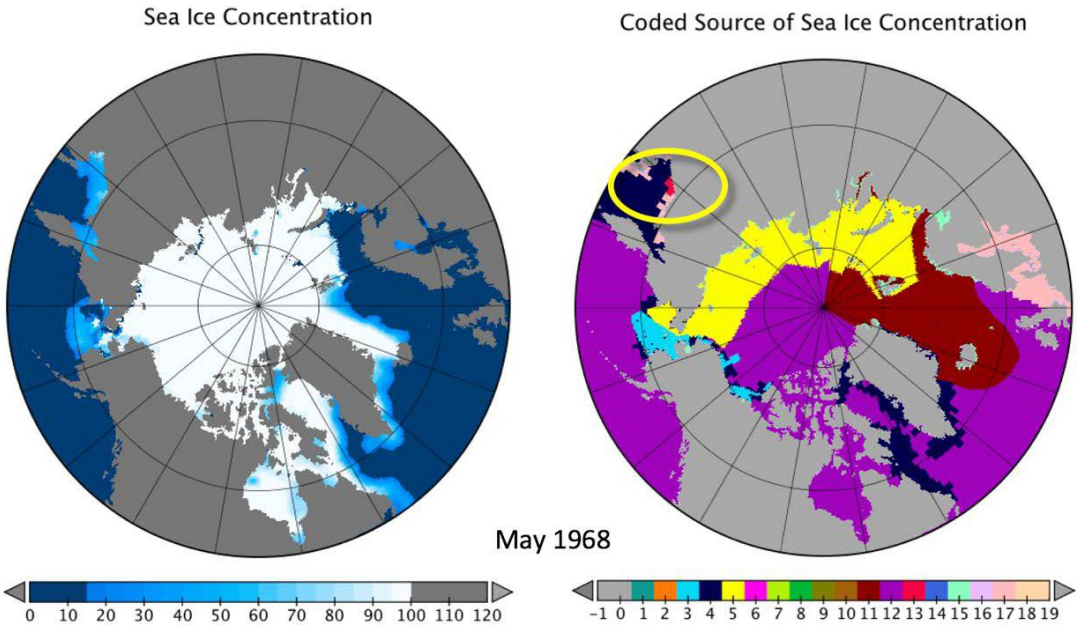


Figure 16. The first example of Source 13, data from the Japan Meteorological Agency (JMA)

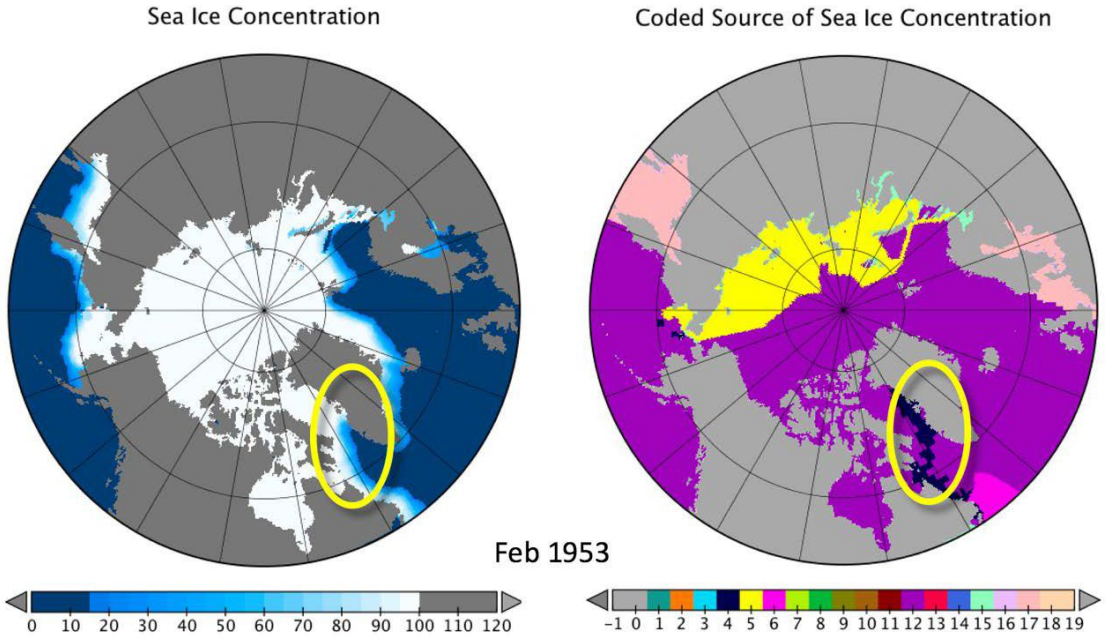


Figure 17. The first example of Source 4, Naval Oceanographic Office (NAVO) yearbook data, occurs in February 1953, where it aids in defining the ice edge in Baffin Bay.

