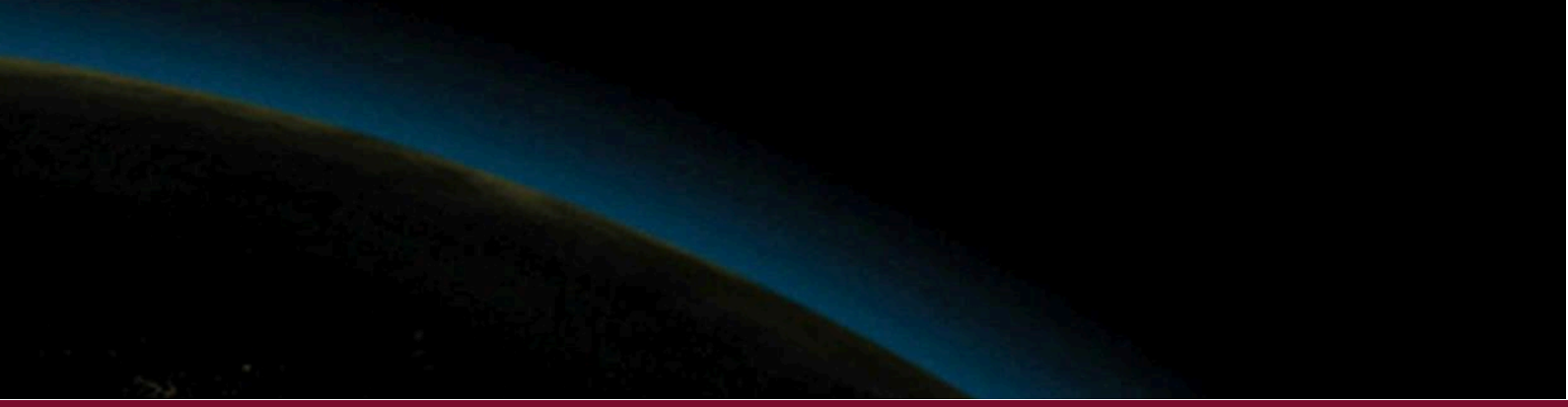

ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

CARRA2: Inventory and final description of all input data for data assimilation and for SST, sea ice, physiography and glacier albedo

Issued by: MET Norway/Per Dahlgren

Date: 18/09/2024

Ref: C3S2_D361a.2.1.2-InputDataAssim-v3.docx

Official reference number service contract: 2022/C3S2_361a_METNorway/SC1

PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

IMPLEMENTED BY ECMWF

Contributors

DMI
Bjarne Amstrup
Kristian Pagh Nielsen
Jacob L Høyer
Kent Lauritsen
Ida L Olsen
Pia Englyst

MET Norway
Harald Schyberg
Per Dahlgren
Mariken Homleid
Mari Anne Killie

IMO
Bolli Palmason

FMI
Reima Eresmaa

GEUS
Jason Box

Table of Contents

1. Introduction

This document is the deliverable D361a.2.1.2 for the framework contract *C3S2 361a: Next Generation Copernicus pan-Arctic Regional Climate Reanalysis.* The code system and the dataset to be produced is referred to as CARRA2.

With the CARRA systems, regional reanalysis adds value to the global reanalysis ERA5 by increased horizontal spatial resolution and by using input data sets either not used in the global system or used with less details. In particular, as an Arctic reanalysis, we have prepared enhancements by special attention to adapting surface input data sets for "cold surfaces" such as snow, sea ice and glaciers. With CARRA2 we evolve upon the previous CARRA1 system in many respects, impacting the input data sets to be used. Main such development elements include (see in italics the elements directly related to the input data):

- Extension of the reanalysis domain to a much larger area to provide pan-Arctic coverage, which contributes to a better reanalysis database for studies and applications covering the entire Arctic,
- planned extension of the reanalysis to a period of 40 years (1985-2025),
- *improvement in observation data use in upper air data assimilation by introduction of variational quality control*,
- *a higher resolution (300m) physiographic database (PGD) with ECOCLIMAP Second Generation (SG) with significantly improved data quality and representativeness for the surface conditions*,
- *with the extension to pan-Arctic domain, updated time consistency efforts for satellite-based surface input data for sea surface temperature, sea ice concentration, snow cover and glacier albedo*.

To prepare for possibly going further back, we will as far as possible prepare input data sets starting from 1982. In addition to these main improvement/extension elements, we also consider whether there are recent updated versions of the climatological input data records from existing external providers.

In the next section we give an overview of the input data sets and sources used, how they are prepared or interfaced to our system. Here we cover input data sets for upper-air data assimilation, for surface data assimilation as well as for physiography and surface properties. We also describe input data on radiative forcing from greenhouse gases and aerosols. A summary list of the various input data sets with references is provided in the Appendix 1.

2. CARRA2 input data set processing and use

1.1 Observations for upper-air and surface data assimilation

1.1.1 Conventional observations from the ECMWF MARS archive

We use conventional observations available in the MARS archive at ECMWF. Conventional observations include SYNOP surface pressure over land, SYNOP-SHIP surface pressure over sea, BUOY surface pressure over sea, TEMP profiles of temperature, wind and humidity, and aircraft (AIREP, AMDAR, ACAR) measurements of wind and temperature.

For 10m SYNOP-SHIP winds over sea, we will not use those after identifying issues caused by their instrument height corrections which cause negative impacts from them on the analysis quality. Our system also does not use METAR observations. These have lower precision than similar SYNOP observations, are often redundant with SYNOP observations and have not been used operationally before in the HARMONIE-AROME system. However, METARs do fill gaps in the observing system in some places (e.g. Alaska) and were used in ERA5.

Up to 2002 the ERA40 BUFR archive is used (MARS class e4). After that the data are fetched from MARS class OD using STREAM=DA until 20040629 and then STREAM=DCDA until 20131118 and from then STREAM=LWDA.

As in CARRA1, radiosonde temperatures are bias-corrected using pre-defined corrections (*Haimberger et al, 2012, Hersbach et al, 2020*). These corrections are available from the start of the CARRA2 reanalysis period to the end of 2016. From 2017 and onwards no corrections are applied to radiosonde data. The corrections available after 2016 are the operational ECMWF corrections but these are not applied in any HARMONIE-AROME operational configuration, and therefore we do not apply them in CARRA2 The modern day radiosonde network inside the CARRA2 domain from 2017 is assumed to be of high quality.

The system is prepared for ingestion of conventional data directly from the MARS archive without any pre-processing steps.

1.1.2 Conventional local surface observations from non-MARS/GTS sources

In CARRA1 a big effort was put into collecting supplementary and quality-assured longer-latency observation station data archived locally at each of the Nordic weather services. This data set also includes observations over the Greenland ice sheet from ASIAQ coastal stations (Greenland Survey, Greenland), PROMICE (Programme for Monitoring of the Greenland Ice Sheet via GEUS, Denmark) and GC-NET (Greenland Climate Network, USA/Denmark). This data set (which we refer to as the “LOCOBS” data set) is available for use in CARRA2 (also shared with CERRA) and is going to be used in the reanalysis spin up and production streams.

