Danmarks Meteorologiske Institut

DMI Report 22-13
Danish Regional Reanalysis

Final scientific report of the 2021 National Center for Climate Research Work Package 1.1.2 Danish Regional Reanalysis

Xiaohua Yang, Bjarne Amstrup, Carlos Peralta Aros, Kasper Hintz

Danmarks Meteorologiske Institut

Kolofon

Serietitel
DMI Report 22-13

Titel
Danish Regional Reanalysis

Undertitel
Final scientific report of the 2021 National Centre for Climate Research Work Package 1.1.2, DANRA

Forfatter(e)

Xiaohua Yang, Bjarne Amstrup, Carlos Peralta Aros, Kasper Hintz

Ansvarlig institution
Danmarks Meteorologiske Institut

Sprog
English

Emneord
High-resolution atmospheric reanalysis, Denmark, Harmonie-arome

Url
https://www.dmi.dk/publikationer/

ISSN
2445-9127

ISBN
978-87-7478-715-0

Versionsdato
15. januar 2022

Link til hjemmeside
www.dmi.dk

Danmarks Meteorologiske Institut

Contents

1. Summary

Since late 2020 a regional reanalysis for Denmark, the DANish ReAnalysis (DANRA), has been conducted at the National Center for Climate Research (NCKF), with a target to produce a high resolution atmospheric reanalysis for 70 years during 1950 and 2020, using the HARMONIE-arome numerical weather prediction system. Up to the present, a total of 12 years worth of reanalysis data has been generated, excluding preparation assimilation needed for system spin-up.

With a continuous, high-resolution gridded data coverage in time and space, DANRA portrays the Danish climate and its evolution in recent periods with an unprecedented high fidelity. DANRA also produces near real time updates of the atmospheric states to monitor the latest climate trends in Denmark. From continuous verification against observation data, the high resolution DANRA is seen to be clearly superior to other reanalysis products for the Danish area, since the latter are produced by model systems with coarser resolution or without data assimilation. In particular, DANRA is found to be more skillful in representing extreme weather such as winter storms and summer cloud bursts, making it a superior data-set for use in regional climate studies over the Danish area as well as other applications in the Green transition.

Presently, the DANRA team is working on the delivery of the first 30 years of the reanalysis, 1990-2020. For computation efficiency, the reanalysis production is split on 9 time slices distributed over the analysis period to allow parallel execution. Assuming a similar level of staff resources, with the present pace, it is estimated that the full 70 year DANRA reanalysis can be finished by 2024, with the first 30 years released within one year.

Work has been on going to establish the necessary software infrastructure, danra-pp, through the automatization of post-processing operations via containerisation to generate relevant statistical products to extract climate information from the dataset, and to ready the DANRA dataset for use by external users in different applications.

2. DANRA during 2021

Since late 2020 a regional reanalysis for Denmark, the DANish ReAnalysis (DANRA), has been in production at the National Center for Climate Research (NCKF), with a target to release a high resolution atmospheric reanalysis for 70 years from 1950 until 2020, using the HARMONIE-AROME numerical weather prediction system.

Making atmospheric reanalysis entails running an assimilation forecast using a fixed version of a state-of-the-art NWP model for a sufficiently long period (~30 years or longer) which is of relevance from climate perspectives. Observation data, which ideally has been collected and quality assured, add reliability to the quality of the dataset. The resulting dataset, in the form of gridded atmospheric states, which are continuous in space and time, provides a basis for climate statistics that forms a knowledge basis about climate states in the recent past. A re-analysis project involves as a minimum the preparation of input data (models, orographic and physiographic database and observations), configuration and running of NWP models with data assimilation and short-range forecast, monitoring and diagnosis of production data, post-processing of essential climate variables, dissemination and outreaching activities including data dissemination, user guidance and support.

During 2020, a successful pilot phase has been conducted, resulting in the establishment of the DANRA system and a test production of 6 years worth of re-analysis. In the scientific report by Yang et al 2021, the DANRA pilot study is documented, demonstrating a satisfactory performance with the provisional reanalysis data-set. During 2021, quality assurance work for the recent 30 years of observation data in the Danish area was initiated, enabling DANRA to enter the production phase with significantly enhanced observation data. Up to the present, a total of 12 years worth of reanalysis data has been generated, excluding preparation assimilation needed for system spin-up.

With a continuous, high resolution gridded data coverage in time and space, DANRA portrays the Danish climate states and its evolution in recent periods with an unprecedented high fidelity. Apart from reanalysis for historical periods, DANRA also started producing near real time updates of the atmospheric states to facilitate monitoring of the latest climate trends in Denmark.

