# Seasonal Antarctic Sea Ice Extent Reconstructions, 1905-2020, Version 1

## USER GUIDE

**How to Cite These Data**

As a condition of using these data, you must include a citation:

Fogt, R. L., M. N. Raphael, and M. S. Handcock. 2023. *Seasonal Antarctic Sea Ice Extent Reconstructions, 1905-2020, Version 1*. [Indicate subset used]. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. https://doi.org/10.7265/55x7-we68. [Date Accessed].

FOR QUESTIONS ABOUT THESE DATA, CONTACT NSIDC@NSIDC.ORG

FOR CURRENT INFORMATION, VISIT https://nsidc.org/data/G10039

National Snow and Ice Data Center

# TABLE OF CONTENTS

# 1 DATA DESCRIPTION

## 1.1 Summary

These data are reconstructed seasonal Antarctic sea ice extent (SIE) for 1905 through 2020. Reconstruction ensembles and metadata (in the form of parameters that describe each member of the ensemble) are given by season for five geographic sectors as well as for the entire Antarctic. The reconstruction method uses temperature and pressure observations that begin in 1905 as well as climate indices. Reconstructions have been calibrated to satellite-period observations.

Reliable satellite sea ice observations of Antarctic sea ice extent begin in 1978, with limited satellite data available before then. Fogt et al. (2022a) derived these data in order to have a longer series that would make it possible to assess the significance of recent (~2015) large changes in extent in the context of longer-term variability. At time of publication, the resulting data are unique in being a 116-year-long seasonally resolved reconstruction of Antarctic-wide sea ice extent.

The reconstructions are highly skillful and robust, showing little sensitivity to the particular predictor data used. They are, as the data set authors write in Fogt et al. (2022a), “appropriate to use in further research on evaluating climate models and interpreting the causes of the regime shifts depicted in these reconstructions.“

## 1.2 Parameters

Antarctic sea ice extent (SIE) is the geophysical parameter provided in this data set. It is provided along with uncertainty estimates and metadata that describe aspects of the reconstruction process that produced the SIE data.

## 1.3 Spatial Information

Reconstructed SIE ensembles are given for the entire Southern Ocean, as well as for five contiguous sectors (listed below). The sectors are defined based on their spatial decorrelation scale as discussed in Raphael and Hobbs (2014). The longitude bounds for the sectors are as follows:

- Amundsen-Bellingshausen Seas (250° E to 290° E)
- Weddell Sea (250° E to 346° E)
- King Hakon VII (346° E to 71° E)
- East Antarctica (71° E to 162° E)
- Ross-Amundsen Sea (162° E to 250° E).

Note that gridded mask files for these sectors are available in Arctic and Antarctic Regional Masks for Sea Ice and Related Data Products, Version 1 data set as the Raphael and Hobbs ("RH") masks.

Figure 1 shows the locations and names of stations with pressure and temperature records that were used, as well as the location of the sector boundaries.


Figure 1. Southern hemisphere long-term pressure and temperature observation locations with station names shown with sea ice sectors. ABS stands for Amundsen-Bellingshausen. After Extended Data Figure 1 in Fogt et al. (2022a).

# 1.4 Temporal Information

## 1.4.1 Coverage

Reconstructed sea ice extent numbers cover from 1905 through 2020. Extent numbers directly from observations are provided from 1979 to 2020 when satellite data are available.

## 1.4.2 Resolution

Data are seasonal: December, January, February (DJF); March, April, May (MAM); June, July, August (JJA); and September, October, November (SON).

# 1.5 File Information

## 1.5.1 Format

The data files are available in comma-separated values format (.csv) and NetCDF-3 format (.nc) and for the summed best fit SIE there is also an Excel workbook (.xlsx) available. We have also provided the NCAR Command Language (.ncl) scripts used to generate some of the figures in Fogt et al. (2022a), which can be found in the `FigureCode_from_Fogt22` directory.

## 1.5.2 File Contents and Directory Structure

Data and metadata are organized in six directories. A seventh directory has the NCAR Command Language code. The directory names are as follows:

- Observations
- Reconstruction_Ensembles
- Reconstruction_Metadata
- Reconstruction_BestFit_and_Mean
- Antarctic-wide_BestFit_and_Mean_of_Sector_Sums
- Antarctic-wide_Sum_of_Sector_BestFits
- FigureCode_from_Fogt22

The directories are used as subsections below with a general description of their contents. In all files, sea ice extent is given in $10^6$ $km^2$.

### 1.5.2.1 Observations Directory

This directory contains files of seasonal SIE derived from passive microwave data. The files are provided in both CSV text and NetCDF-3 format. In all the files, the *missing* data value of -999.900 is listed for the years 1905 through 1978 because no usable satellite data exists for those years. For 1979 through 2020, the seasonal SIE is derived from the daily values in the NOAA/NSIDC Sea Ice Concentration Climate Data record (Meier et al., 2021).