There now exists updates to LOCOBS to be used in CARRA2 with:

1. Quality corrected and complementary observations from 5 stations in the Svalbard area: Verlegenhuken (80.0592N, 16.25E), Edgeøya (78.2508N, 22.8225E), Karl XII-øya (80.653N, 25.008E), Kvitøya (80.1058N, 31.4643E) og Gråhuken (79.7875N, 14.4708E). For the first four of these stations data has been quality corrected, and additional observations from periods not present in the CARRA1 data set have been added. Gråhuken is an additional station.

2. Stations over Denmark, Faroe Island and Greenland from the Danish Meteorological Institute (DMI). Most of these observations are additions to the currently available LOCOBS data set.

3. A new delivery of local observation data from Iceland. These are mainly small corrections to the original LOCOBS data set.

4. Improved version of some of the Greenland campaign observation data set which is a combination of PROMICE and GC-NET. These data were already used in CARRA1 but are undergoing further analysis and checking now. This includes e.g. revised coordinates and improved quality control. These data have very recently been made available but have to undergo some further processing by personnel at DMI before they can be used in CARRA2.

For 1-3, we have created an updated LOCOBS data set with these updates in it. The Greenland observations mentioned in point 4 require some more work before the updated quality control and position information can be included in the LOCOBS data set used in CARRA2. Therefore, we will start production with the original version (as in CARRA1) without those updates and consider inclusion later.

1.1.3 Microwave radiances

All observations used are already stored at ECMWF. Microwave (MW) radiances from the Microwave Sounding Unit (MSU), the Advanced Microwave sounding Unit A and B (AMSU-A and AMSU-B), the Microwave Humidity Sounder (MHS) are used. On top of that we also use the Advanced Technology Microwave Sounder (ATMS) and the Microwave Humidity Sounder (MWHS2) from the Fengyun-3 satellites. The system is prepared for ingestion of these data directly from the MARS archive without any pre-processing steps.

- MSU radiances will be used from the beginning of production until 2006 from satellites NOAA 6-14
- AMSU-A used from 1998 onwards from satellites: NOAA 15,16,17,18,19, AQUA, METOP-A,B,C
- AMSU-B used from 2000-2009 from: NOAA 16,17
- MHS used from 2005 onwards from: NOAA 18,19, METOP-A,B,C
- ATMS used from 2012 onwards from: S-NPP, NOAA 20,21
- MWHS2 used from 2016 onwards from: FY-3C,D,E

1.1.4 Infrared radiances

In the previous service contracts (CARRA1, CERRA), historical and near real time IASI data files from the ECMWF archives were pre-processed before the data was used in the data assimilation. The pre-processing consisted of reducing the number of channels and only keeping data north of 20 degrees latitude. This pre-processing is continued for near real time data in the CARRA-TU service contract and there is no need to do any additional processing of IASI in CARRA2 since the pre-processed data are archived on ECFS. CARRA2 also uses CrIS radiances, and these data are also already in the ECMWF archives.

- IASI radiances used from 2007 onwards from: METOP-A,B,C
- CrIS radiances used from 2012 from: S-NPP, NOAA 20,21

1.1.5 Atmospheric Motion Vectors (AMV)

The bulk of AMV data used in CARRA2 are from polar orbiting satellites. Several reprocessed data sets exist at ECMWF in ECFS and MARS and were used in CARRA1 and will be used in CARRA2 as well. Since the domain of CARRA2 extends further south than the CARRA1 domains we at first thought AMVs from geostationary satellites could be used as well. It however turned out that all geostationary AMVs are blacklisted in HARMONIE north of 50 degrees latitude. With doubt on whether the system can benefit from these, we have chosen not to use them.

From the start of the CARRA2 reanalysis reprocessed AMVs from NOAA satellites are used and in 2002 the first MODIS (from the TERRA satellite) data appears. From 2006 AMVs from METOP satellites comes in and from 2014 there are so-called dual-AMVs from METOP. The availability of AMV data from polar orbiting satellites is well described in *Hersbach et al, 2020* Figure 6. From 2015 CARRA2 will also use AMVs derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument on the S-NPP satellite. VIIRS AMVs are also available on the subsequent NOAA-20 and NOAA-21 satellites.

1.1.6 Scatterometer

HARMONIE-AROME uses reprocessed scatterometer data from the EUMETSAT OSI-SAF products. In CARRA1, scatterometer winds, reprocessed by EUMETSAT were retrieved and archived on ECFS. The reprocessed data set covers the period starting from 1993 with data from the ERS satellites. It also contains reprocessed winds from the satellites QuikSCAT (1999-2009), METOP-A (2007-2014) and Oceansat-2 (2009-2014). From 2014, operational OSISAF data from the METOP satellites will be used. The data set stored on ECFS is global and is used in CARRA2 as well. For the NRT stream, the corresponding data input infrastructure prepared in the CARRA-Timely Updates contract is used.

1.1.7 Radio occultation

For CARRA2 the plan for radio occultation was to use data following the recommendation from the DMI team behind production of the radio occultation data set of the ROM SAF (the EUMETSAT Radio Occultation Meteorology Satellite Application Facility): They recommended using the latest reprocessed radio occultation Climate Data Record (CDR) data set, ROM SAF CDR v2.0, which was due to be released end of 2023. This is assumed to be an improvement over the datas et ROM SAF CDR v1.0 as used in CARRA1. It turned out that the v2.0 data set is delayed and is rescheduled to be released at the end of 2024, or more probably only in 2025.

Hence, the new improved data set was not, and will not, be ready in time for CARRA2 production and we use instead a fallback solution which is to use the already existing v1.0 data set for the CARRA2 production.

1.1.8 Satellite snow observations

In CARRA1 CryoClim satellite imager-based snow extent was used in addition to snow depth observations. CryoClim is a global, optical snow product with 5 km resolution based on historical AVHRR GAC data (EUMETSAT CM SAF Clara A2, A2.1). A Bayesian approach is used to combine information from optical and infrared AVHRR channels to estimate the probabilities for the classes “snow”, “snow-free land” and “cloud”.

CryoClim has been updated for use in CARRA2, with the input satellite climate data record CLARA-A3 from EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF). This data set was released in April 2023 (DOI: 10.5676/EUM_SAF_CM/CLARA_AVHRR/V003). CLARA-A3 covers 1979 – 2020, and the data record continues after 2020 with the production of an interim climate data record (ICDR). The ICDR data is downloaded at MET Norway with a 14 days (or less) delay, allowing regular updates of the CryoClim snow cover product for use in CARRA2.