From continuous verification against observation data, the high resolution DANRA is seen to be clearly superior to other reanalysis products for Danish area, which are produced by model systems either with coarser resolution or without data assimilation capability. DANRA is found to be competitive in representing extreme weather events such as winter storms and summer cloud bursts, making it an superior data-set for use in regional climate studies over the Danish area as well as other applications in the Green transition, in particular wind and solar energy/power.

Quality assurance of observation data from Denmark

In routine weather forecasting, the backbone in-situ observation data for the near-surface atmosphere are those from the WMO-sponsored international meteorological observation

network Global Telecommunication System (GTS). In most reanalysis systems, including those used by the Copernicus flagship reanalysis ERA5, the observation archive from GTS data is the only source of surface data over Denmark that is available for data assimilation. Due to historical reasons, the station density of the GTS surface network (SYNOP) has been quite thin compared to the actual data density as collected by national weather services, although the data exchange has improved greatly over the past decades. For DANRA, it has been necessary to put substantial effort to collect additional observation data in order to enhance the fidelity of the reanalysis.

The use of observation data, especially data collected from historical periods, including those additional, (hereafter referred to as “non-GTS”) data, involves a significant amount of work on quality assurance. For the Danish region, the non-GTS data can be extracted from DMI archives in bufr format, and from the DMI observational database (dmiDB). These also contain 3rd party data. Some of these stations started to deliver 10-min data around 2000. This contrasts with the situation with the GTS data as archived at the ECMWF, which has data at most every hour.

An extensive quality control has been applied to the historical data including both GTS and non-GTS data. In particular, quality for some of the data from the 1990s tends to be quite poor and needs extensive preprocessing before injection to data assimilation. During the quality control step, manual inspection of monthly plots of pressure, temperature, relative humidity and wind data for each station is made. As reference, corresponding data from models (ERA-5 or the operational DMI model at the time) are compared. This has been shown particularly helpful to detect anomalies with pressure data, since the mean sea level pressure (mslp) from the short forecast model is normally consistent in time and of good quality. With model data as reference, it becomes easy to detect anomalies with some stations data in mslp due to bad barometer height or bad calibration. Corrections can be made accordingly to improve data quality to make data usable, or reject it. Additionally, in a few cases mslp measurements are missing for a longer period whereas surface pressure records exist. For such situations, using 2 meter temperature, the mslp can be rescued using a revised barometer height and the DMI approach to derive mslp from surface pressure. The approach has been applied to a few stations in the 1990s and also a few 3rd party stations in recent years. Fig 1 illustrates an example using the 3rd party station 06017.

Prior to 2000, there had been cases with poor mslp due to the absence of 2 meter temperature. These data are discarded. Also for the pre-2000 period, sometimes the same data (pressure, relative humidity, temperature and wind) appear in 2 consecutive observation times 3 hours apart and in some cases 1 hour apart. These measurements are most likely wrong. In many cases, this is found to occur to 1 or 2 of the 3 datasets, but not all 3. In such cases, data from the ‘normal behaving streams’ can be selected. In some cases, the bufr stream may contain 2 sets of observations for the same time point. Most often, it is obvious which one to pick. It also happens that the data for one or all parameters are the same for a longer period and those data are discarded as well. Another type of errors where data is discarded are cases where a sudden jump occurs. That is typically for 2 meters temperature and pressure data. Care is taken for sudden jumps for 2 meters temperature, which are not weather related.