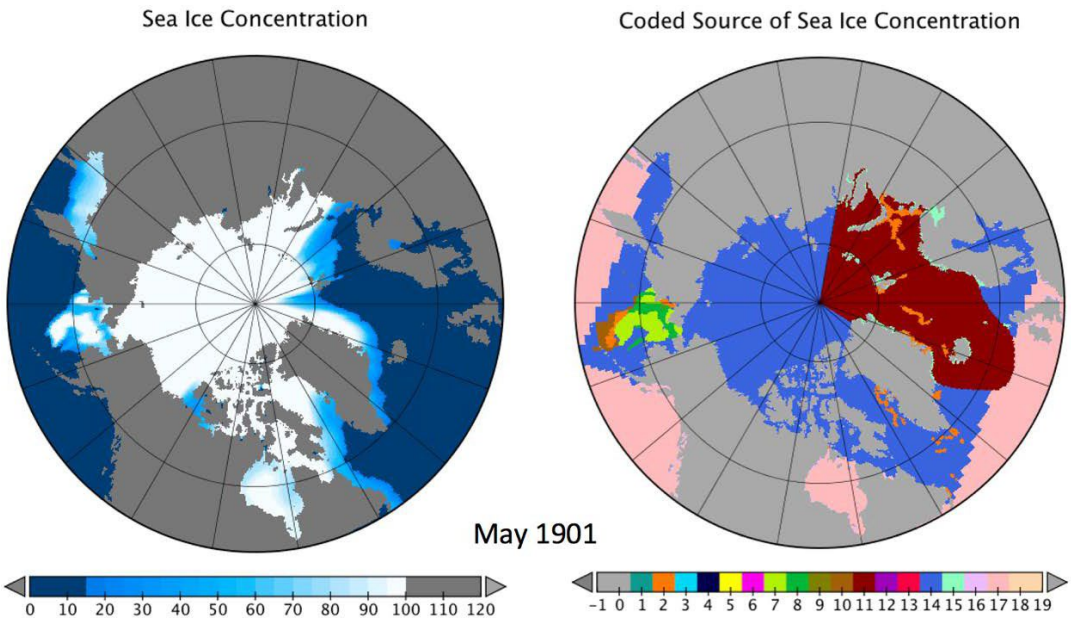


Figure 18. The first example of Source 14, the Kelly ice extent grids, occurs in May1901. This coverage is typical through 1952.

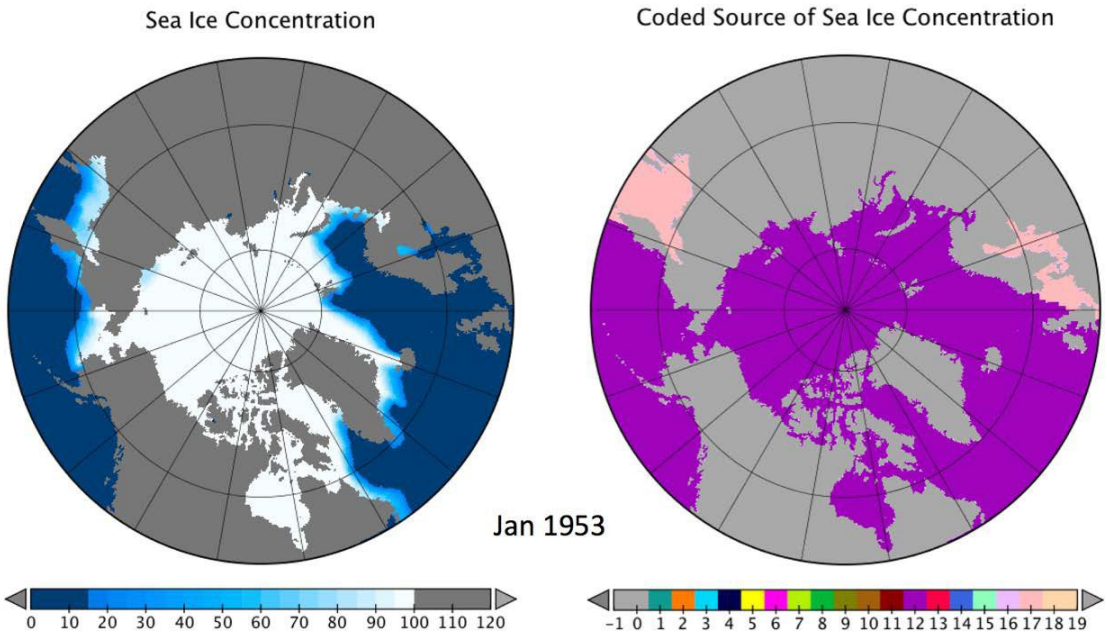


Figure 19. In January of 1953, the only source is Walsh and Johnson (Source 12). Analog filling of spatial gaps is used elsewhere.

19 Appendix 2: Known erroneous data

June 1962 (Figure 20) has a small patch of ice at concentrations of 6%-10% in the mid-Atlantic Ocean. Source 4, NAVO Yearbooks, is the source of this erroneous ice.

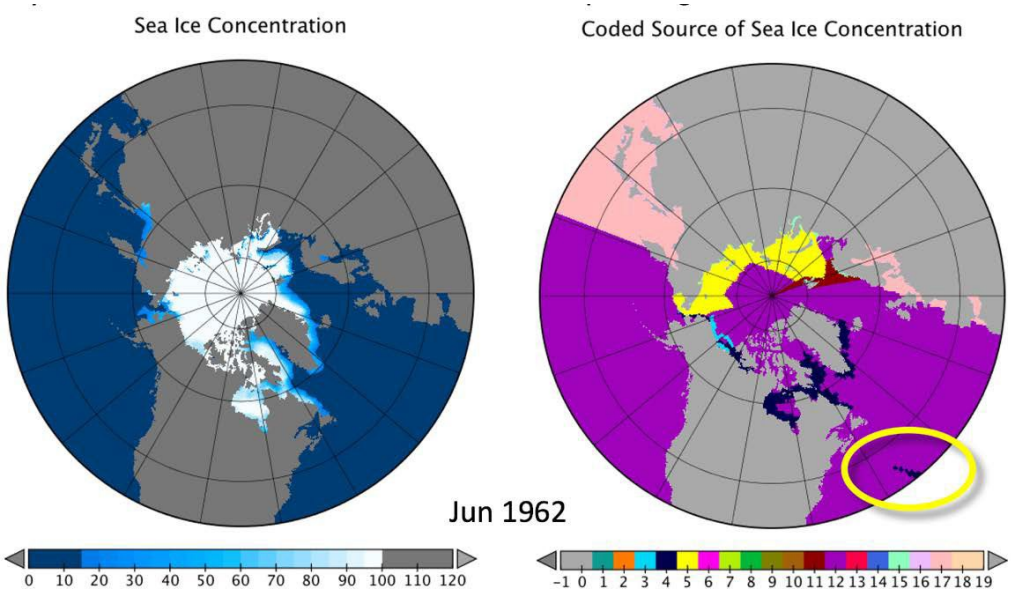


Figure 20. June 1962 has falsely indicated ice at low concentrations (<15%) in the mid-Atlantic.