Updated CryoClim data covering the period 1985 to end of July 2023 has been made ready for use at CARRA2 production start and will be prolonged when needed.

1.2 Physiography

Physiographic data in the Arctic in the past often had lower quality and more issues than at lower latitudes, so exploiting more recent publicly available data sets has had some emphasis in CARRA2. For the physiographic data base, significant upgrades from CARRA1 and adaptions for the larger pan-Arctic domain were done. The following data sets are used for the physiography in CARRA2:

- **Orography:** Digital elevation model (DEM) at 150 m resolution using Copernicus GLO-90 data with a few updates using ArcticDEM
- **Land cover:** ECOCLIMAP-SG at 300 m resolution with updates to glacier mask (RGI), lakes and vegetation over Greenland, Iceland, Svalbard and Faroe Islands

- **Lakes:** 300 m global depth map (GLDB)
- **Soil:** Global database (SoilGrids) of sand and clay fractions at 250 m resolution and soil organic carbon (SOC)

The input digital elevation map (DEM) used for the model orography is the upscaled Copernicus GLO-90 DEM to 150 m resolution. This upscaling was from the original 90 m resolution to make file sizes more tractable. A few known errors in this data base were fixed with the ArcticDEM 90 m mosaic data set.

The reference HARMONIE-AROME version 43 system introduces the *ECOCLIMAP Second Generation* (SG) database for an improved physiographic representation of surface properties as needed in the HARMONIE-AROME surface scheme. While the ECOCLIMAP-SG data is a general improvement over those of the predecessor ECOCLIMAP-II with a higher resolution (ca 300 m) with various corrections and updates, there were many errors at high latitude areas that needed to be addressed. The glacier extents were updated with Randolph Glacier Inventory (RGI) and kept fixed throughout the whole reanalysis period. The RGI data available aims to have the glacier extents set as close as possible to the year 2000, although its input data used can represent a substantial range of dates in some regions. We consider this year to be reasonably representative for the CARRA2 reanalysis period. Lake masks and vegetation covers were updated with different data sets over Greenland, Iceland, Svalbard and Faroe Islands (see table in Appendix 1). Tree heights were updated over Iceland with local data. In some cases, the coastlines needed to be improved with land-cover updates - especially in Greenland and Svalbard.

1.3 Glacier surface albedo

For the CARRA2 reanalysis glacier surface albedo will be based on data from three sources:

- Data from AVHRR APP-X is available from 1981 to 2000 and in CARRA2 this data set will be used from the start of production until 1999.
- Between 2000 and 2020 the data will be from MODIS.
- The glacier surface albedo data from 2021 will be based on Sentinel-3 satellite data. Mainly from the OLCI instrument. Here the S3-SICE algorithm is used for the retrieval, which is available at https://github.com/GEUS-SICE/SICE/.

Using daily varying broadband albedo values for the full reanalysis time series in this way gives a more realistic and consistent data set for the radiative exchange scheme than relying on fixed climatological fields for the whole or parts of the period. Albeit from different sensors and satellites, all these three sources provide similar information. Kohkanovski et al (2020) demonstrated good consistency between albedos derived from MODIS and from Sentinel-3 when comparing to albedos derived from surface stations.

Figures 1 and 2 illustrate the products from AVHRR and Sentinel to be used, and for CARRA2 the data set has been extended to also cover glaciers outside the CARRA1 domains, such as Alaska and the Canadian archipelago.

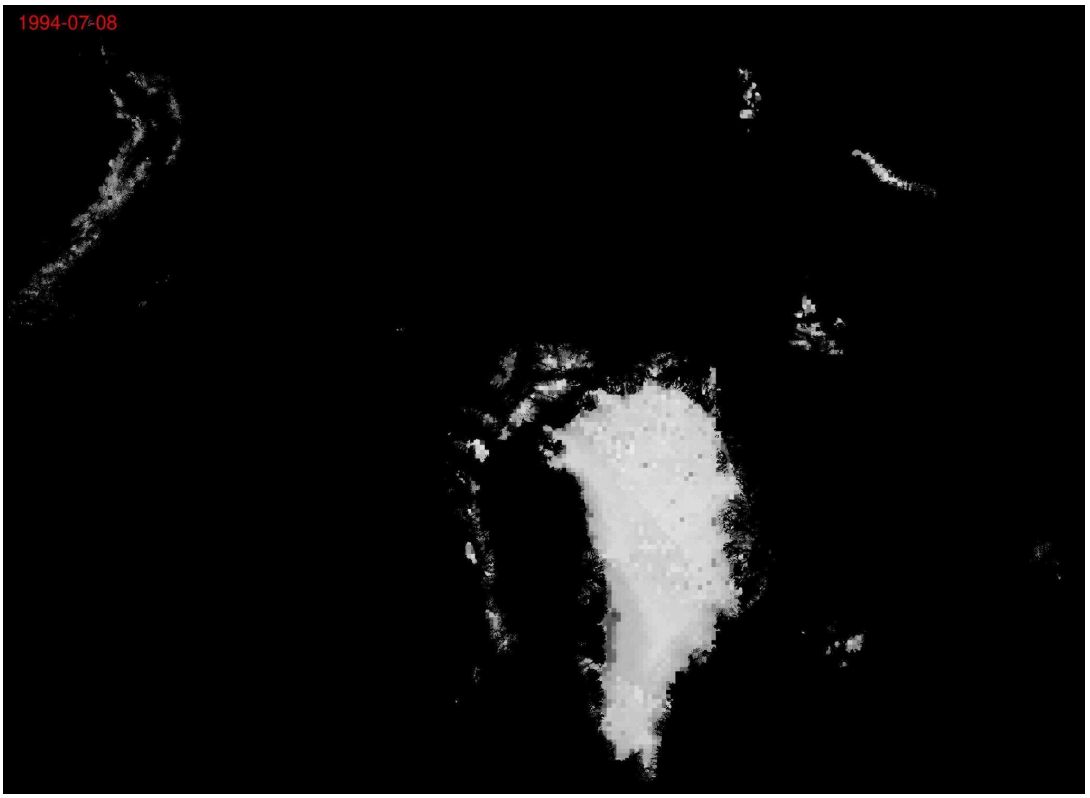

Figure 1. Glacier albedos derived from AVHRR data with the SICE algorithm for the 8th of July 1994.