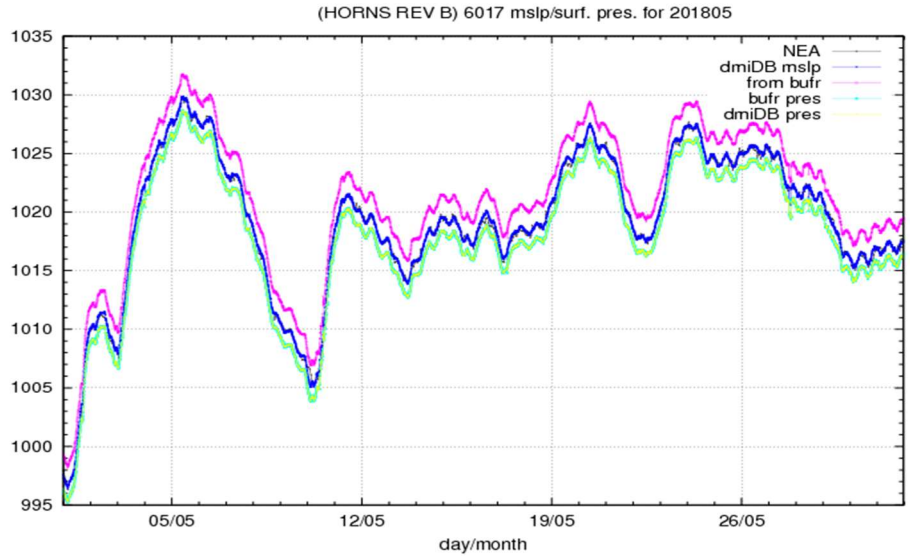

Figure 1. The pressure time series for mslp at station 06077 during May 2018 illustrating the impact due to bad barometer height. The data "from bufr" are original mslp data extracted from bufr (and dmiDB). The blue line is made using surface pressure and 2 meters temperature combined with a revised barometer height. The standard DMI approach has been applied to reduce surface pressure to mslp. It fits the HARMONIE-NEA mslp data very well.

Figure 2 demonstrates the validation intercomparison between reanalysis dataset for 2 m temperature during Jan 2011, for which DANRA has been run with and without addition of the above-mentioned quality assured non-GTS data. In addition, the coarser resolution global reanalysis ERA5 is also compared. The comparison shows a clear benefit of adding good quality observation data. In addition, the added value of the high resolution DANRA data over that of the global ERA5 dataset is obvious.

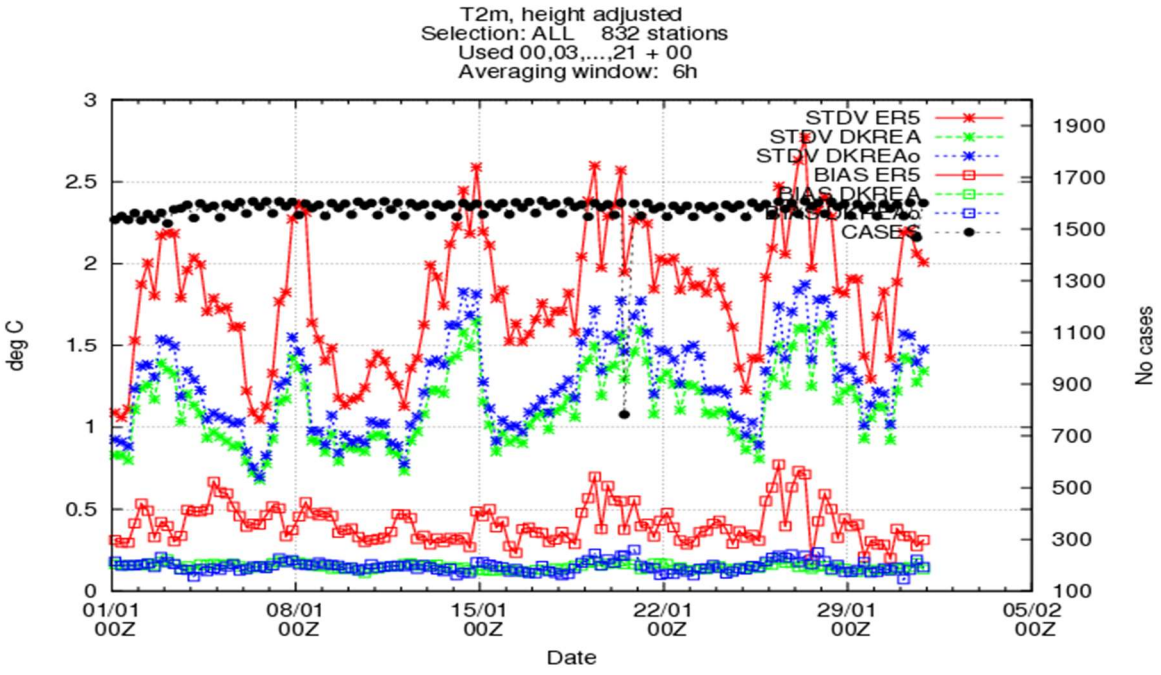

Figure 2. Time series of averaged departure between reanalysis data and observed screen level temperature (2 m) for the model area, Jan 2011. Upper curves for standard deviation in K and lower curves for bias. The red color denotes those by global reanalysis ERA5, the green color by DANRA, the blue color the same as for Green but without non-GTS observation. Note that when interpolating gridded model temperature to observed points, a standard lapse rate correction of -6.4 degree/1000 m has been applied to take into account the difference between model orography and measurement station heights at orographic area.

