RESEARCH ARTICLE

Value of the Copernicus Arctic Regional Reanalysis (CARRA) in representing near-surface temperature and wind speed in the north-east European Arctic

Morten Køltzow,[1] Harald Schyberg,[1] Eivind Støylen[1] & Xiaohua Yang[2]

[1]Norwegian Meteorological Institute, Oslo, Norway; [2]Denmark Meteorological Institute, Copenhagen, Denmark

Abstract

The representation of 2-m air temperature and 10-m wind speed in the high-resolution (with a 2.5-km grid spacing) Copernicus Arctic Regional Reanalysis (CARRA) and the coarser resolution (ca. 31-km grid spacing) global European Center for Medium-range Weather Forecasts fifth-generation reanalysis (ERA5) for Svalbard, northern Norway, Sweden and Finland is evaluated against observations. The largest differences between the two reanalyses are found in regions with complex terrain and coastlines, and over the sea ice for temperature in winter. In most aspects, CARRA outperforms ERA5 in its agreement with the observations, but the value added by CARRA varies with region and season. Furthermore, the added value by CARRA is seen for both parameters but is more pronounced for temperature than wind speed. CARRA is in better agreement with observations in terms of general evaluation metrics like bias and standard deviation of the errors, is more similar to the observed spatial and temporal variability and better captures local extremes. A better representation of high-impact weather like polar lows, vessel icing and warm spells during winter is also demonstrated. Finally, it is shown that a substantial part of the difference between reanalyses and observations is due to representativeness issues, that is, sub-grid variability, which cannot be represented in gridded data. This representativeness error is larger in ERA5 than in CARRA, but the fraction of the total error is estimated to be similar in the two analyses for temperature but larger in ERA5 for wind speed.

Keywords

High-resolution; weather; climate; polar low; northern Fennoscandia; Svalbard

Correspondence

Morten Køltzow, Norwegian Meteorological Institute, Henrik Mohns Plass 1, NO-0371 Oslo, Norway. E-mail: famo@met.no

Abbreviations

CARRA: C3S Arctic Regional Reanalysis
C3S: Copernicus Climate Change Service
CoNo: coast and fjords of Norway
ECMWF: European Center for Medium-range Weather Forecasts
ERA5: ECMWF Reanalysis, fifth generation
ETS: equitable threat score
GoBothnia: Gulf of Bothnia coast
GTS: Global Telecommunication System
MAE: mean absolute error
MSLP: mean sea-level pressure
NWP: numerical weather prediction
NSF inland: inland and valleys in Norway, Sweden and Finland
T2m: 2-m air temperature
WS10: 10-m wind speed
SD: standard deviation
SDE: standard deviation of error
SYNOP: synoptical observations

Introduction

Reanalyses combine a comprehensive assimilation of historical observations with state-of-the-art descriptions of relevant physical processes in a frozen version of NWP models. By this, they produce a best estimate of historical weather and climate at arbitrary locations, which is consistent in space, time and between parameters, and is therefore suitable for climate monitoring, climate and meteorological research, and as input to various other types of products, models and research.

This study investigates the value added by the newly released high-resolution regional CARRA reanalysis (Yang et al. 2020) compared to the extensively used global C3S ERA5 reanalysis (Hersbach et al. 2020), which is used as a host model on the lateral boundaries of the CARRA reanalysis. We focus on near-surface temperature and wind speed over the north-east European Arctic. The parameters are chosen because of their importance for climate and high-impact weather, and because near-surface temperatures are constrained through the assimilation process in both reanalyses, while wind speeds are not assimilated and can, therefore, be treated as independent observations. To limit the amount of details and discussion, the study considers the eastern of the two CARRA domains: CARRA-East, covering Svalbard, Barents Sea, northern Norway, Sweden and Finland (Fig. 1). The results are, however, expected to be qualitatively similar in the western domain of CARRA: CARRA-West, which includes Greenland and Iceland.