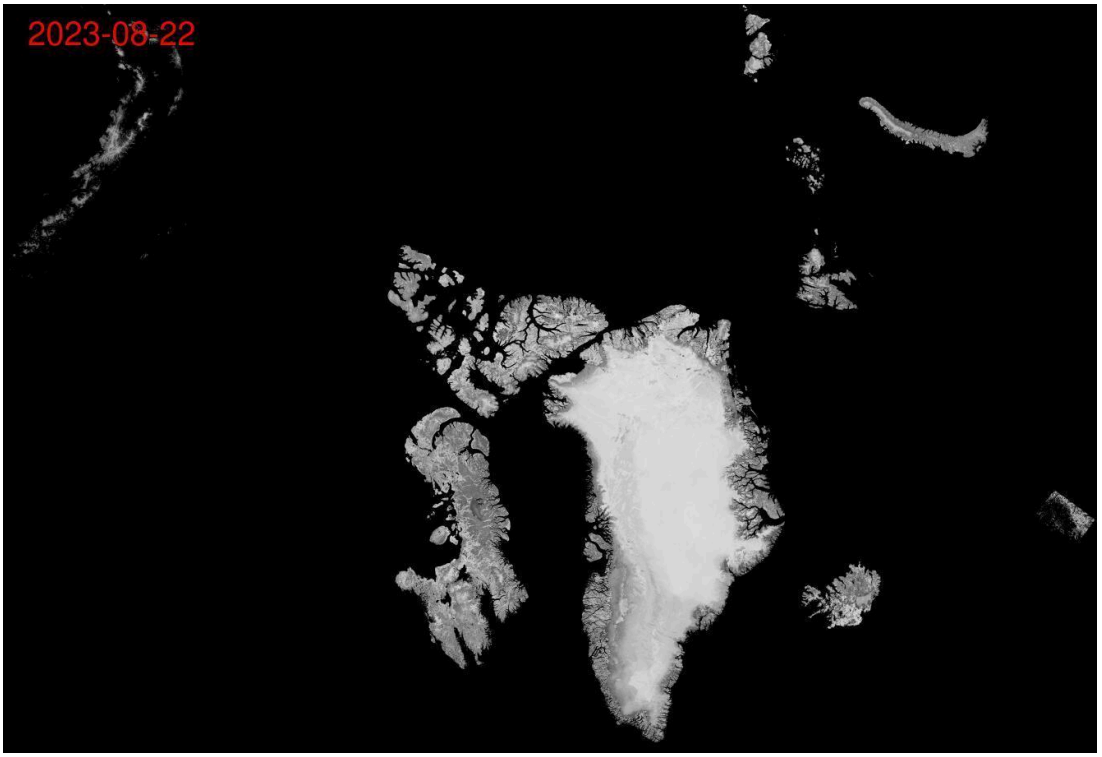

Figure 2. Glacier albedos derived from Sentinel-3 OLCI data with the SICE algorithm for the 22nd of August 2023.

The plan is to use broadband albedo data for all grid boxes that include glaciers, as was done in CARRA1. Here, we modify the UV-VIS spectral albedo to fit the broadband satellite-derived albedo. For the Sentinel-3 albedos, the infrared snow albedo estimation also depends on the optical snow grain diameter. This is made possible by the usage of the more advanced Explicit Snow scheme in the CARRA2 setup.

1.4 Sea surface temperature and sea ice

High resolution sea surface temperature (SST) and sea ice data were used in CARRA1 by combining several reprocessed data sets. For CARRA2, the same approach will be used, but extension and reprocessing are used to keep a consistent SST and sea ice concentration (SIC) product to suit the needs for an extended spatial and temporal coverage and also due to availability of new and updated external products. For the Baltic Sea and surrounding areas, this is handled separately with regional historical SST and SIC analysis products (Høyer, 2016, Høyer and Karigali, 2016). For SIC, different datasets are used and combined with sea ice chart information over time to ensure that the best available product is always used. For that reason, it is important to ensure temporal and spatial consistency. To ensure consistency, several filters have been developed and applied and the temporal and spatial consistency has been investigated (see section “Processing and filtering of SST and SIC”). Here, it is found that the CARRA2 sea ice extent aligns very well with the climate data record from the Ocean and Sea Ice Satellite Application Facility of EUMETSAT, OSI-450a CDR. The following section describes the different input data that has been used.

1.4.1 Input data

SST: For SST we use the newest version (i.e. version 3) from ESA SST CCI Level 4 Analysis Climate Data Record (CDR) for the period 1982-2021 and an integrated climate data record (iCDR v3) after 2021. Both the CDR and the ICDR are generated using the same software and algorithms developed as part of the ESA SST CCI. The Copernicus Marine Service (CMEMS) regional Baltic Sea - Sea Surface Temperature Reprocessed analysis (produced by DMI) is always used in the North Sea and Baltic Sea.

SIC: For SIC a combination of three passive microwave (PMW) satellite SIC climate data records are used in a prioritized order as listed below:

- OSI-458 is the third version of the EUMETSAT OSISAF SIC CDR based on medium-resolution (15–25 km) passive microwave satellite data from the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E; June 2002 to October 2011) and AMSR2 (July 2012 to May 2017). For 2021-2023 a temporal extension of OSI-458 is used.
- SICCI-HR-SIC is a high(er) resolution SIC CDR from ESA Sea Ice CCI (SICCI) covering the period 1991-2020. It is an advanced “pan-sharpened” version of OSI-450a which means that the high-frequency, high(er) resolution near-90 GHz imagery of the SSM/I and SSMIS satellite missions has been used to obtain higher resolution SIC fields of 12.5 km (compared to the original OSI-450a fields of 25 km).

- OSI-450-a is the third version of the EUMETSAT OSISAF SIC CDR based on coarse resolution (30-60 km) passive microwave satellite sensors (SMMR, SSM/I and SSMIS). It covers 1978-2020.

Apart from passive microwave SIC products, the following sea ice chart information (digitized ice map from operational sea ice monitoring services) is also used: For the North Sea and Baltic Sea, regional sea ice concentration is always used from the CMEMS Baltic Sea ice concentration, extent, and classification dataset. This is a gridded product based on sea ice charts produced by FMI and SMHI ice analysts based on manual interpretation of satellite data and ground truth, which is available from the beginning of the sea ice season 1980-1981 and until today at the nominal resolution of 1 km and weekly temporal availability.

Landfast ice (sea ice that is immobile due to its attachment to a coast) is included from the U.S. National Ice Center (NIC) weekly or biweekly ice-chart time series, which means that the CARRA2 SIC is set to a 100 % in these cases. The NIC ice charts are moreover used to filter the passive microwave sea ice concentration data (see Filter 4 below).

For an overview of the SST and SIC Input data see Figure 3 below.