The ice concentration field derived from the DMI chart for April 1901 (Source 2), obtained from *Arctic Sea Ice Concentration and Extent from Danish Meteorological Institute Sea Ice Charts, 1901-1956*, is erroneous (Figure 21). The ice edge in the Pacific was missed when the April 1901 ice chart was created in the source data set. (Figure 22).

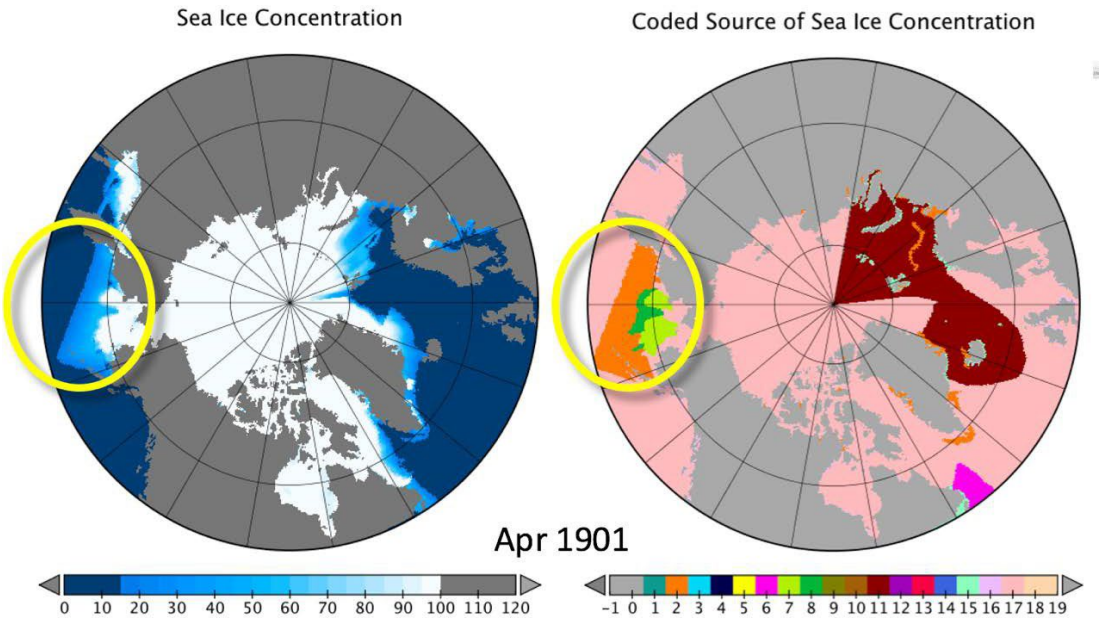


Figure 21. April 1901 erroneously shows ice in the north Pacific.

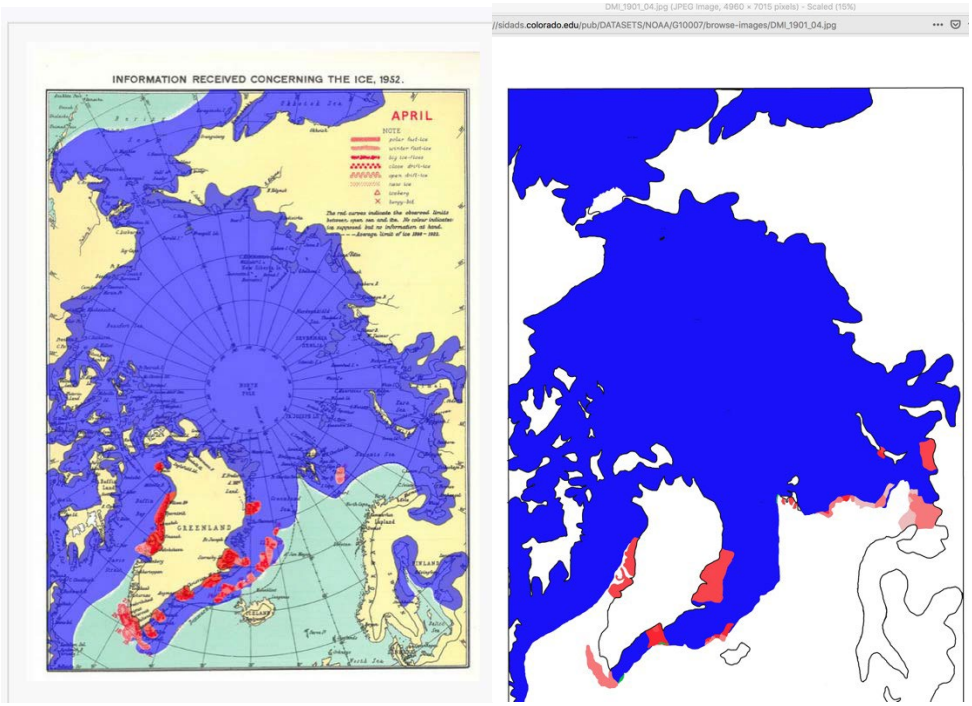


Figure 22. A figure from the User Guide for Underhill et al., 2014, (left) shows a DMI chart underlying areas interpreted has having sea ice of some designated concentration. These are marked in shades of red. The darker blue area traces the ice edge where there were no ship observations but the DMI map makers had assumed that there was ice. A browse image from the data set (right) shows how ice concentration and ice edge were coded for the map from April 1901 (right). Clearly, the ice edge in the Pacific sector was missed.

Figure 23 illustrates a case of a misaligned data source. This case was discovered after the processing for V2 was complete. It will be corrected in the next update. However, there may be other cases in which we have overlooked misaligned source data.