Reanalysis production and continuous monitoring

Production of DANRA is carried out at the High Performance Computation Facility (HPCF) at the ECMWF using the computation resource allocated to member states. In order to speed up the production, a total of 9 time slices, each with approximately 5 years, have been set up for parallel execution. Following the conventional practice in reanalysis such as Copernicus global reanalysis ERA5 and the Arctic regional reanalysis (CARRA), for each individual time slice, the DANRA reanalysis is initiated one year in advance of the targeted start, with the first year of provisional reanalysis conducted with the sole purpose to ensure an adequate 'warm-up' for deep soil model states. Up to the end of 2021, a total of 21 years of reanalysis have been generated, taking off a total of 9 year for the warm-up period, a 12 year worth of DANRA reanalysis has been produced.

The DANRA production is monitored in near real time via the website hosted on hirlam.org on https://hirlam.org/portal/oprint/WebgraF/DKREA/ mainly through the WebGraf interface. The monitoring site updates daily the production status including weekly and monthly production throughput and expected completion date. Fig 3 below shows a snapshot of the Gantt chart indicating progress of the DANRA reanalysis for each time slice.

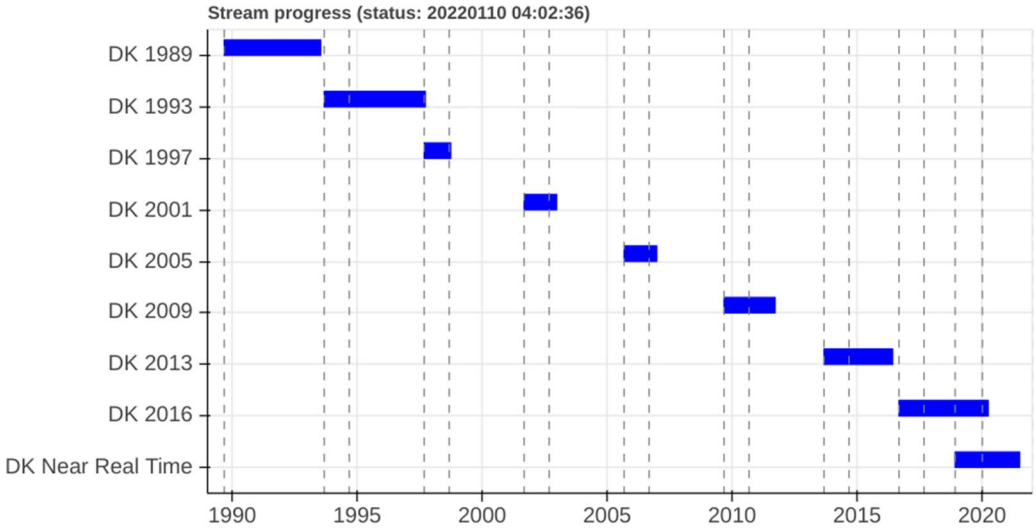

Figure 3. Gantt chart showing the overall DANRA production progress for the whole re-analysis period. The plot is interactive and shows the start date, current date for each production stream and weekly speed on each bar.

DANRA produces hourly essential climate variables, which enables the reproduction of continuous weather states for Denmark covering the entire reanalysis period with hourly temporal resolution. During production monitoring, weather charts for temperature, wind, pressure and precipitation every 3h are generated along production for each production stream. They can be visualized in the DANRA production interface https://hirlam.org/portal/oprint/WebgraF/DKREA/Prod/index.html?choice_ind=WeatherMap.

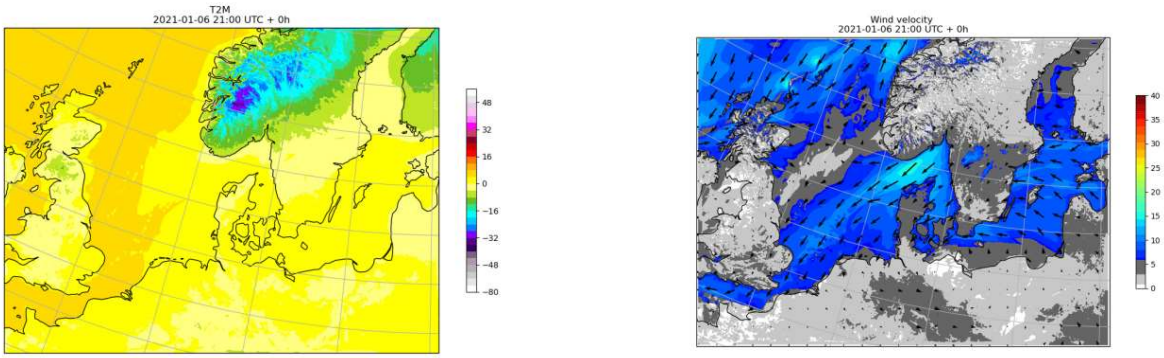

Figure 4. Weather charts of temperature and wind speed from DANRA. These charts, also including 3h accumulated precipitation, are produced along reanalysis production and presented at the CARRA webpage.