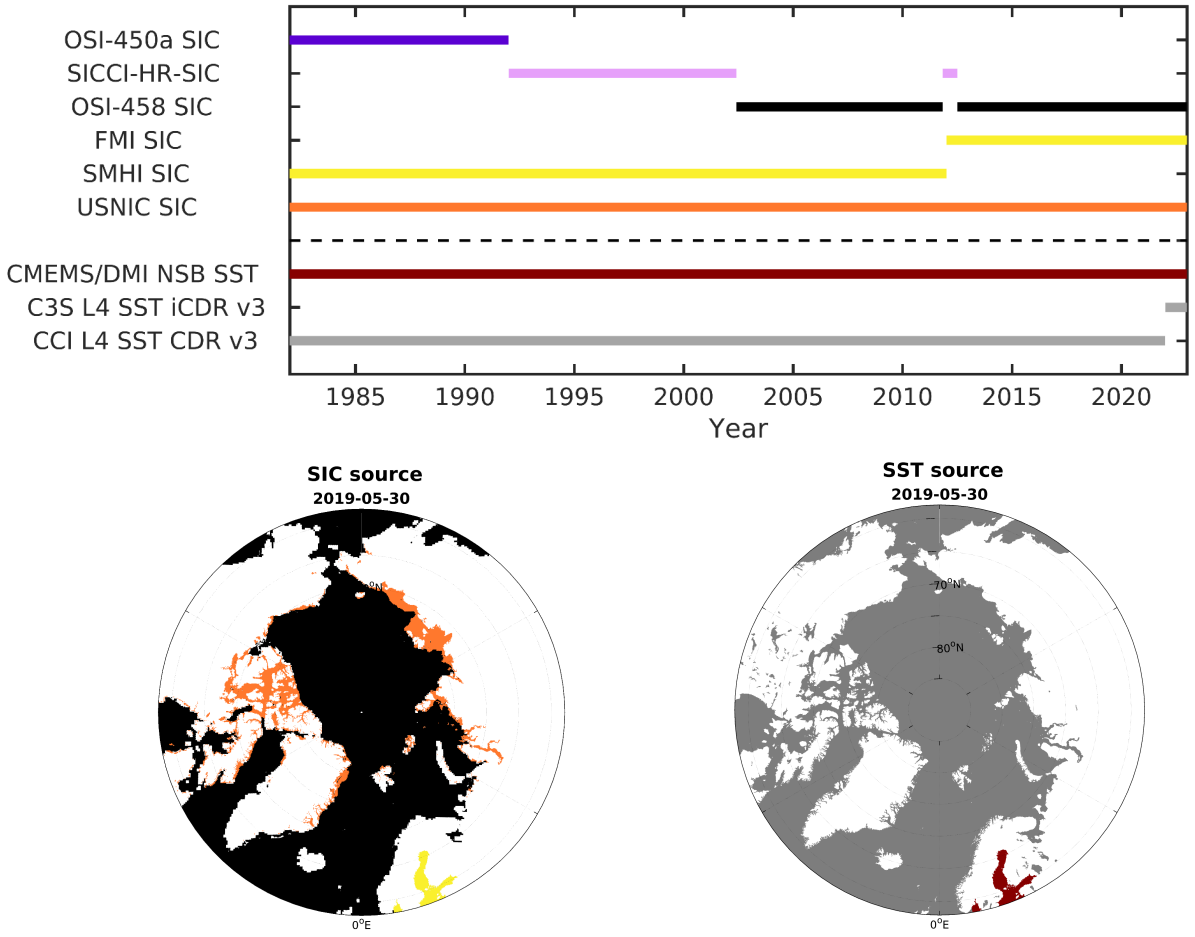

Figure 3. The top plot illustrates the temporal coverage of the SIC and SST data as they are combined and included to generate the SIC and SST fields for CARRA2. The maps provide examples of the spatial coverage of the different products that are combined for 2019-05-30 for SIC and SST, respectively. The colors in the maps correspond to the colors/products listed in the top plot.

1.4.2 Processing and filtering of SST and SIC

All input data is regridded to a 0.05-degree regular latitude-longitude grid by using bilinear interpolation for PMW SIC data, whereas ice chart information is regridded using 2D nearest neighbor interpolation. Subsequently, the data is combined as shown in Figure 3. The bar plot illustrates how the data is temporally combined while the maps illustrate how the data is spatially combined for one day. The PMW SIC data is afterwards filtered to increase the accuracy and consistency of the final SIC used in CARRA2.

The following 4 filters are applied in the order below and the sea ice concentration is set to zero if any of the conditions are true:

Filter 1: The grid cell is less than 25 km from the coast and the minimum sea ice concentration within a 12.5 km radius is zero.
Filter 2: The grid cell is less than 25 km from the coast and the SST value is above 3 degrees Celsius.
Filter 3: The SST value is above 8 degrees Celsius.
Filter 4: The two subsequent sea ice charts, which are closest in time, have recorded a sea ice concentration of zero for grid cells that are less than 75 km from the coast.

As an additional filter, an inter-sensor bias adjustment for OSI-450-a and SICCI-HR-SIC is performed using OSI-458 as reference.

After filtering, the data is extrapolated into fjords, lakes and near coastal areas, which has not already been filled with data from the ice-charts. An iterative extrapolation method is used where each gridpoint with a null value is assigned the mean value of the 9 neighboring grid cells. A land mask is applied between each step to ensure that values are not extrapolated over land.

After filtering we obtain a higher consistency between the SIC products. As shown in Figure 4 we observe similar trends in sea ice extent (SIE) between OSI-450-a and CARRA2, and we also see a very close proximity between the actual sea ice extent values of CARRA2 and OSI-458.

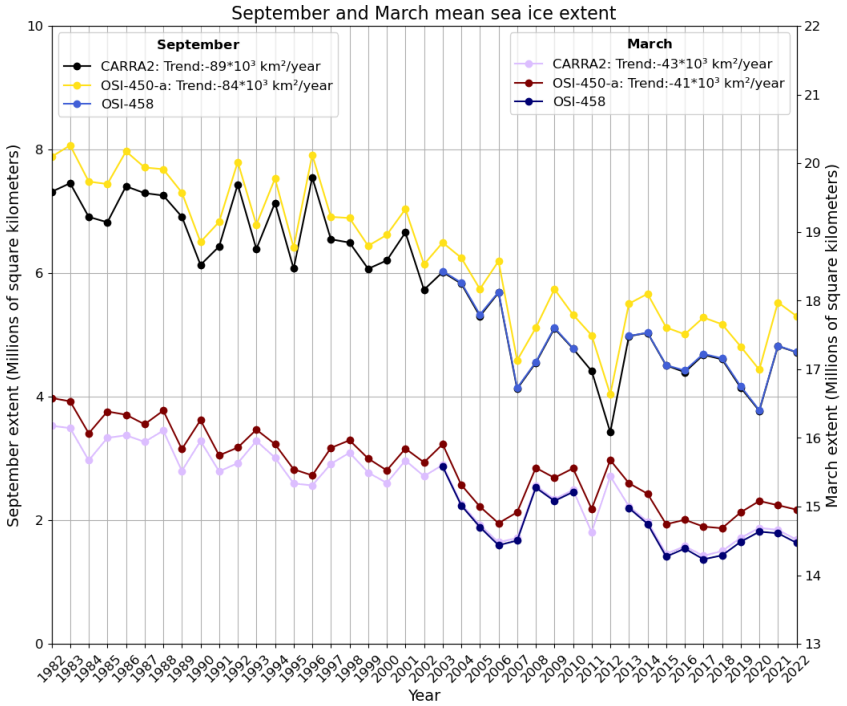

Figure 4. Monthly mean sea ice extent from 1982-2022 of the CARRA2, OSI-450-a and OSI-458 sea ice products for March and September, respectively.