**CSV format**

There are 24 SIE observation CSV files: for each of the five sectors and for the whole region (Antarctic-wide) there is a file for each of the four seasons (see Table 8 for a list of the sectors and seasons). The files contain two columns: `Year` and `Observed_Sea_Ice_Extent_(x10^6 km^2)`, which are the 4-digit year of the observation and the sea ice extent in millions of sq km, respectively.

**NetCDF format**

There are four NetCDF files, one for each season with seven variables within each for each of the five sectors and the whole region (Antarctic_wide) plus a time variable. The variables are described in Table 1.

Table 1. Variables in the NetCDF files in the Observations directory.

| Variable | Description |
|---|---|
| Amundsen_Bellingshausen_obs | Amundsen Bellingshausen observed sea ice extent in $10^6$ $km^2$ |
| East_Antarctica_obs | East Antarctica observed sea ice extent in $10^6$ $km^2$ |
| King_Hakon_obs | King Hakon observed sea ice extent in $10^6$ $km^2$ |
| Ross_Amundsen_obs | Ross Amundsen observed sea ice extent in $10^6$ $km^2$ |
| Weddell_obs | Weddell observed sea ice extent in $10^6$ $km^2$ |
| Antarctic_wide | Antarctic wide observed sea ice extent in $10^6$ $km^2$ |
| time | Years for which reconstructions were run |

### 1.5.2.2 Reconstruction_Ensembles Directory

This directory contains files of seasonal SIE from the reconstruction ensembles. Each file contains time series of SIE from different ensemble members. The number of ensemble members in each file varies and the files are available in both CSV and NetCDF format.

**CSV format**

There are 24 reconstruction ensemble CSV files: for each of the five sectors and for the whole region (Antarctic-wide) there is a file for each of the four seasons (see Table 8 for a list of the sectors and seasons). The first row of each file is a header denoting the Year column and the sector and season the data correspond to. The first column of each file is Year which contains the 4-digit year of the reconstruction. Each subsequent column has SIE given by a member of the ensemble of reconstructions for that season and sector, or in the case of the Antarctic-wide files, for the Antarctic overall. The number of columns in each file varies – different sectors and different seasons for the same sector have differing numbers of SIE reconstructions. The number of columns is determined by the number of ensemble members that are selected based on the strength of the predictor networks' correlation with observed SIE (see the Processing section for further details). Figure 2 shows a sample of the CSV files.

```
Year,Amundsen_Bellingshausen_DJF,,,,
1905,0.454,0.538,0.576,0.638,0.549
1906,0.530,0.695,0.570,0.584,0.585
1907,0.547,0.403,0.470,0.528,0.522
```

Figure 2. The first few lines for the DJF season for the Amundsen-Bellingshausen sector (Amundsen_Bellingshausen_DJF_recon_ensembles.csv

**NetCDF format**

There are 24 reconstruction ensemble NetCDF files: for each of the five sectors and for the whole region (Antarctic-wide) there is a file for each of the four seasons (see Table 8 for a list of the sectors and seasons). In each file there are three variables as described in Table 2 below.

Table 2. Variables in the NetCDF files in the Reconstruction_Ensembles directory.

| Variable | Description |
|---|---|
| ens_member | Number of ensemble members from the reconstruction |
| ens_recons | Sea ice extent for each of the years and ensemble members in $10^6$ $km^2$ |
| time | Years for which reconstructions were run |

### 1.5.2.3 Reconstruction_Metadata Directory

This directory contains files with the metadata for each ensemble member in the corresponding files in the `Reconstruction_Ensembles` directory (see section 1.5.2.2 above). Files are only in CSV format.

There are 24 reconstruction metadata CSV files: for each of the five sectors and the whole region (Antarctic-wide) there is a file for each of the four seasons (see Table 8 for a list of the sectors and seasons). Each row in a file holds metadata describing the data layer used, the skill, and other factors of the reconstruction responsible for the SIE found in the corresponding column in the Reconstruction Ensemble file for the corresponding sector and season. There are as many rows in these files as there are columns in the corresponding Reconstruction Ensemble files. Table 3 describes the metadata file columns and Figure 3 shows a sample of the files.

```
Master Stats For All Data Layers and Networks.....,,,,,,,,,,
Station_Name,Data_Layer,%Network,#factors,cal r,val r,val r (dt),RE,CE,Best?
Bellingshausen-Amundsen_Sea,Obs,All,12,0.8514,0.7459,0.7462,0.7245,0.5448,no
Bellingshausen-Amundsen_Sea,Obs,10pct,9,0.8648,0.7681,0.7698,0.7461,0.5784,no
Bellingshausen-Amundsen_Sea,IPO,10pct,13,0.8909,0.7779,0.7869,0.7923,0.5854,no
```

Figure 3. The first few lines of the `Amundsen_Bellingshausen_JJA_recon_master_stats.csv` file.