Citation: *Polar Research* 2022, *41*, 8002, http://dx.doi.org/10.33265/polar.v41.8002

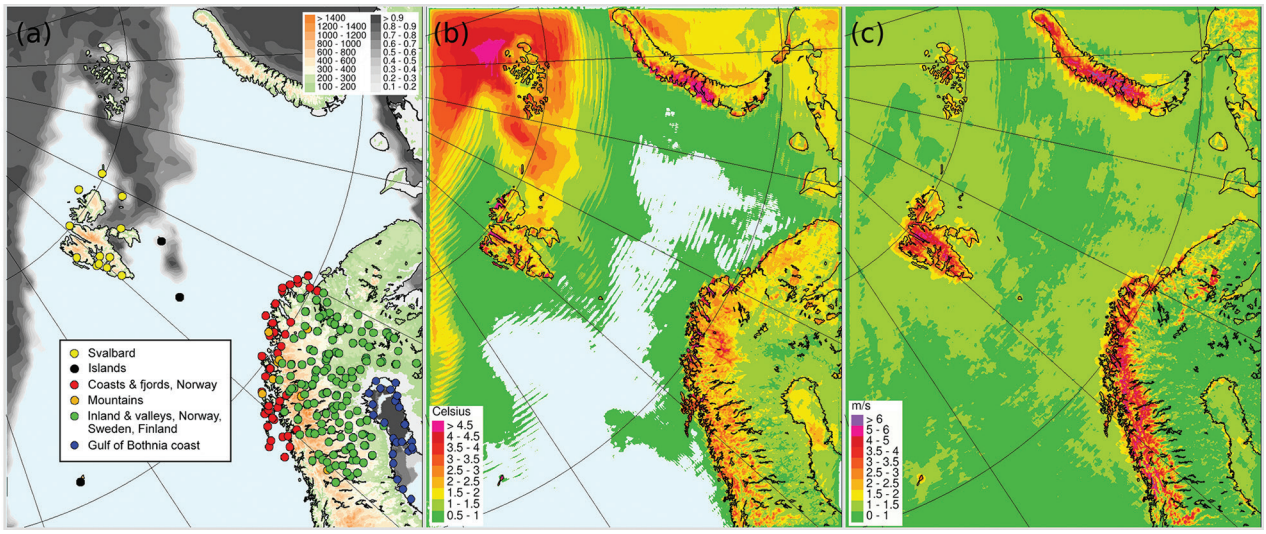

Fig. 1 (a) Region of comparison, indicating the observation sites applied for verification, with topography from the CARRA data set and sea ice from 1 March 2018. In addition, mean absolute difference between CARRA and ERA5 averaged over the Year of the Polar Prediction Special Observing Period 1 (February and March 2018) for (b) T2m and (c) WS10.

The fifth generation of atmospheric reanalysis produced by ECMWF, ERA5 benefits from a substantial development in model physics, core dynamics and data assimilation from the operational NWP system at ECMWF, in addition to enhanced horizontal and temporal resolution compared to its predecessors. ERA5 is, therefore, one of the leading global reanalyses. Many studies have evaluated ERA5 in general (e.g., references in Hersbach et al. 2020) and also its performance for the Arctic and high-latitudes (e.g., Batrak & Müller 2019; Betts et al. 2019; Graham et al. 2019; Tetzner et al. 2019; Wang et al. 2019; Delhasse et al. 2020; Sheridan et al. 2020). In general, ERA5 performs well compared to other reanalyses but pronounced weaknesses remain in the representation of the Arctic climate system. For example, there is a warm bias over sea ice (e.g., Batrak & Müller 2019; Wang et al. 2019; Demchev et al. 2020; Graham et al. 2019), and certain characteristics of temperature extremes are not well represented (Sheridan et al. 2020). The same is true for wind speed associated with polar lows (Moreno-Ibanez et al. 2021).

Global reanalyses require enormous computer capacity, which puts limitations on the affordable spatial resolution in the underlying NWP model. Even if the spatial resolution of global reanalyses has improved, ERA5 still employs a relatively coarse grid spacing of ca. 31 km. The resolution is, therefore, too coarse to describe many small- and meso-scale details, in particular, the complex topography and coastlines, as found at Greenland, Svalbard and Norway, among other places. The relatively coarse resolution of global reanalyses, therefore, provides good arguments for regional reanalyses. Recently, the added value of regional reanalyses has been documented in mid-latitudes (e.g., Kaiser-Weiss et al. 2019; Kaspar et al. 2020; Keller & Wahl 2021) and in the Arctic (e.g., Bromwich et al. 2018). Furthermore, the approach with regional high-resolution model systems has also proven highly useful in daily weather forecasting of near-surface variables in the European Arctic (e.g., Müller et al. 2017; Yang et al. 2018; Køltzow et al. 2019). It is, therefore, expected that the recently released high-resolution regional CARRA data set will add value to ERA5 because: (1) the horizontal resolution (2.5-km vs. 31-km grid spacing in CARRA and ERA5, respectively) has been substantially improved; (2) more local observation data have been collected and used in the assimilation process in CARRA; and (3) the treatment of cold surfaces in CARRA has improved (Yang et al. 2020).