1.5 Time varying radiative forcing from aerosols and greenhouse gases

The variation in radiative impact of aerosols is recognized to have an impact on past climate change, where mainly large volcanic eruptions have affected the radiative balance on the global scale. In the CARRA2 time period, the only such major event was the Mount Pinatubo eruption in 1991. The stratospheric sulfate aerosols affect the solar irradiance reaching the surface. At most, the related aerosol optical depth was 0.17 as a global average in this event. The main impact for the Arctic reanalysis, however, comes from the use of SSTs and lateral boundary conditions that are implicitly affected by the Pinatubo eruption - even when we don't add the correct stratospheric aerosols in the radiative transfer scheme. On top of that we have included a realistic time varying aerosol load applying the generator for stratospheric aerosol optical depth of Toohey et al, 2016. This addition of stratospheric optical depth is inserted for the period of the Pinatubo eruption and the following years where this extra aerosol load was expected to have a considerable impact, in the period 1991-1995.

For time variation in greenhouse gases the HARMONIE-AROME system reference has used a time series provided back in 2000 and therefore not the latest greenhouse gas data and scenarios available. The CARRA2 system has been updated to use CMIP6 historical greenhouse gas data up to 2014 and scenarios afterwards. A northern hemisphere asymmetry in the CH_4 data was included by multiplying these CMIP6-based values with 1.05. The reason for this was comparisons with observations (Kristian Pagh Nielsen, personal communication).

2. Monitoring and consideration on time consistency

As outlined for several of the input data types above, we need to make use of different input sources in different time periods of the reanalysis. An overview of times for change events in surface input data is given in Appendix 2 below. Although we have used approaches in the data preparation and preprocessing to minimize time consistency issues as far as feasible, it is of interest to particularly monitor the times with such input change occurrences. We will then do checks with respects to biases and jumps in the produced reanalysis. For this we will monitor time series, and in particular verification statistics in the post-production verification. This will be used to detect jumps in the time-series or other issues and will enable further investigation and possible adjustments in the use of the input data (and possibly reruns) when issues are seen. When such issues are detected and affect the data set, this will be documented in the user guidance document.

3. Summary

The present deliverable is the inventory and final description of all input data for data assimilation and for SST, sea ice, physiography, glacier albedo and time varying radiative forcing for CARRA2. In particular, we have described inclusion and adaptation of new data sets, updated and extended data sets relative to what was available in CARRA1. This also comprises extensions necessary to adapt to a large pan-Arctic domain. Such extended use of input data sets, in particular for surface and screen

level information, is part of the basis to enable CARRA2 to add information and value with respect to the host model ERA5.

Fetching routines and interfaces have been prepared and are in place for all these data sets to be used in the CARRA2 production. For reference and summary, we provide below a list of literature references and in the Appendix 1 an overview table of the input data sets used in CARRA2. In the overview table we have specifically marked with an asterisk the data types which constitute a significant addition, extension or increased resolution relative to what is used in the global reanalysis ERA5.

Literature references for data sets

Box, J.E., D. van As, K. Steffen, 2017. Greenland, Canadian and Icelandic land ice albedo grids (2000-2016), GEUS Bulletin, 38, 53-56. 10.34194/geusb.v38.4414.

Danielson, J.J., and Gesch, D.B., 2011, Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073, 26 p.

Gleisner, H., Lauritsen, K. B., Nielsen, J. K., and Syndergaard, S., 2020: Evaluation of the 15-year ROM SAF monthly mean GPS radio occultation climate data record, Atmos. Meas. Tech. 2020 vol 13, no 6, pp 3081-3098. DOI 10.5194/amt-13-3081-2020 . Available from https://amt.copernicus.org/articles/13/3081/2020/ .

GLO-90, 2024: produced using Copernicus WorldDEM™-90 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018 provided under COPERNICUS by the European Union and ESA; all rights reserved. https://esa.maps.eox.at/copernicus-dem_eula.pdf

Haimberger, L., Tavolato, C. and Sperka, S. (2012) Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighboring stations. Journal of Climate, 25, 8108–8131.

Hersbach, H, Bell, B, Berrisford, P, et al. The ERA5 global reanalysis. Q J R Meteorol Soc. 2020; 146: 1999–2049. https://doi.org/10.1002/qj.3803 .

Homleid, M, H. Körnich, M. Lindskog and E. Olsson, 2018: Surface input data, Temperature, relative humidity and snow. C3S deliverable report C3S_D322_Lot2_MetNorway/SC1.

Homleid, Mariken and Mari Anne Killie, 2013: HARMONIE snow analysis experiments with additional observations, MET report 06/2013, available on https://www.met.no/publikasjoner/met-report/met-report-2013 .

Høyer, J. L., 2016: Product user Manual For Baltic Sea SST Level 4 Product SST_BAL_SST_L4_NRT_OBSERVATIONS_010_007_b, CMEMS user manual: [online] Available from: http://cmems-resources.cls.fr/documents/PUM/CMEMS-OSI-PUM-010-007-b.pdf .

Høyer, J. L. and Karagali, I., 2016: Sea Surface Temperature Climate Data Record for the North Sea and Baltic Sea, J. Clim., 29(7), 2529–2541, doi:10.1175/JCLI-D-15-0663.1.

Jacek Andrzej Urbański, Monitoring and classification of high Arctic lakes in the Svalbard Islands using remote sensing, International Journal of Applied Earth Observation and Geoinformation, Volume 112, 2022, 102911, ISSN 1569-8432, https://doi.org/10.1016/j.jag.2022.102911

Johansen, B., Tømmervik, H. & Karlsen, S.R. 2009. Vegetasjonskart over Svalbard basert på satellittdata. Dokumentasjon av metoder og vegetasjonsbeskrivelser. NINA Rapport 456. 54 s. Tromsø, mars, 2009

Killie, M.A. Steinar Eastwood, S. and, Sørensen, A. M., 2018: Improvement and Validation of the Optical Component Used for Snow Mapping in CryoClim. MET report 12/2018, ISSN 2387-4201. Available from https://www.met.no/publikasjoner/met-report/met-report-2018/_/attachment/download/5e28741e-ccb3-4dd9-8e12-fb131661fc13:fb74c03d869d5180e4de8ff90cda39cc03549e89/MET-report-12-2018.pdf .