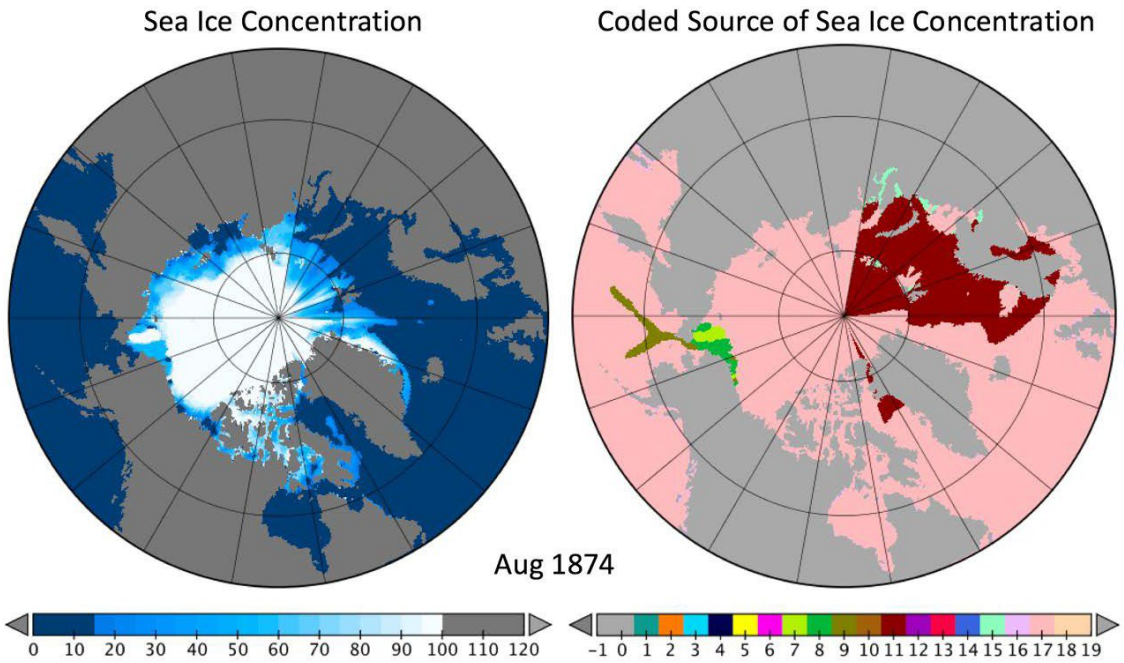


Figure 23. This example from August 1874 illustrates how certain data sets can be misregistered. The ACSYS data source grid, Source 11, (right) is not properly geolocated. Notice how the shape of Iceland appears as a hole in the ACSYS source coverage about 20 degrees to the east. The dislocated concentration values result in an oddly contoured concentration field (left).

20 Appendix 3: Differences in sea ice extent and area between Version 1 and Version 2, and notes on the extent and area time series.

The processing code, and the grids of source data concentration values, that were used to produce CW91 and SIBT1850 V1 are not available to the team that worked on V2. This limited our ability to diagnose sources of V1/V2 differences, and to thoroughly explore the reasons behind some of the characteristics of the extent and area time series.

20.1 V1/V2 differences

In general, the V2 extent and area values are slightly higher than those derived from V1 (Figure 24). To create V1.1, inconsistencies in the V1.0 land mask were addressed by applying an "if ever land, always land" rule to grid cells. This had the effect of slightly expanding the land mask, so that the area covered by ocean cells, and thus the area that could contain sea ice, was proportionally reduced. The improved V2 land mask leaves more ocean cells in which ice may occur. In total, about 450,000 sq km of coastal ocean are added. V2 ice extent and area differ from extent and area derived using V1, simply because the V1 and V2 land areas are not the same. This difference is greatest in the winter months, when ice reaches a greater expanse of coastline.

The largest V1-V2 differences in the historical period are for winter months when temporal fill is used (indicated by yellow shading in the plots).

In the satellite era, another factor comes into play. The CDR V3 *goddard_merged_seaice_conc* source used for SIBT1850V2 has higher values than the CDR V2 used for SIBT1850V1. This accounts for between about 30% and 70% of the greater SIBT1850V2 extent values for March, for example.