Here we present a detailed evaluation of the representation of T2m and WS10 in CARRA and ERA5. First, the CARRA and ERA5 data sets, and point observations used for verification are briefly described. Then, we present the results of the evaluation and we compare CARRA and ERA5.

Methods and data

In this study, reanalyses are evaluated by comparison with point observations, and the added value of CARRA data in comparison to ERA5 is based on the relative reduction in errors of T2m and WS10.

Citation: *Polar Research* 2022, *41*, 8002, http://dx.doi.org/10.33265/polar.v41.8002

ERA5

ERA5 is based on the Integrated Forecast System cycle 41r2, which was used to produce the ECMWF's operational forecasts in 2016. Hersbach et al. (2020) described the general set-up and provided a summary of improvements in the model system compared to the ECMWF's previous reanalysis (ERA-interim). These improvements benefit from a decade of developments in the ECMWF's operational model system. The atmospheric initialization is done by a four-dimensional variational data assimilation approach, while the initialization of the surface is done by optimal interpolation for snow and soil temperature and a simplified Extended Kalman Filter for soil moisture. A number of remote sensing measurements are assimilated together with conventional observations distributed by GTS. ERA5 provides hourly output with 31-km grid spacing and 137 vertical levels from 1950 until today and can be downloaded from the online Copernicus Climate Data Store (Copernicus Climate Change Service 2017).

CARRA

CARRA is based on the HARMONIE-Arome cycle 40h1.1. A similar version was used for operational forecasting at the Danish Meteorological Institute and the Norwegian Meteorological Institute until the beginning of 2021. The production of CARRA-West (Greenland/Iceland) and CARRA-East (Fig. 1), therefore, benefits from the work done for operational weather forecasting in these regions. In addition, further tailoring of the description and specification of cold surfaces has been done in a number of ways in the preparation for the production of CARRA (Yang et al. 2020). The atmospheric assimilation is done by a three-dimensional variational data assimilation approach, while the surface assimilation (snow, soil moisture and temperature) is done by optimal interpolation. A number of remote sensing measurements, conventional observations distributed by GTS and additional surface observations collected from the national institutes are used. CARRA produces 3-hourly analyses with a 2.5-km grid spacing and 65 vertical levels. CARRA is forced with ERA5 analyses at its lateral boundaries. CARRA can be downloaded from the Copernicus Climate Data Store (https://climate.copernicus.eu/climate-data-store).

Observations

In this study, CARRA and ERA5 are evaluated with SYNOP observations for the period 1998–2020. Eventually CARRA will cover a period from September 1990 to September 2025. Although the production of the full data set was not completed at the time of the analyses, our results represent a substantial part of the total data set to become available. The majority of observations used here were distributed via GTS in near real-time and have, therefore, been available for the surface assimilation in both reanalyses. However, a substantial part of the observations was also collected from local sources, for example, for periods before they were distributed via GTS, and therefore, they are available only in CARRA during production. Furthermore, the usage of the observations in the assimilation process in the reanalyses depends on system characteristics (e.g., assimilation choices, resolution, and land–sea mask) and varies between ERA5 and CARRA. Both CARRA and ERA5 apply observations of T2m in their assimilation schemes, while WS10 is not assimilated. In addition, the observations have undergone additional quality control after the distribution via GTS and in the preparation of CARRA (Yang et al. 2020). The observations used in this evaluation, and also in the production of CARRA, may, therefore, occasionally deviate from what was originally distributed via GTS.