Kokhanovsky A, Box JE, Vandecrux B, Mankoff KD, Lamare M, Smirnov A and Kern M, 2020: The Determination of Snow Albedo from Satellite Measurements Using Fast Atmospheric Correction Technique. Remote Sensing 12(2), 234.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S. and Pedersen, L. T., 2019: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13(1), 49–78, doi:10.5194/tc-13-49-2019.

Lindskog M, Dahlbom M, Thorsteinsson S, Dahlgren P, Randriamampianina R, Bojarova J, 2012: ATOVS Processing and Usage in the HARMONIE Reference System. HIRLAM Newsletter 59: 33-43, available via http://www.hirlam.org.

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., Good, S., McLaren., A., Rayner, N. and Donlon, C., 2016: ESA Sea Surface Temperature Climate Change Initiative (ESA SST CCI): Analysis long term product version 1.1, , doi:10.5285/2262690A-B588-4704-B459-39E05527B59A .

Mojtaba Karami, Andreas Westergaard-Nielsen, Signe Normand, Urs A. Treier, Bo Elberling, Birger U. Hansen, A phenology-based approach to the classification of Arctic tundra ecosystems in Greenland, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 146, 2018, Pages 518-529, ISSN 0924-2716, https://doi.org/10.1016/j.isprsjprs.2018.11.005

OSI SAF (2016): ASCAT L2 12.5 km winds data record release 1 - Metop, EUMETSAT SAF on Ocean and Sea Ice, DOI: 10.15770/EUM_SAF_OSI_0007. http://dx.doi.org/10.15770/EUM_SAF_OSI_0007 .

OSI SAF (2019): ASCAT Coastal Winds at 12.5 km Swath Grid - Metop, EUMETSAT SAF on Ocean and Sea Ice, Collection ID: EO:EUM:DAT:METOP:OSI-104, https://navigator.eumetsat.int/product/EO:EUM:DAT:METOP:OSI-104 .

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

Porter, Claire, et al., 2023, “ArcticDEM, Version 4.1”, https://doi.org/10.7910/DVN/3VDC4W, Harvard Dataverse, V1, 2024-02-05

Randriamampianina, R.; Iversen, T.; Storto, A., 2011: Exploring the assimilation of IASI radiances in forecasting polar lows. Q. J. R. Meteor. Soc., doi:10.1002/qj.838, 137, 1700–1715.

Randriamampianina, R., Aspenes, T., Mile, M. and Schyberg, H., 2017: Impact of atmospheric motion vectors (AMV) on raid update cycling (RUC) and rapid refresh (RR) systems. MET report 04/2017, ISSN 2387-4201. Available from Norwegian Meteorological Institute at https://www.met.no/publikasjoner/met-report/met-report-2017.

RGI 7.0 Consortium, 2023. Randolph Glacier Inventory - A Dataset of Global Glacier Outlines, Version 7.0. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi:10.5067/f6jmovy5navz. Online access: https://doi.org/10.5067/f6jmovy5navz

Solberg, Rune; Rudjord, Øystein; Salberg, Arnt Børre; Killie, Mari Anne; Eastwood, Steinar; Breivik, Lars-Anders, 2017: Advancement of global snow mapping in CryoClim, Sentinel4CryoClim Phase 1, Deliverables 1-6. Norsk Regnesentral, Oslo. NR-notat SAMBA/10/17.

Tonboe, R. T., Eastwood, S., Lavergne, T., Sørensen, A. M., Rathmann, N., Dybkjær, G., Pedersen, L. T., Høyer, J. L. and Kern, S., 2016: The EUMETSAT sea ice concentration climate data record, The Cryosphere, 10(5), 2275–2290, doi:10.5194/tc-10-2275-2016.

Toohey et al., 2016: Easy Volcanic Aerosol (EVA v1.0): An idealized forcing generator for climate simulations. Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-8.

Toptunova, O., Choulga, M., and Kurzeneva, E.: Status and progress in global lake database developments, Adv. Sci. Res., 16, 57–61, https://doi.org/10.5194/asr-16-57-2019, 2019.

Toudal Pedersen, L., Dybkjær, G., Eastwood, S., Heygster, G., Ivanova, N., Kern, S., Lavergne, T., Saldo, R., Sandven, S., Sørensen, A. and Tonboe, R., 2017: ESA Sea Ice Climate Change Initiative (Sea_Ice_cci): Sea Ice Concentration climate data record from the AMSR-E and AMSR-2 instruments at 25km grid spacing, version 2.0, , doi:10.5285/c61bfe88-873b-44d8-9b0e-6a0ee884ad95.

Valkonen, T., H. Schyberg, and J. Figa-Saldana, 2016: Assimilating Advanced Scatterometer winds in a high-resolution limited area model over northern Europe. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., PP (99), 1–12, doi:10.1109/JSTARS.2016.2602889.

Yang, X., et al., 2020: C3S Arctic regional reanalysis - Full System documentation. Available at https://confluence.ecmwf.int/display/CKB/Copernicus+Arctic+Regional+Reanalysis+%28CARRA%29%3A+Full+system+documentation.

Appendix 1 - summary list of CARRA2 input data sets

Type of input data	Data source	Time period covered	Comments	Relevant reference(s) for data set
Conventional observations: SYNOP, SHIP, DRIBU, TEMP, AIRCRAFT	MARS archive. ERA40 BUFR 1982-2002. After that the operational archive	1982 - present		Hersbach et al, 2020 Haimberger et al, 2012
Local surface observations from non-MARS/GTS sources*	National Weather Services quality controlled archives, three Greenland measurement programmes	1982 - present	Data collected by the CARRA team, not used in global reanalysis, with addons relative to CARRA1 (see above)	Homleid et al, 2018
Microwave radiances: MSU, AMSU-A, AMSU-B, MHS, ATMS, MWHS2	MARS archive, same source as ERA5	1982 – present. For individual instruments: MSU: 1982-2006. AMSU-A: 1998 - present. AMSU-B: 2000-2009. MHS: 2005 - present ATMS: 2012 - present MHWS2: 2016 - present	Some addons to CARRA1	Lindskog et al, 2012
Infrared radiances: IASI, CrIS	MARS archive, same source as ERA5	2007 – present. For individual instruments: IASI: 2007 – present CrIS: 2012- present		Randriamampianina et al, 2011