Danra output streams and data storage

The raw output from the DANRA production runs in ECMWF and it is post-processed daily on the ECMWF HPCF (Cray cca). Selected fields at specific levels (surface or pressure) are extracted and saved in files in GRIB1 format. The files are split in short/long forecasts (3/18h) and analysis. A complete list of the available parameters can be found on https://dmidk.github.io/danra-docs/fields.html. The files are packed in compressed tar balls and transferred to the DMI file archive isilon. The transfer is done daily using the ecaccess gateway at DMI.

Development of value-added post processing module danra-pp

The computation of various climate statistics including those of monthly and annual means of Essential Climate Variables (ECV) such as temperature, wind and rainfall can be conducted, using the available DANRA time series. Development of a post-processing module is ongoing. Until recently (December 2021) the main focus has been on creating infrastructure and dataflow for the post-processing of DANRA data. Since DANRA data is being produced continuously, there is also a need for the post-processing module to work continuously in parallel as new data arrives. Therefore, a file monitor system is needed. To solve this, a database has been created with an API endpoint that stores all information about which DANRA data has been processed (Fig 4). The post-processing module then queries that database for information when new data arrives and updates it respectively.

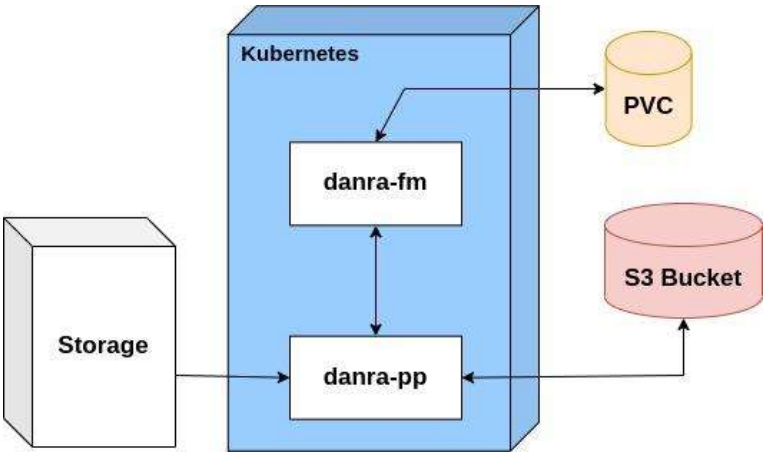

Figure 5. Overall structure of the post-processing module, danra-pp, and the connection with general infrastructure. danra-fm is a database with an API endpoint that stores information of what has already been processed. danra-fm is connected to a persistent volume claim (PVC) to avoid data losses in case of the container stopping.

Figure 5 shows how the post-processing module, danra-pp, is connected to the general infrastructure. As described, raw compressed data from DANRA is archived at DMI storage. danra-pp is a container which reads data from storage and at the same time, queries the danra-fm container, to get information about the file it has found. If the requested parameter has not been computed before danra-pp continues and updates the temporary results stored in a S3 bucket.

It has been decided to complete the module first for a single parameter and a single type of statistics. To begin with, monthly means of 2-meter temperature are being produced. However, more parameters and types of statistics are easily added hereafter. When a dataset is complete, it will be moved to a public storage instance in NetCDF format for download accessibility.

Outlook for 2022

Porting of the DANRA reanalysis system will be done during 2022 to move the reanalysis production to the new HPCF at the new ECMWF Computer center in Bologna, Italy. With the present pace, it is technically feasible to complete within one year the first full 30 year reanalysis between 1991 and 2020, provided that resource input is kept at the similar level as in 2021. Another crucial factor in pace-setting DANRA production is the work to quality assure observation data for the remaining time slices of the 30 year period, which sums up to 18 years. It is envisaged that, in cooperation with the Niels Bohr Institute at the Copenhagen University, a master student can investigate the application of machine learning in such work, with the ultimate goal to speed up the preparation and cleaning of the observational data. Availability of a continuous DANRA data stream also enables its use by downstream applications such as those in the energy sectors. If feasible, outreaching activities will be initiated to promote the use of the DANRA data.

Reference

Yang, X, K. S. Hintz, C. P. Aros and B. Amstrup, 2021, Danish Regional Atmospheric Reanalysis: Final scientific report of the 2020 NCKF Work Package 3.2.1, Regional Reanalysis Pilot. DMI report 21-31, 2021.