Table 3. Contents of the CSV files in the Reconstruction_Metadata directory.

| Column | Description |
|---|---|
| Station_Name (Note that this designates *sector*, not *station*. See Figure 1) | The sector for which reconstruction statistics are provided. When the ensemble member is the reconstruction for the entire Antarctic, this column is filled with `totalsie`. |
| Data_Layer | The predictor data used for this ensemble member reconstruction. See Table 4 for an explanation of the names used for each data layer. |
| %Network | The threshold for probability (P-value) of significant correlation of the predictor data with observed SIE. Data layers meeting the threshold are admitted to the indicated network. |
| #factors | Number of Principal Components (PCs) that go into the reconstruction represented by this row. |
| cal r | Calibration correlation. This is the correlation of this ensemble member reconstruction with observed SIE. |
| val r | Validation correlation. This is the correlation with observed SIE of an entirely predicted and independent time series, made with this data layer. See Fogt et al. (2022a). |
| val r (dt) | The validation correlation with all linear trends removed in both the validation reconstruction and SIE time series. |
| RE | Reduction of Error. RE is a measure of the goodness of fit of the reconstruction, compared to the climatological mean. RE ranges from negative infinity to +1; +1 is a perfect reconstruction, RE>0 indicates the reconstruction is performing better than the climatological mean. |
| CE | Coefficient of efficiency. The CE is like the RE, but based on the independent validation reconstruction. CE ranges from negative infinity to +1, with the same indications for skill as the RE. Because it is based on validation reconstruction, it is always less than or equal to the RE.<br>CE is computed using val r. CE is used in a process to select a subset of possible PCs and thereby limit overfitting. See Fogt et al. (2022a). |
| Best? | A “yes” in this column indicates the row that holds the metadata describing the most skillful reconstruction ensemble member, usually that with the highest CE. |

Table 4. Data_Layer Names

| Abbreviation | Full name or description for observations or climate index |
|---|---|
| Obs | Temperature and pressure observations, primarily from UCAR ds 570.0 |
| IPO | Interdecadal Pacific Oscillation |
| AMO | Atlantic Multidecadal Oscillation |
| PDO | Pacific Decadal Oscillation |
| SOI | Southern Oscillation Index |
| SAM | Southern Annular Mode |
| Nino1.2 | Niño SST index calculated using the NOAA Extended Reconstructed Sea Surface Temperature dataset, version 5 (ERSSTv5) averaged over 0°–10° S, 270° E–280° E |
| Nino3.4 | As above, averaged over 5° N–5° S, 210° E–270° E |
| Nino3 | As above, averaged over 5° N–5° S, 190° E–240° E |
| Nino4 | As above, averaged over 5° N–5° S, 160° E–210° E |

More information about the indices in Table 4, can be found in the NCAR Climate Data Guide. The Climate Data Guide also has explanations of the Niño SST indices.

### 1.5.2.4 Reconstruction_BestFit_and_Mean Directory

This directory has files with time series of SIE from the following; observations, the best fit ensemble member, the mean of all ensemble members, with 95% lower and upper confidence intervals. (The austral winter (JJA) sector data in these files are plotted in Figure 4 of Fogt et al. (2022a).). There are 24 of each CSV and NetCDF files: for each of the five sectors and for the whole region (Antarctic-wide) there is a file for each of the four seasons (see Table 8 for a list of the sectors and seasons). Table 5 describes the file contents and Figure 4 of this document shows a sample of the CSV files. For more detailed information, see the Methods section of Fogt et al. (2022a).

```
Year,Observed_Sea_Ice_Extent,Best_Fit_Recon,Ensemble_Mean_Recon,95%_Lower_CI,95%_Upper_CI
NOTE:Sea Ice Extent is in 10^6 km^2,,,,,
1905,-999.900,0.780,0.691,0.475,0.922
1906,-999.900,0.683,0.629,0.473,0.825
1907,-999.900,0.353,0.384,0.153,0.615
```

Figure 4. The first few lines of the CSV file for the MAM season for the Amundsen-Bellingshausen sector (`Amundsen_Bellingshausen_MAM_best_fit_recons.csv`).

Table 5. Contents of the CSV and NetCDF files in the Reconstruction_BestFitandMean directory.

| CSV Column Name | NetCDF Variable | Description |
|---|---|---|
| Year | time | A row for each of 116 years from 1905 through 2020 |
| Observed_Sea_Ice_Extent | obs_sea_ice_extent | Sea ice extent derived from passive microwave sea ice concentration data in $10^6$ $km^2$. |
| Best_Fit_Recon | best_fit_recon | Reconstructed sea ice extent given by the highest-performing ensemble member. |
| Ensemble_Mean_Recon | ens_mean_recon | The average of all ensemble member reconstructions. (These individual reconstructions can be found in the Reconstruction_Ensembles directory). |
| 95%_Lower_CI | lower_CI | Lower limit of 95% confidence interval for reconstruction uncertainty, given by *the greater of* 1.96 times the standard deviation of the ensemble members or 1.96 times the standard deviation of the residuals between the best fit reconstruction ensemble member and the observed data (this value is a constant). |
| 95%_Upper_CI | upper_CI | As above, for the upper limit of the 95% confidence interval. |