It is well known that the quality of NWP systems varies geographically, and therefore, we stratify the evaluation into six regions in this study (Fig. 1a) for which we expect relatively homogeneous weather: Svalbard (12 observation sites); North Atlantic islands (three observation sites, hereafter islands); CoNo (41 observation sites); NSF inland (140 observation sites); mountains (10 observation sites); and coastal stations of GoBothnia (26 observation sites). Observation sites were subjectively assigned to regions, taking into account elevation and distance to the coast. Several of the same regions have previously proven to highlight important geographical differences in forecast quality (Køltzow et al. 2019).

Results and discussion

Different aspects of the value added by CARRA as compared to ERA5 are evaluated in this section. First, we compare CARRA and ERA5 for a limited time slice to identify typical differences. Then, a general evaluation against observations is performed, followed by an intercomparison of spatial variability and temporal variability. This is followed by an investigation of how well local extremes and some selected high-impact events are represented. Finally, we discuss how much of the difference between reanalyses and observations arises from representativeness errors, that is, sub-grid variability causing deviations from the grid box average represented in the reanalyses.

Potential added value

An impression of where CARRA has potential added value compared to ERA5 is illustrated by the mean absolute difference in Fig. 1b, c for T2m and WS10 for a winter period. The largest mean absolute difference for T2m, and thereby also the largest potential added value, is found over complex topography (Fig. 1a) and along coastlines, for example, in Svalbard, Novaya Semlya and CoNo and on mountains. A close examination shows that the differences are particularly large in some of the fjords in Svalbard and CoNo. In the more homogeneous terrain in NSF inland, the mean absolute difference is also evident but is less pronounced. In addition, the potential added value is high over the sea ice north of Svalbard (Fig. 1a; CARRA is systematically colder than ERA5, not shown), while the least potential added value is seen over the ocean.

Some of the same patterns seen for T2m are also seen for WS10, with large potential added value over complex terrain and along coastlines. However, over more homogeneous terrain in north Sweden and Finland and over sea ice and open sea, there are more modest differences between CARRA and ERA5. The spatial patterns in mean absolute difference appear similar during summer, albeit with smaller values (not shown), but with no pronounced potential added value for T2m over sea ice. The spatial patterns in mean absolute difference highlight the importance of resolution for both T2m and WS10. In addition, the differences in T2m over sea ice may be due to the more sophisticated representation of sea ice and snow on sea ice in CARRA. The inclusion of a snow layer on top of the sea ice is missing in many reanalyses but its presence (as done in CARRA) has shown to improve on a common warm bias over sea ice seen in many reanalyses (e.g., Batrak & Müller 2019).

General evaluation against observations

The annual cycle of T2m systematic errors (biases) for different regions is shown in Fig. 2. In CARRA, the temperature biases are small, but a modest exaggeration of the annual cycle is seen in the mountains (ca. 0.5°C cold/warm bias in winter/summer) and a ca. −0.5°C cold bias in summer for GoBothnia. It should be noted that even if the average biases in the regions are small, the biases may vary between observation sites within each region (not shown). The T2m biases in ERA5 are more pronounced than in CARRA for several of the regions. A substantial cold bias is seen for CoNo and Svalbard in winter (ca. −1.5°C) and slightly less in summer (ca. −1°C). These biases can at least partly be explained by the coarser resolution of ERA5, making an accurate representation of the complex coastline difficult. However, compared to the ECMWF's previous reanalysis, these specific biases are reduced by modifications to the radiation scheme (Hogan & Bozzo 2015; Hersbach et al. 2020). Coarser resolution is also a likely explanation for the underestimation of the annual cycle at islands (warm bias ca. 0.5°C in winter and cold bias ca. −0.5°C in summer), which are not properly resolved, because the land–sea mask shows 0–10% land for the