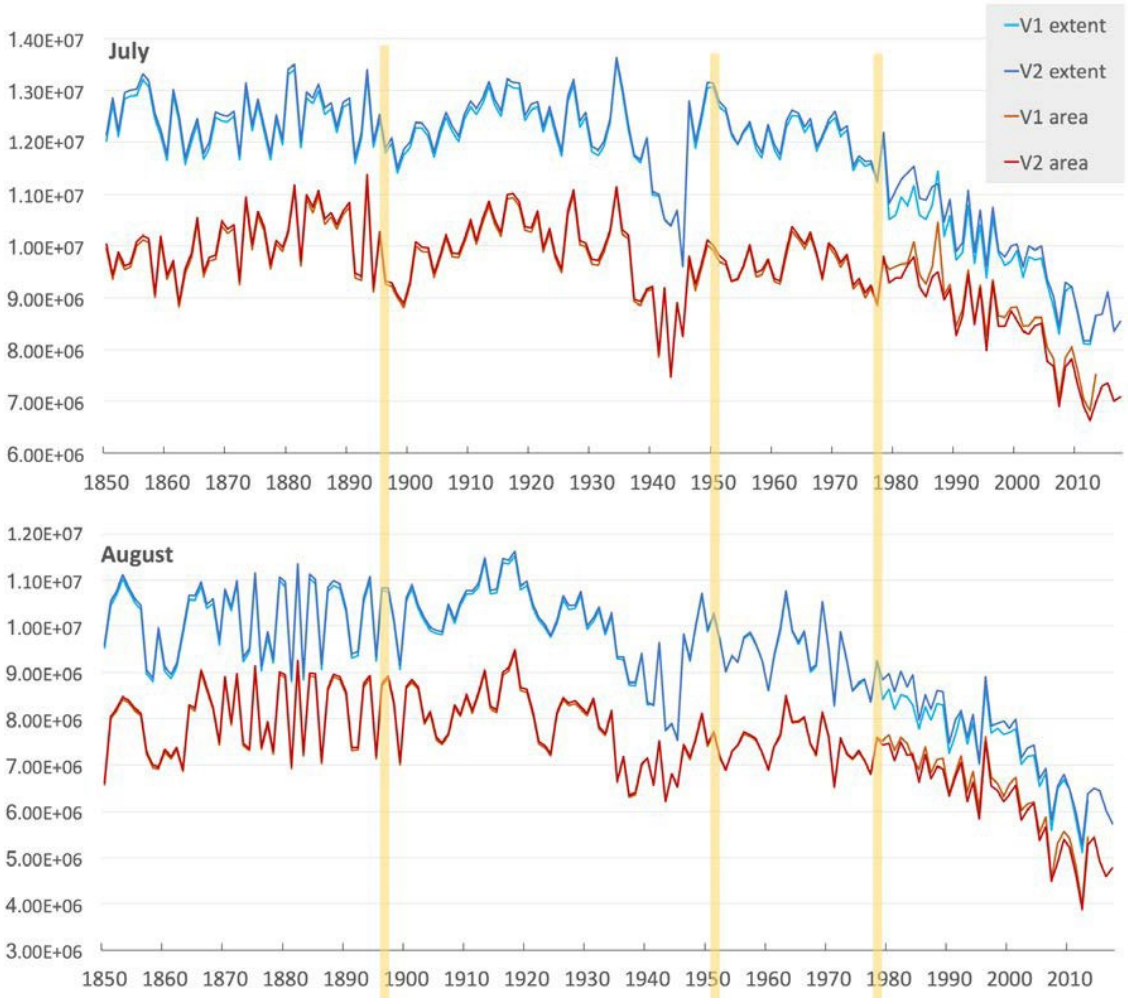
In summary, most of the higher V2 extent and area values are the result of the changed land mask, with the change in input CDR source data version being an additional factor in the satellite era. There is a seasonal effect in the change attributable to the improved land mask.

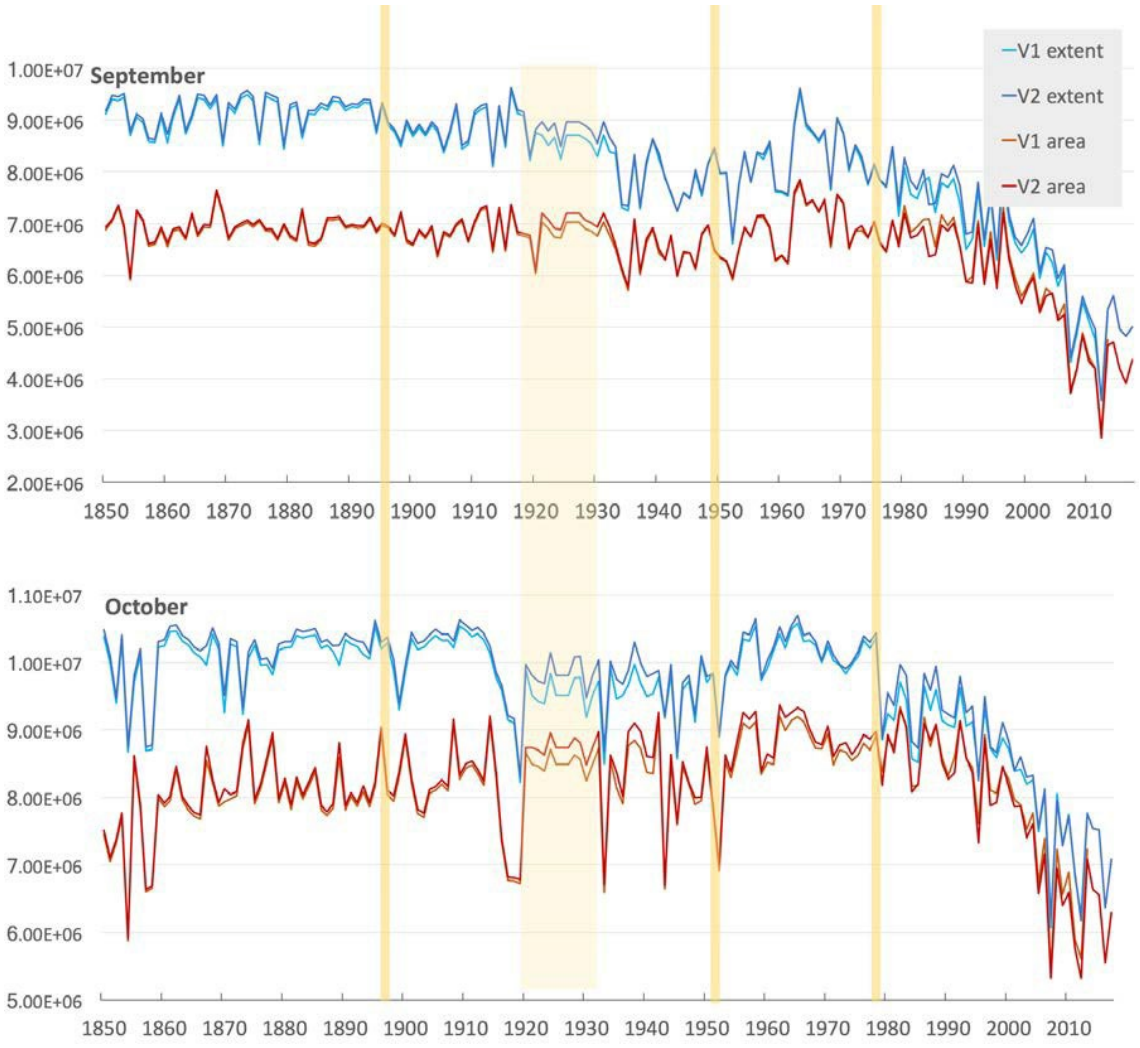
20.2 General characteristics of the time series