### 1.5.2.5 Antarctic-wide_BestFit_and_Mean_of_Sector_Sums Directory

These files contain the SIE reconstruction that is the best fit of all possible sums of sector reconstructions, along with the SIE time series that is the mean of all possible sums, the standard deviation of these, and the total number of ensembles used. There are four of each CSV and NetCDF files – one for each of the four seasons. The total number of ensembles used in each season are given at the top of the CSV files or listed in the `long_name` attribute for the `best_fit_total_recon` variable in the NetCDF file. The number ranges from a minimum of 234,900 for season DJF to 1,159,200 for season JJA. The file contents are described in more detail in Table 6 and a sample of the CSV file is shown in Figure 5.

```
#ens 234900,,,
Year,Best-Fit_Total_Recon,Total_Ensmean,Total_EnsStdDev
1905,6.6240,6.3823,0.3318
1906,6.3970,6.1642,0.3740
1907,6.3650,6.3836,0.2706
```

Figure 5. The first few lines of the CSV file for the DJF season (`DJF_total_best_fit_and_mean.csv`).

Table 6. Contents of the CSV and NetCDF files in the Antarctic-wide_BestFit_and_Mean_of_Sector_Sums directory.

| CSV Column Name | NetCDF Variable | Description |
|---|---|---|
| Year | time | A row for each of 116 years from 1905 - 2020 |
| Best-Fit_Total_Recon | best_fit_total_recon | The best fit of all possible sums of sector reconstructions in $10^6$ $km^2$ |
| Total_Ensmean | total_ens_mean | Mean of the best fit of all possible sums of sector reconstructions in $10^6$ $km^2$ |
| Total_EnsStdDev | total_ens_stddev | Standard deviation of the best fit of all possible sums of sector reconstructions |

### 1.5.2.6 Antarctic-wide_Sum_of_Sector_BestFits Directory

There are nine files in this directory: four CSV files (one for each of the four seasons), four NetCDF files (one for each of the four seasons), and one Excel (.xlsx) file that contains four worksheets (one for each season) in that single file. Each CSV file contains columns and each NetCDF file contains variables with the best fit ensemble member for each of the five sectors along with SIE given by the sum of these across all sectors, these are described in more detail in Table 7 and a sample of a CSV file is shown in Figure 6.

**Note**: For most applications that require Antarctic-wide SIE, the authors recommend using the files in this `Antarctic-wide_Sum_of_Sector_BestFits` directory. That is, use either the values in the `Antarctic-wide_BF_Sum` column from the CSV files or the values in the `Antarctic_wide_sum_best_fits` variable from the NetCDF files (Table 7). For most applications, these are preferable to the values in the `Best-Fit_Total_Recon` variable from the files in the `Reconstruction_BestFit_and_Mean` directory. See Fogt et al. (2022a) for more information about how the summed sector version compares with other representations of Antarctic-wide sea ice extent in this data set.

```
Year,Amundsen_Bellingshausen_BF_recon,East_Antarctica_BF_recon,King_Hakon_BF_recon,
Ross_Amundsen_BF_recon,Weddell_BF_recon,Antarctic-wide_BF_Sum
1905,0.538,0.711,1.261,1.431,2.351,6.292
1906,0.695,0.871,0.865,1.910,1.777,6.118
1907,0.403,0.711,0.988,1.951,1.593,5.646
```

Figure 6. The first few lines of the file for the SON season (`Antarctic_wide_SON_summed_sector_best_fit.csv`)

Table 7. Contents of the CSV and NetCDF files in the Antarctic-wide_Sum_of_Sector_BestFits directory.

| CSV Column Name | NetCDF Variable | Description |
|---|---|---|
| Year | time | Years for which reconstructions were run |
| Amundsen_Bellingshausen_BF_recon | Amundsen_Bellingshausen_best_fit_recon | Best fit sea ice extent from reconstructions for Amundsen Bellingshausen sector in $10^6$ km$^2$ |
| East_Antarctica_BF_recon | East_Antarctic_best_fit_recon | Best fit sea ice extent from reconstructions for East Antarctica sector in $10^6$ km$^2$ |
| King_Hakon_BF_recon | King_Hakon_best_fit_recon | Best fit sea ice extent from reconstructions for King Hakon sector in $10^6$ km$^2$ |
| Ross_Amundsen_BF_recon | Ross_Amundsen_best_fit_recon | Best fit sea ice extent from reconstructions for Ross Amundsen sector in $10^6$ km$^2$ |
| Weddell_BF_recon | Weddell_best_fit_recon | Best fit sea ice extent from reconstructions for Weddell sector in $10^6$ km$^2$ |
| Antarctic-wide_BF_Sum | Antarctic_wide_sum_best_fits | Sum of best fit sea ice extent from all five sectors in $10^6$ km$^2$ |
| N/A | Antarctic_wide_obs | Antarctic-wide observed sea ice extent from the NOAA/NSIDC Sea Ice Concentration CDR in $10^6$ km$^2$ (Meier et al., 2021) |

#### 1.5.2.7 FigureCode_from_Fogt22 Directory

This directory holds NCAR command language code that was used by the authors to create the figures in Fogt et al. (2022a). NOAA@NSIDC has not tested this code. We include it here because some users of these data may find it useful.