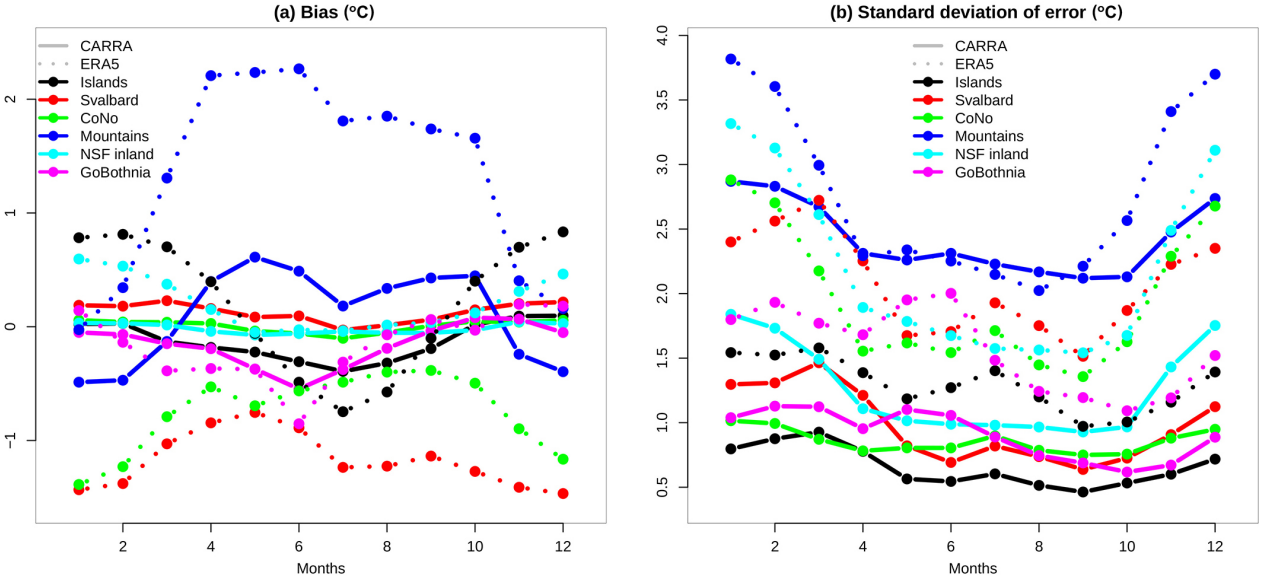

Fig. 2 (a) Annual cycle of T2m bias and (b) SDE for T2m for CARRA and ERA5 compared with observations for different regions: islands, Svalbard, CoNo, mountains, NSF inland and GoBothnia.

Citation: *Polar Research* 2022, *41*, 8002, http://dx.doi.org/10.33265/polar.v41.8002

three locations. This results in reduced variability and limitations in the use of surface observations in the data assimilation. Furthermore, ERA5 shows a pronounced warm bias (ca. 2°C) from April to October in the mountains. On average, the height difference between the applied observation sites and ERA5 model topography is 278 m in this region (because of the coarse resolution in ERA5), which with a simple height correction (0.65 °C/100 m) will reduce the summer bias by ca. −1.8°C. However, given the more complex vertical temperature profile in winter, no simple correction is possible at this time of the year. A similar height correction for CARRA, which better resolves the topography, gives a correction of ca. −0.7°C in the mountains. The impact of height differences when comparing the reanalyses with observations is further discussed later in this article. A moderate warm bias (ca. 0.5–1°C) during winter is in addition found at NSF inland for ERA5. This is an area for which episodic, very stable stratification of the atmosphere usually gives cold conditions and large errors in NWP systems (e.g., Atlaskin & Vihma 2012; Sandu et al. 2013). Recently, it has been shown that at least a part of this warm bias can be improved by implementing a multilayer snow scheme in the ECMWF's Integrated Forecast System (Arduini et al. 2019).

The annual cycle of the T2m unsystematic errors, measured by SDE, is shown in Fig. 2. The SDE is consistently less in CARRA than in ERA5, with the exception of being similar in the mountains during spring/summer. The largest difference is found in CoNo during winter, with ca. 1°C and ca. 3°C in CARRA and ERA5, respectively. Furthermore, both CARRA and ERA5 experience larger SDEs during winter than in summer. In general, the SDE is smaller in winter for regions that are locally influenced by the prescribed sea-surface temperature (islands, CoNo, GoBothnia) and that are also on average warmer, with less variability, than other regions (NSF inland, Svalbard, mountains).

The regional biases for WS10 in CARRA (Fig. 3) vary from an underestimation in the mountains (ca. −1 m/s) and GoBothnia (ca. −0.5 m/s), a minor positive bias for NSF inland and CoNo, and a more pronounced overestimation on islands and in Svalbard (ca. 1 m/s). In general, the positive wind biases found in CARRA are more pronounced in winter than in summer. For most regions and months the WS10 biases in ERA5 are similar or larger than for CARRA in absolute value, but the sign of the biases changes between the reanalyses in certain regions, for example, Svalbard and islands. The largest biases in ERA5 are found in the mountains (−3 to −4 m/s), Svalbard (−1 to −1.5 m/s) and islands (+1 to +1.5 m/s).