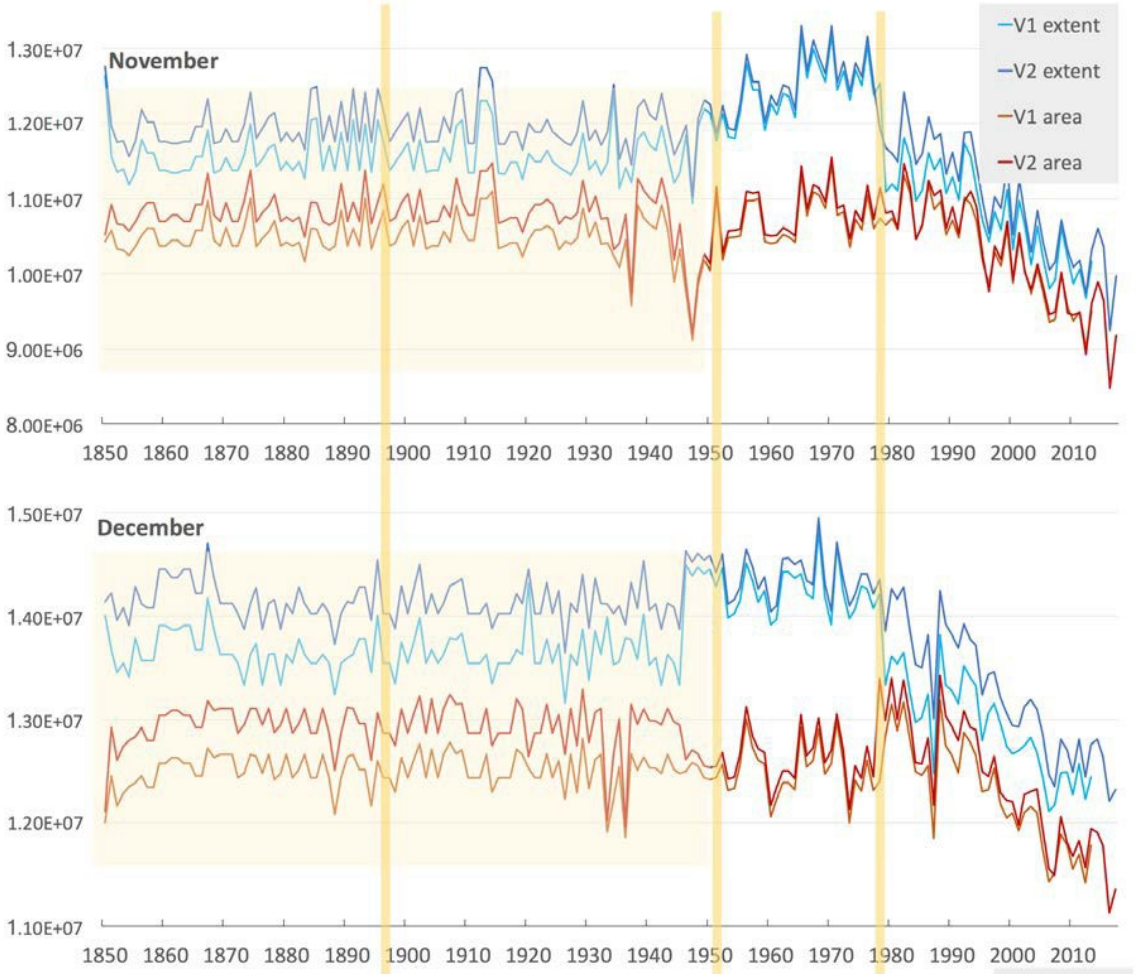
There are heterogeneities in the ratio of ice area/extent ratio as one moves from the pre-satellite period to the passive microwave period. More specifically, the area is a larger fraction of extent in the satellite data. While the reasons for it are not apparent, the extents are considered much more reliable than the concentration-dependent areas because much of the information from earlier decades was from ships or coastal observations.