### 1.5.3 Naming Convention

The naming convention for the files in the following four directories is described in Table 8:

`Observations`

`Reconstruction_Ensembles`

`Reconstruction_Metadata`

`Reconstruction_BestFit_and_Mean`

**Naming Convention:**

`[sector]_[season]_[data].[ext]`

**Example:**

`Amundsen_Bellingshausen_DJF_recon_ensembles.txt`

Table 8. Naming Convention for the files in the following directories: Observations, Reconstruction_Ensembles, Reconstruction_Metadata, and Reconstruction_BestFit_and_Mean.

| Variable Name | Description |
|---|---|
| sector | One of the following:<br>Amundsen_Bellinghausen<br>East_Antarctica<br>King_Hakon<br>Ross_Amundsen<br>Weddell<br>Antarctic_wide (all 5 sectors, entire Antarctic)<br>Antarctic_all_sectors (only applies to the NetCDF observation files) |
| season | One of the four seasons:<br>DJF (December, January, February)<br>MAM (March, April, May)<br>JJA (June, July, August)<br>SON (September, October, November) |

| Variable Name | Description |
|---|---|
| data | Depending on the directory it can be one of the following:<br>observations<br>recon_ensembles<br>recon_master_stats<br>best_fit_recons |
| .ext | File extension:<br>.csv – identifies this file as CSV text file (applies to all data)<br>.nc – identifies this as a NetCDF file (applies to observations, recon_ensembles and best_fit_recons) |

The naming convention for the files in the `Antarctic-wide_BestFit_and_Mean_of_Sector_Sums` directory is the following:

```
[season]_total_best_fit_and_mean.ext
```

where season is a three-letter abbreviation of the season and `.ext` is the file extension as described in Table 8.

The naming convention for the files in `Antarctic-wide_Sum_of_Sector_BestFits` directory is the following:

```
Antarctic_wide_[season]_summed_sector_best_fit.ext
```

where season is a three-letter abbreviation of the season and `.ext` is the file extension as described in Table 8. There is also an Excel workbook that contains all the seasons (DJF, MAM, JJA, SON):

```
Antarctic-wide_all_seasons_summed_sector_best_fit.xlsx
```

### 1.5.4 Example Graph of Parameters

Figure 2 in Fogt et al. (2022a), titled Seasonal Antarctic total sea ice extent comparisons, 1905–2020, has plots of some of the parameters in these files.  Figure 7 of this document plots the same parameters.

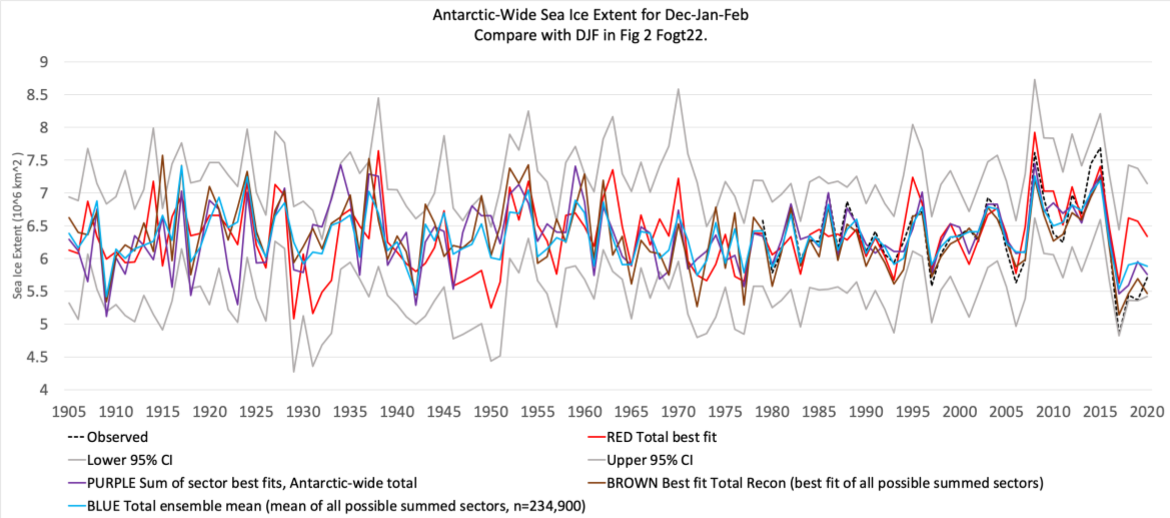


Figure 7. December-January-February SIE from selected parameters in files within the directories Antarctic-wide_BestFit_and_Mean_of_Sector_Sums (brown and blue lines), Antarctic-wide_Sum_of_Sector_BestFits (purple line), and Reconstruction_BestFitandMean (red line and confidence interval lines).