The SDEs for WS10 are smaller in CARRA than ERA5 for NSF inland, GoBothnia and CoNo, similar at Svalbard and islands, and larger in the mountains. However, SDE can be expected to scale with WS10 itself and is, therefore, expected to be lower in ERA5, which underestimates WS10 seen in the mountains. Also, the annual cycle of SDEs shows larger errors in winter than in summer as the wind speed and its variability increase.

The agreement between reanalyses and observations vary with region. In general, CARRA is in better agreement

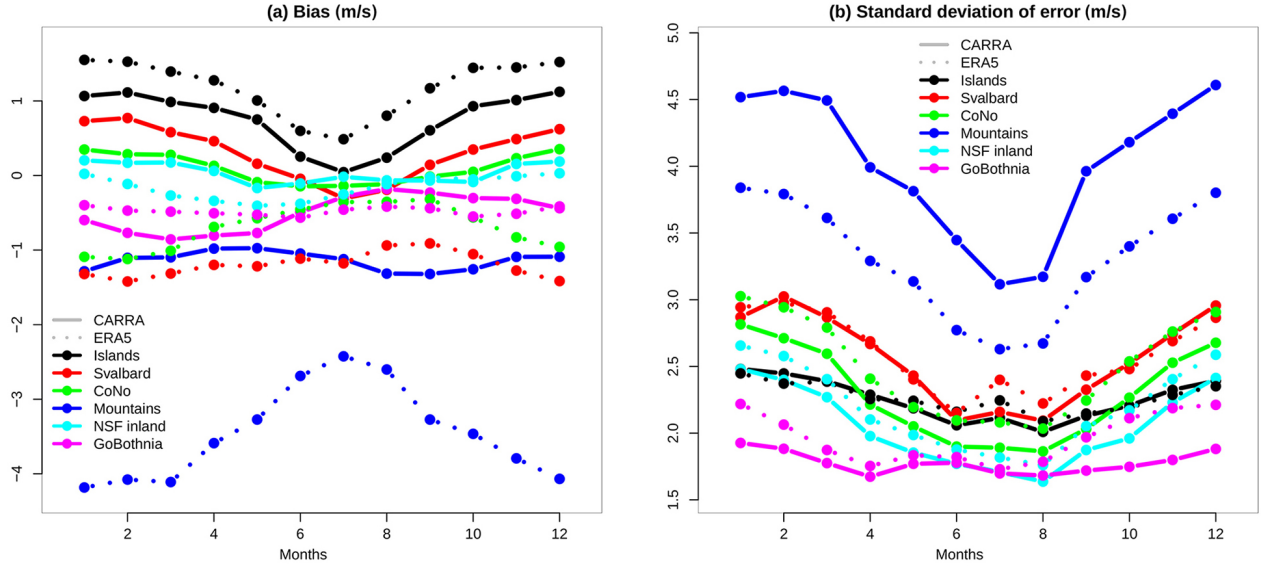

Fig. 3 Annual cycle of (a) WS10 bias and (b) SDE for WS10 for CARRA and ERA5 compared with observations for different regions: islands, Svalbard, CoNo, mountains, NSF inland and GoBothnia.

with the observations than ERA5 for both systematic (bias) and unsystematic (SDE) errors. The differences also follow the patterns of potential added value shown in Fig. 1 to a certain degree, that is, the objective added value of CARRA is high in CoNo and Svalbard and for the biases in the mountains. Unfortunately, the potential added value in T2m seen over the sea ice is difficult to evaluate with the observation data set applied in this study. However, results from Karl XII-øya, a small island in north-eastern Svalbard (80.65°N, 25.0°E) and represented as an ocean point in both reanalyses, hint towards a better representation in CARRA (bias ca. 2.0°C) than in ERA5 (bias ca. 4.0°C) for days when the observation site is surrounded by sea ice. However, more observational data from the sea ice and analyses are required to conclude on this topic.