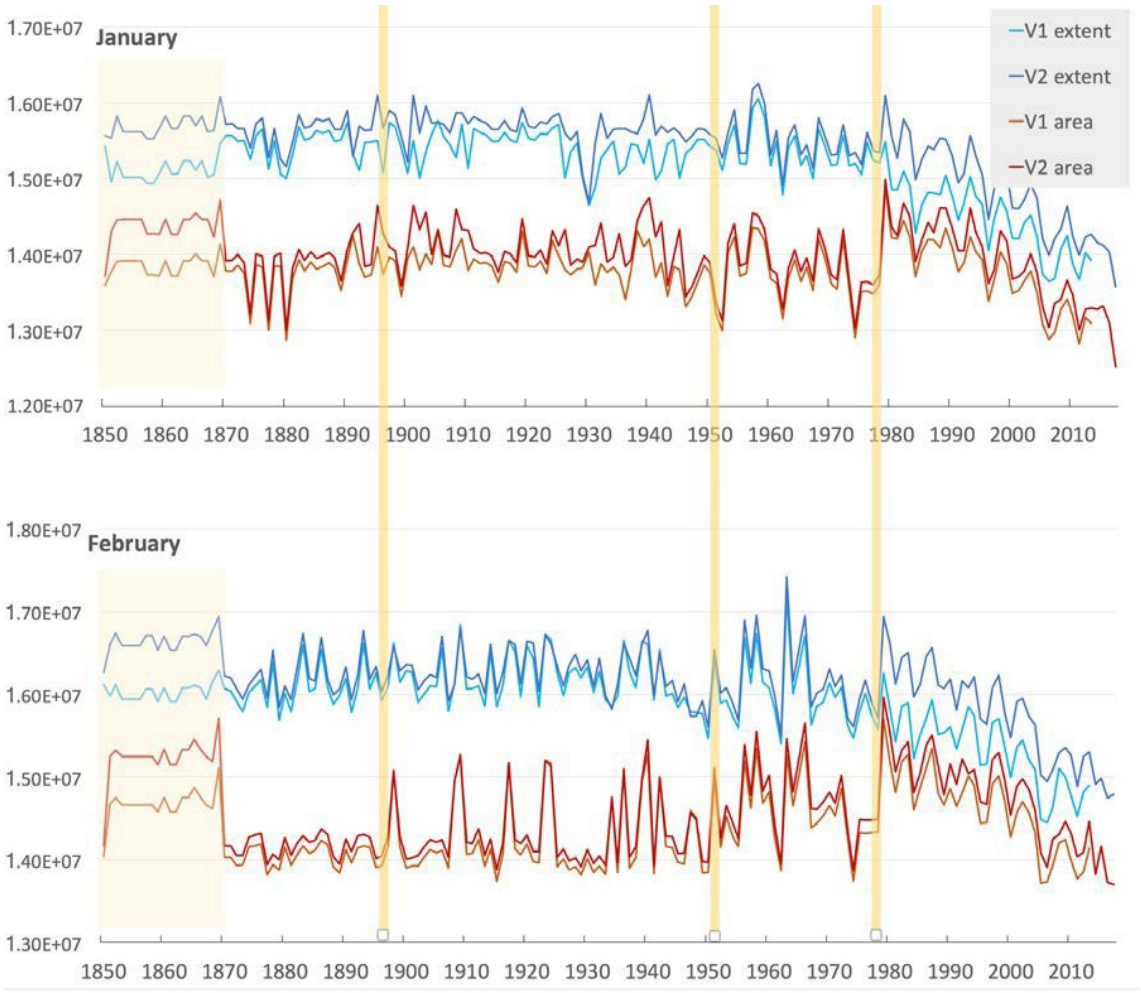
For some months, like October, February, March, and April, the average area over most of the historical record appears smaller than that of the later part of the historical period (1953-1978) or the satellite era. Again, the reasons for this are not apparent, but we suspect the ice concentration fields are biased low as a result of the way the sparse source data from this time has been digitized and used.

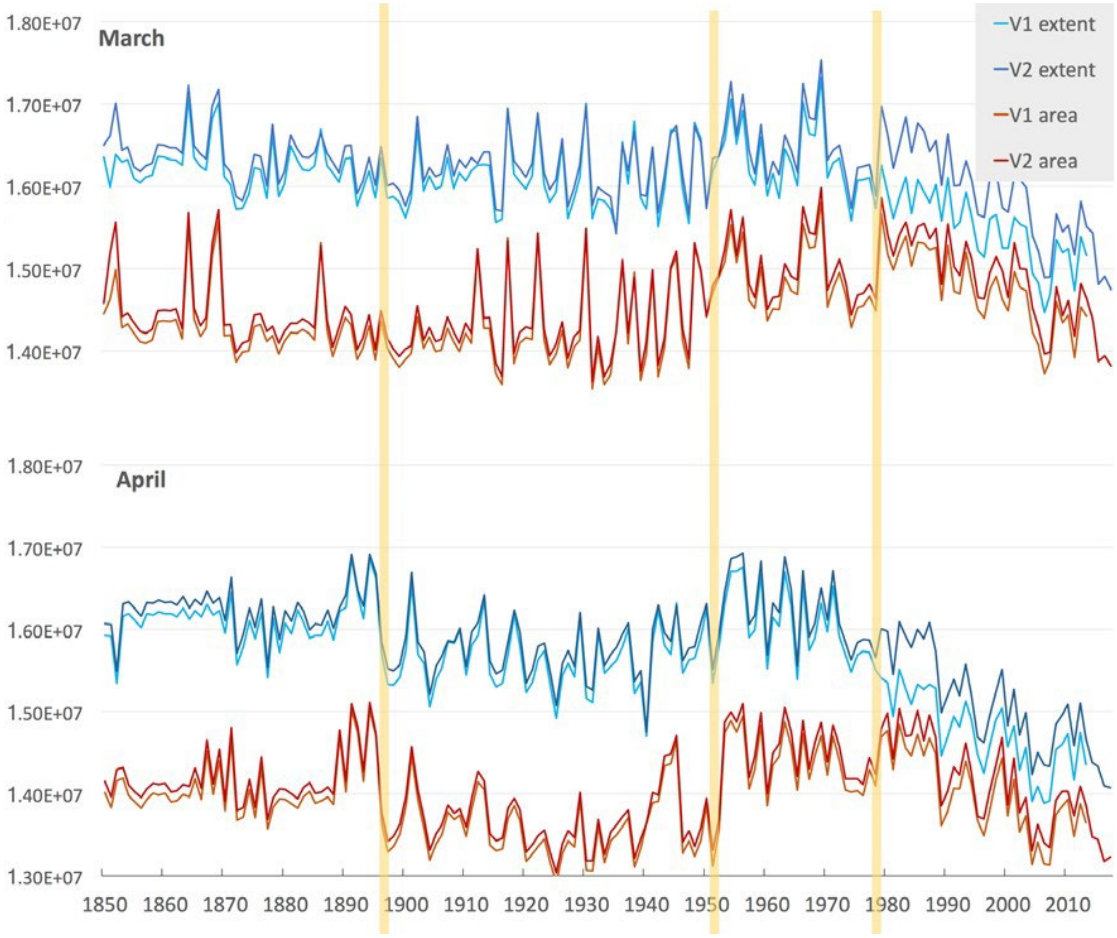
Coverage in the summer months of the 1900s through the 1930s is relatively good, with information from the Danish Meteorological Institute maps (Source 2) and the Kelly grids (Source 14) providing the majority of data. There are no data from either of these sources during the war years of 1941 through 1946. (The montage files are helpful if one wishes to quickly assess data source coverage.) This lack of data necessitates that analog spatial fill (Source 17) be used. Shortcomings in the analog fill process are the reason behind the unrealistic dip in extent and area especially noticeable in Julys of years around 1940.