# 2 DATA ACQUISITION AND PROCESSING

## 2.1 Acquisition

These data were acquired by NOAA@NSIDC in November 2022 by downloading from the Figshare site referenced in Fogt et al. (2022a).

All the files in `Antarctic-wide_Sum_of_Sector_BestFits` were compiled at NSIDC using downloaded data.

## 2.2 Processing

### 2.2.1 Overview

Here we summarize the material in the Methods section of Fogt et al. (2022a). See that publication for a complete description of the processing steps that produced the reconstructions.

The reconstructions make use of the strong connection between regional and large-scale climate variability, represented by station meteorological data, climate indices, and sea ice. An overview of processing steps are as follows with more detail in section 2.2.2:

1. Obtain predictor data by obtaining station and index data for 1905-2020.
2. Derive observed SIE data. These are the calibration data.

3. Compute correlations between predictor data and seasonal mean SIE.
4. Select predictor data.
5. Perform principal component (PC) analysis of selected predictor data. PC weights describe the relationship between PCs and individual predictor data layers.
6. Regress a subset of PCs onto the 1979–2020 SIE time series. Regression coefficients describe the relationship between PCs and SIE.
7. Extended PCs back to 1905 with the regression coefficients to produce a reconstruction. There will be one reconstruction ensemble member for each predictor data layer.
8. Assign skill levels: Once seasonal reconstructions have been made for each of the five individual sectors and for Antarctica as a whole, the skill level of each ensemble member reconstruction is estimated and metrics for skill as well as measures of variability are included in the metadata files.

### 2.2.2 Detailed Description

#### 2.2.2.1 Obtain predictor data

Reconstructions are primarily based on sea-level pressure and surface temperature records from UCAR World Monthly Surface Station Climatology, also known as UCAR ds570.0. The authors use the 30 stations shown in Figure 1. Some station records were patched, updated, or corrected as described in Fogt et al. (2022a). Nine indices of modes of climate variability were also used. These are listed in Table 4. Seven of the nine are indices of ocean variability and may depict patterns that influence sea ice extent that the land-station-based pressure and temperature records cannot capture.

#### 2.2.2.2 Derive observed SIE data

Fogt et al. (2022a) obtain daily sea ice extent for the calibration period of 1972-2020 by summing the area of grid cells filled by sea ice at greater than 15% concentration in the passive-microwave-based NOAA/NSIDC Sea Ice Concentration Climate Data Record (Meier et al., 2021).

#### 2.2.2.3 Compute correlations between predictor data and seasonal mean SIE

Seasonal means of the station data are calculated for all 30 stations. Seasonal means of all nine indices are calculated. Linear trends in the time series of temperature, pressure, and indices are removed. This is referred to as detrending.

These time series are correlated with the detrended seasonal SIE for each sector individually and for the Antarctic overall. Sea ice extent takes time to respond to changes in the ocean and the atmosphere. Therefore, the authors correlate predictor data, which lead the Antarctic sea ice extent data for up to one season, at increments of one month. For example, if reconstructing sea ice extent for JJA, the correlations of predictors with JJA sea ice extent are computed for predictor data averaged in JJA (no lag), May–July (MJJ, one-month lead), April–June (AMJ, two-month lead) and MAM (three-month or one-season lead). The authors note that reconstructions are skillful only when this lag effect is accounted for.

#### 2.2.2.4 Select predictor data

Select the predictor data to use for each sector and season's reconstruction as well as for seasonal Antarctic-wide SIE reconstructions. It may be helpful to refer to section 1.5.2.3 Reconstruction_Metadata that describes the reconstruction metadata, the contents of `Amundsen_Bellingshausen_JJA_recon_master_stats.csv` (Figure 3), and Table 3 while reading this description.

Each reconstruction, for example that described by the metadata in `Amundsen_Bellingshausen_JJA_recon_master_stats.csv` (Figure 3) has one ensemble member that uses all temperature and pressure data. These are combined in a single layer called `Obs`, which stands for observations, in the metadata files. This `Obs` layer has 240 variables: all 60 station variables, (30 for temperature plus 30 for pressure), times 4 lags = 240. The lags account for variables leading the sea ice up to one season at one-month increments (that is, JJA sea ice would use JJA observations, MJJ observations, AMJ observations, and MAM observations).

In the next step, only timeseries of `Obs` that are correlated with observed SIE at p<10% are retained. If any climate index, and its monthly seasonal leads, are correlated within the p<10% threshold, a separate reconstruction that uses these data is added to the 10% network.

Four levels of significance are tested: 90%, 95%, 97.5%, and 99%. Altogether, there is a possibility of up to 41 ensemble members for each reconstruction. (Ten layers times four significance-level networks plus the one ensemble member that uses all station observations regardless of how well those observations correlate with SIE).

In the example shown in Figure 3, after processing is completed, there are 23 ensemble members for the Amundsen and Bellingshausen Seas June-August reconstruction, each represented by a line in the reconstruction metadata text file.