Spatial and temporal variability

The finer grid spacing of CARRA suggests that small-scale spatial patterns should be better represented in CARRA than in ERA5. Ideally, high-resolution gridded observations should be used to evaluate this. However, in the absence of such a data set, we apply point observations following the approach of Marzban et al. (2009). Firstly, the correlations between all observation sites for T2m and WS10 in the observations, CARRA and ERA5 are calculated. Then, the correlations are averaged in bins as a function of distance and plotted as variograms showing the (de)correlation as a function of distance. A rapid decorrelation with distance will then indicate strong dominance of small-scale features.

The decorrelation of T2m with distance (Fig. 4) is more pronounced in the observations than in CARRA and ERA5. However, the decorrelation curve for CARRA is in better agreement with the observations than that in ERA5. For distances larger than a few tens of kilometres, there is a difference between summer (slower) and winter (faster) decorrelation in the observations indicating more small-scale features in winter. However, this seasonal difference is less pronounced in both reanalyses than in the observations. Also, for WS10 both reanalyses underestimate the decorrelation with distance, but CARRA is in better agreement with the observations than that in ERA5 (Fig. 4). Furthermore, the decorrelation for WS10 with distance is larger in summer (more small-scale features) than winter in the observations, opposite to what was seen for T2m. However, for WS10 the difference between summer and winter is better captured by both CARRA and ERA5. The very rapid decorrelation for the shortest distances (seen for T2m, and in particular for WS10) can contribute to representativeness issues of point observations when compared to gridded data (e.g., sub-grid variability) and will be further discussed below.

The temporal variability of T2m and WS10 in CARRA, ERA5 and the observations is compared at each observation site by their SD. The T2m variability during winter in CARRA agrees in general very well with the variability of the observations (Fig. 5), both for individual observation sites, but also averaged over regions. The only pronounced deviation between CARRA and the observed variability is for NSF inland, where observation sites with large (less) observed variability are underestimated (overestimated), resulting in an, on average, modest underestimation by CARRA for the region. Also, ERA5 reproduces the temporal variability for T2m reasonably well (Fig. 5), but with slightly larger deviations than CARRA. The deficiencies for NSF inland are more pronounced in ERA5,

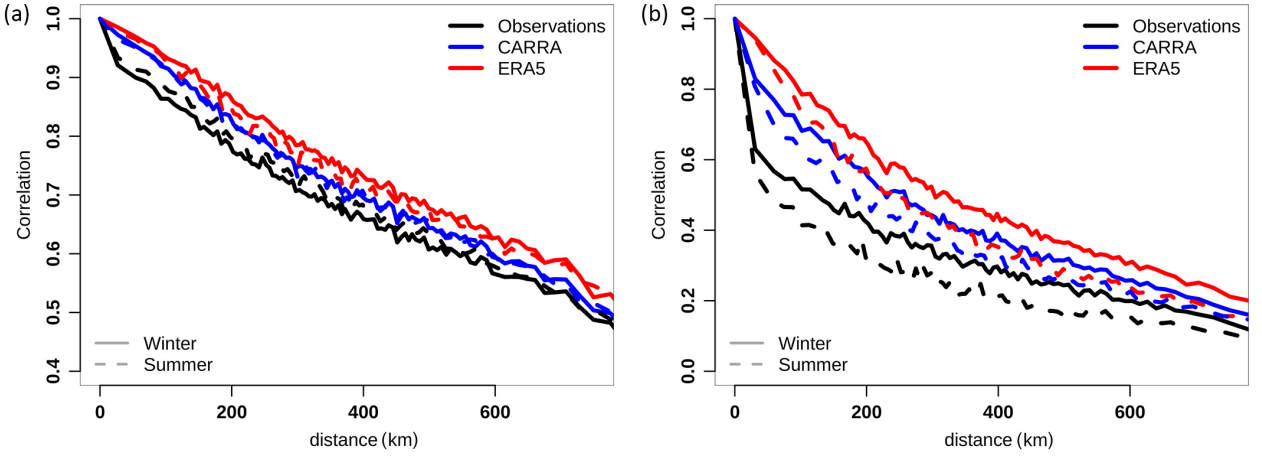

Fig. 4 Variograms showing spatial correlation between sites in observations and reanalysis of (a) T2m and (b) WS10. Correlation as a function of distance between SYNOP sites is calculated, and the average over stations with similar distances are plotted.