September
1.00E+07
9.00E+06
8.00E+06
7.00E+06
6.00E+06
5.00E+06
4.00E+06
3.00E+06
2.00E+06
1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010
V1 extent
V2 extent
V1 area
V2 area
October
1.10E+07
1.00E+07
9.00E+06
8.00E+06
7.00E+06
6.00E+06
5.00E+06
1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010


November
December
V1 extent
V2 extent
V1 area
V2 area
1.30E+07
1.20E+07
1.10E+07
1.00E+07
9.00E+06
8.00E+06
1.50E+07
1.40E+07
1.30E+07
1.20E+07
1.10E+07
1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010


January
February
V1 extent
V2 extent
V1 area
V2 area
1.70E+07
1.60E+07
1.50E+07
1.40E+07
1.30E+07
1.20E+07
1.80E+07
1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010


March
April
V1 extent
V2 extent
V1 area
V2 area
1.80E+07
1.70E+07
1.60E+07
1.50E+07
1.40E+07
1.30E+07
1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010

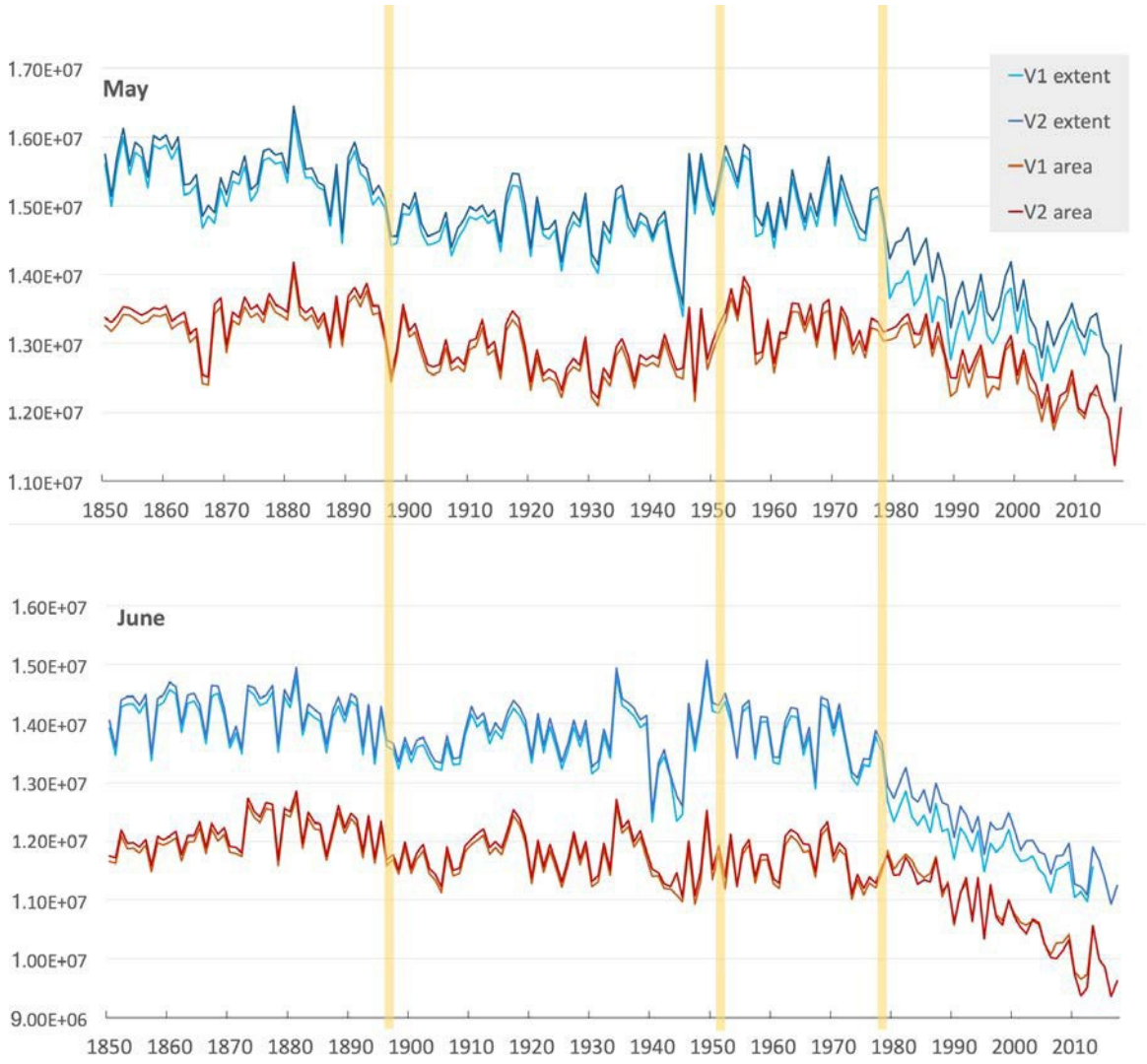


Figure 24. July (top) through June (bottom) extent (blue) and area (red) for the entire time series. Both V1 and V2 values are plotted; V2 values are the darker line and are generally larger than V1 values. The leftmost vertical yellow line marks when data from the Danish Meteorological Institute begins in around 1900. The middle vertical yellow line marks approximately when the Arctic-wide coverage of Walsh and Johnson (Source 12) begins in 1953. The rightmost yellow line marks when satellite-only coverage begins in 1979. The yellow shading that appears in parts of the September through February series indicates when temporal filling was required.