#### 2.2.2.5 Perform principal component (PC) analysis of selected predictor data

Each selected variable is standardized over the period 1905–2020, including predictor data that lead the sea ice by up to one season. PC analysis is then used to quantify the relationships that the PCs share with the predictor data, by means of the PC expansion weights.

#### 2.2.2.6 Regress a subset of PCs onto the 1979–2020 sea ice extent time series

The correlation between all PCs and observed SIE is calculated. The PCs are sorted by the absolute magnitude of their correlation with observed SIE. A subset of PCs is selected through iteration:

- The highest-correlated PC is regressed onto observed SIE to give one possible reconstruction.
- The next-highest-correlated PC is regressed onto observed SIE to give a second possible reconstruction.
- A coefficient of efficiency (CE) is computed for each possible reconstruction.
- This continues until all PCs have been regressed.

As PCs are added, the CE will rise, reach a peak, and then sharply decline. The cutoff for the subset of PCs that will be used is determined when the regression model that has the highest CE is reached. This process reduces the possibility of overfitting with too many PCs. The PCs are then regressed onto the calibration period SIE described in section 2.2.2.2. The resulting regression coefficients describe the relationship between PCs and SIE.

#### 2.2.2.7 Extend PCs back to 1905 and obtain a reconstruction

Fogt et al. (2022a) describe these steps:

> The reconstruction can then be obtained using the relationships that the PCs share with the sea ice extent time series (the regression coefficients) and the relationships that the PCs share with the individual predictor data (the PC expansion weights). Combining the regression and PC weights through matrix multiplications yields a value for each predictor variable, sometimes called the beta weights. These beta weights can then be used on the full length of the observed predictor data to obtain the reconstruction back to 1905. Alternatively, the PCs can similarly be extended back to 1905 from the predictor data and these PCs used with the regression coefficients to perform the reconstruction. Page 63.

#### 2.2.2.8 Assign skill levels and check for overfitting

Each reconstruction has as many ensemble members as there were selected predictor data layers. Metrics for the skill of each ensemble member as well as measures of variability are included in the metadata files.

Here we describe a process that takes place when the subset of PCs for each reconstruction is selected. The authors use a leave-one-out cross validation procedure to arrive at the ideal number of PCs to retain. They repeat the PC regression model one time for each year in the observed record (calibration period) from 1979 to 2020 (42 times), each time leaving out a moving window gap of 5 years of SIE. The center year SIE is assumed to be independent and is predicted using the beta weights generated in the reconstruction model.

The predicted values constitute an entirely predicted and independent time series called the validation reconstruction. The validation reconstruction is used to calculate the skill metrics: validation correlation (val r) and coefficient of efficiency (CE).

From Fogt et al. (2022a):

> This reconstruction is correlated to the observed sea ice extent data (giving the skill metric that we call validation correlation, or val r) and used separately to compute the CE, as discussed in [Cook et al., 1999]. After the CE is determined, the entire validation procedure is repeated by adding one more PC at a time to the regression model. The reconstruction that yields the highest CE is retained as the best-performing ensemble reconstruction for that particular data network and data layer. Selecting the highest CE aids in limiting model overfitting, as the CE will reach a peak, sharply decline and often go very negative as more PCs are added to the regression; this sudden decline is a key indicator of model overfitting despite the overall calibration correlation continuing to increase. Furthermore, retaining only a subset of the PCs makes this regression approach superior to multiple linear regression, as we effectively filter out unwanted noise by retaining only the most strongly correlated PCs. The entire procedure is then reproduced for each data network and data layer to yield at least five ensemble reconstruction members for each season and sector.
>
> Once all ensemble reconstructions are produced, the best-fit ensemble member (one for each season and sector of Antarctic sea ice extent) is first selected by extracting the reconstruction with the highest CE across all the ensemble reconstructions. To perform an additional check for model overfitting, the variance of the reconstruction in the period 1905–1978 is compared with the reconstruction variance in 1979–2020 and the observed variance in 1979–2020. If the reconstruction variance is significantly ($P < 0.01$) larger than the observed variance, there is a possibility of overfitting occurring. In these few cases, the best-fit reconstruction was selected as the next-highest CE where the variances were not statistically significantly different from the observed variance during 1979–2020, if possible. In some cases, there were no alternative reconstruction ensemble members with equal variances, and the highest CE was still selected as the best-fit reconstruction. This procedure was repeated for each season and sector to create the full array of ensemble and best-fit sector reconstructions. Pages 63-64.

The seasonal reconstructions for each sector and for the Antarctic overall were produced separately following the above PC regression model.

Each sector has at least five and usually many more ensemble member reconstructions. The variability in these can be used to estimate uncertainty. From Fogt et al. (2022a), one can

> …check how the sums of all possible ensemble member combinations across the sector compare. These combinations constitute more than 200,000 possible combinations for each season (there are over one million possible combinations in JJA). We therefore also provide an additional estimate from these combinations by selecting the combination that is most correlated with the observed sea ice extent time series (brown lines in Fig. 2). We use this and the full ensemble spread to further estimate reconstruction uncertainty. *For others seeking to utilize these reconstructions in the future, we recommend using the full ensemble to best characterize the full uncertainty, and the best-fit reconstruction for all the sector reconstructions as the individual member with the highest skill. For total sea ice, the sum of the best-fit sector reconstructions is ideal to use for further study, as this is more reliable than the best-fit reconstruction calibrated to the total sea ice, as described in the main paper. (Emphasis added).* Page 64.