Atmospheric Motion Vectors (AMV)	MARS and ECFS archive, same as used in ERA5.	1982-present	Products from several polar orbiting satellites.	Randriamampianina et al, 2017
Scatterometer	EUMETSAT OSI SAF	1993-present	Several satellites	Valkonen et al, 2016, OSI SAF, 2016 OSI SAF, 2019
Radio occultation	EUMETSAT ROM SAF delivered directly from DMI	2006 - present	Several satellites	Gleisner et al, 2020
Satellite snow observations*	CryoClim based, processed at MET Norway	1982-present	AVHRR satellites, using input data records from EUMETSAT CM SAF	Homleid and Killie, 2013 Solberg et al, 2017 Killie et al, 2018
Physiographic data bases: As the various physiographic data bases have inaccuracies in the Arctic, they are all inspected and manually corrected.				
Orography*	Copernicus GLO-90 DEM with few corrections using ArcticDEM	Fixed	GLO-90 is a digital elevation model (DEM) read in at 150 m resolution (GMTED2010)	GLO-90, 2024, Porter, Claire, et al., 2023
Land cover, height of trees, leaf-area index (LAI), albedo over land, both visible and near infra-red (NIR).	ECOCLIMAP-SG from Météo-France with updates	Most parameters are fixed, but LAI and albedo are in 36 10-day periods	The latest version of ECOCLIMAP-SG, produced at 300m-resolution, with updates to glacier extents (RGI), lakes and vegetation (Greenland, Svalbard, Faroe Islands and Iceland), tree height in Iceland	ECOCLIMAP-SG (https://opensource.umr-cnrm.fr/projects/ecoclimap-sg) RGI 7.0 Consortium, 2023 Mojtaba et al., 2018 Johansen et al., 2009 Jacek, 2022 Iceland tree height based on data from Icelandic National Forestry, their research division at Mógilsá
Lakes	GLDB	Fixed	300 m global depth map (GLDB)	Toptunova et al, 2019
Soil*	SoilGrids from ISRIC – World Soil	Fixed	Global database (SoilGrids) of sand	Poggio et al, 2021

	Information		and clay fractions at 250 m resolution and soil organic carbon (SOC)	
Glacier albedo				
Glacier surface albedo*	Geological Survey of Denmark and Greenland (GEUS)	1982-present	Pre-1997, the Extended AVHRR Polar Pathfinder (APP-x) product is used. For the period 1997 to 2020, MODIS-MOD10A1 glacier albedo data is used. Sentinel-3 is used from 2021.	Box and Steffen, 2017 Kokhanovsky et al. 2020
SST and sea ice: SST and sea ice inputs are post-processed for consistency at DMI and MET Norway				
SST*	CMEMS and C3S	1982-present	Based on ESA CCI SST (up to 2021) and Baltic maps. After 2021, C3S iCDR v3 replaces the EXA CCI product.	Merchant et al, 2016 Høyer, 2016 Høyer and Karigali, 2016
Sea ice*	ESA CCI+, CMEMS Baltic maps, EUMETSAT OSISAF	1982-present		Tonboe et al, 2016 Toudal Pedersen et al, 2017 Lavergne et al, 2019
Time varying radiative forcing				
Aerosols	Data provided in Toohey et al, 2016.	1991-1995	Estimate for the Pinatubo eruption.	Toohey et al., 2016.
Greenhouse gases	CMIP6 historical greenhouse gas data and scenarios.	The full CARRA2 period (1986-2025)	The historical greenhouse gas data are used 1850-2014, scenario data from 2015.	IPCC AR6, 2021.

* Indicates input data types which constitute a significant addition, extension or increased resolution relative to what is used in the global reanalysis ERA5.

Appendix 2 - overview of changes in surface input data sets

Surface and type of data sets	Time for change in input data type	Notes on impact of data change
Satellite snow cover for surface assimilation: Infrared imager on polar orbiting satellites AVHRR.	2019	Snow cover based on infrared imagery is only used where and when such satellite products are available (dependent on sun angle and cloud cover). These observations are then assimilated together with conventional snow observations in the surface assimilation scheme, and the blending implicit in the assimilation procedure reduces the impact of any change in input data resolution. The CryoClim optical product with the same resolution is now available for 1982 - June 2019, based on Fundamental Climate Data Record (FCDR) for AVHRR GAC brightness temperatures and radiances from EUMETSAT CM SAF ("Clara A-2" and "Clara A-2.1"). Post June 2019 EUMETSAT delivers Interim Climate Data Records (ICDR) for AVHRR GAC data, meaning that we set up processing to produce a continuation of the CryoClim optical data set in a way believed to be quite consistent beyond June 2019.
Sea ice concentration: Passive microwave imagers SSM/I, SSMIS and AMSR.	1991	For sea ice concentration (SIC) we will use two off-the-shelf data sets based on microwave polar orbiting imagers, first the EUMETSAT OSI SAF Sea Ice Concentration v3 data record (1978-), and then the ESA CCI+ Sea Ice Concentration v3 data record (1991-) which is based on AMSR with higher spatial resolution. The OSI SAF dataset is consistent over the 1978 - present day. For the post 1991 period, only the CCI+ data record will be used. Regular processing will be performed for the CCI+ record to keep it up to date towards real-time. In CARRA1 we developed a post-processing and filtering procedure of the CCI+ data set to make it consistent with the OSI SAF products. Manual check on the relevant time series on SST and Ice fractions during the 1991 transition period as well as sensitivity experiments were done to evaluate eventual discontinuities. No significantly abnormal features or jumps were detected in those examinations. (This was documented in the CARRA1 evaluation and validation report.) We have updated the filters developed in CARRA1 to generate consistent SST and SIC fields from the different satellite systems for the extension to the full pan-Arctic domain. This will minimize impact due to sensor resolution changes, such as those in 1991 (AMSR phase-in) and 2011-2012 (AMSR phase-out, AMSR2 phase-in).

Sea surface temperature:	None	A new gap-free Arctic SST and Ice surface temperature (IST) has been generated within the Copernicus Marine Service (CMEMS) for the period between 1982 and 2021. The dataset uses the SST products from the ESA CCI SST and the C3S service, and has been validated against drifting buoys, ships and moored buoy observations, and is believed to have a good consistency in time.
Glacier surface albedo: Infrared imagers on polar orbiting satellites AVHRR, MODIS and Sentinel-3.	1997	Pre 1997, the Extended AVHRR Polar Pathfinder (APP-x) product will be used. For the period 1997 to 2017, MODIS-MOD10A1 glacier albedo data will be used for the regions with glaciers. Both AVHRR and MODIS satellites have quite a bit higher resolution than the model mesh width, so the resolution change is not believed to be a main issue. We will do a check for biases between these two data sets for an overlapping period.
	2020	The MODIS satellites are approaching the end of their lifetime, and Sentinel-3 will be used from 2017 in CARRA2. In CARRA1 the period between 2017 and 2020 was covered with MODIS data, and this will allow us to do a comparison of the two data sets in this period in CARRA2 and consider steps to ensure consistency. The new data set, the “SICE" Sentinel-3 albedo data set, will have increased accuracy as compared to the MODIS MOD10A1 data (Kokhanovsky et al. 2020), the latter suffer some bias for low sun angles. The change of the albedo data stream in this connection was also experienced in the CARRA1 reanalysis in 2020 and was assessed then: The overall impact was estimated to be local and insignificant.

ECMWF - Shinfield Park, Reading RG2 9AX,
UK

Contact: https://support.ecmwf.int/