Files containing the sum of the sector best-fit reconstructions are in the directory `Antarctic-wide_Sum_of_Sector_BestFits`. The directory `Reconstruction_Ensembles` contains the full ensembles.

## 2.3 Quality, Errors, and Limitations

The data are presented with several measures of quality and uncertainty. These are described in Fogt et al. (2022a).

The data product authors “recommend using the full ensemble to best characterize the full uncertainty, and the best fit reconstruction for all the sector reconstructions as the individual member with the highest skill. For total sea ice, the sum of the best fit sector reconstructions is ideal to use for further study, as this is more reliable than the best-fit reconstruction calibrated to the total sea ice” as described in Fogt et al. (2022a).

# 3 SOFTWARE AND TOOLS

NCAR Command Language code that can be used with the netCDF format files to plot the figures in Fogt et al. (2022a) is included. NOAA@NSIDC has not tested this code.

# 4 VERSION HISTORY

Table 9. Version History Summary

| Version | Release Date | Description of Changes | Citation |
|---|---|---|---|
| 1 | July 2023 | Initial release | Fogt, R. L., M. N. Raphael, and M. S. Handcock. 2023. *Seasonal Antarctic Sea Ice Extent Reconstructions, 1905-2020, Version 1*. [Indicate subset used]. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. https://doi.org/10.7265/55x7-we68. [Date Accessed]. |

# 5 RELATED WEBSITES

A Figshare site for Antarctic Sea Ice Reconstructions has the initial version of this data collection, posted 2021-11-16. The citation for that data collection is:

Fogt, Ryan (2021): Antarctic Sea Ice Reconstructions. figshare. Collection. https://doi.org/10.6084/m9.figshare.c.5709767.v1

The Fogt (2021) collection on Figshare is not identical to the collection that this User Guide describes. Additional associated or updated data may be found by searching contributions to Figshare by Ryan Fogt at: https://figshare.com/authors/_/2815045

# 6 ACKNOWLEDGMENTS

R. Fogt, M. Raphael, and M. Handcock produced these data under funding from NSF Grant Number 1744998, Collaborative Research: Understanding Contemporary Variability in Antarctic Sea Ice: Ensemble Reconstruction of Sea Ice Extent and Concentration for the 20th Century.

We thank the authors for contributing the data. Publication of these data was made possible by NCEI support to NOAA@NSIDC thorough the NOAA Cooperative Agreement with CIRES, NA17OAR4320101.

# 7 REFERENCES

## 7.1 User Guide References

Fogt, R.L., & Sleinkofer, A.M. (2021). Reconstructing Seasonal Antarctic Sea Ice Extent during the 20th Century [poster]. *Am. Met. Soc., 16th Conf. on Polar Met. and Ocean.* 1-4 June 2021.

Fogt, R.L., Sleinkofer, A.M., Raphael, M.N., and Handcock, M.S. (2022a). A regime shift in seasonal total Antarctic sea ice extent in the twentieth century. *Nat. Clim. Chang*. 12, 54–62. https://doi.org/10.1038/s41558-021-01254-9.

Meier, W.N., Fetterer. F, & Stewart, J.S. (2019). Validation of AMSR2 sea ice fields with an updated U.S. National Ice Center ice chart climatology [poster]. *Am. Met. Soc., 15th Conf. on Polar Met. and Ocean. in Boulder*, CO, 20-24 May 2019.

Meier, W. N., Fetterer, F., Windnagel, A.K., & Stewart, J.S. (2021). NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 3. Boulder, Colorado USA. National Snow and Ice Data Center. https://doi.org/10.7265/efmz-2t65.

Raphael, M. N. & Hobbs, W. (2014). The influence of the large-scale atmospheric circulation on Antarctic sea ice during ice advance and retreat seasons. *Geophys. Res. Lett.* 41, 5037–5045. https://doi.org/10.1002/2014GL060365.

## 7.2 Research Papers That Use These Data

This is not an exhaustive list but is research that has come to our attention that use these data.

Fogt R.L., Dalaiden, Q., and Zarembka, M.S. (2022b). Understanding differences in Antarctic sea-ice-extent reconstructions in the Ross, Amundsen, and Bellingshausen seas since 1900. *Past Global Changes Magazine*. 30(22), 74-75. https://doi.org/10.22498/pages.30.2.74.

# 8 DOCUMENT INFORMATION

## 8.1 Author

F. Fetterer wrote this User Guide based on the material in Fogt et al. (2022a). R. Fogt reviewed it and J. Roebuck and A. Windnagel edited it prior to publication.

## 8.2 Publication Date

July 2023