Citation: *Polar Research* 2022, *41*, 8002, http://dx.doi.org/10.33265/polar.v41.8002

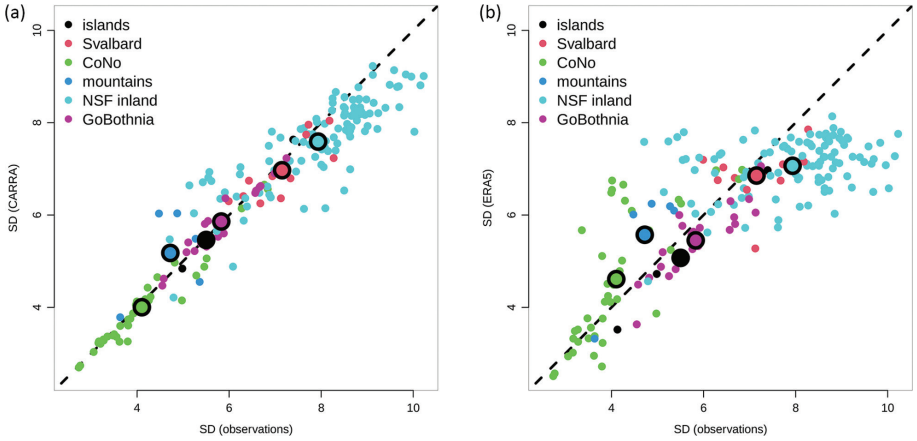

Fig. 5 Temporal variability in winter, comparing the SD at observation sites of (a) CARRA and (b) ERA5 with observations for hourly T2m. Observation sites are coloured by regions: islands, Svalbard, CoNo, mountains, NSF inland and GoBothnia. Average values over the different regions are shown in larger dots ringed in black circles.

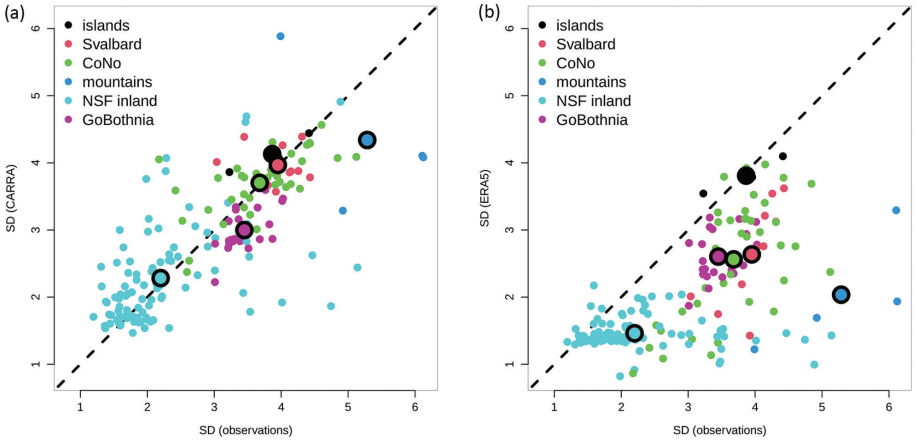

Fig. 6 Temporal variability in winter, comparing the SD at observation sites of (a) CARRA and (b) ERA5 with observations for hourly WS10. Observation sites are coloured by regions: islands, Svalbard, CoNo, mountains, NSF inland and GoBothnia. Average values over the different regions are shown in larger dots ringed in black circles.

and the temporal variability at a number of the observation sites at CoNo is heavily overestimated.

The temporal variability of WS10 in the reanalyses shows less agreement with the observations than that seen for T2m (Fig. 6). CARRA is on average similar to the observations for CoNo, NSF inland, islands and Svalbard, but it underestimates, on average, the variability in the mountains and GoBothnia. Furthermore, the agreement in variability varies substantially between individual observation sites within each region. ERA5 shows a substantial, on average, underestimation of the observed variability in all regions, with the exception of islands. The underestimation is largest in the mountains (related to the general underestimation of WS10 seen in Fig. 3), while in particular in NSF inland, ERA5 is not able to properly represent the inter-site differences. The temporal variability for both T2m and WS10 is less in summer than in winter, but the patterns in agreement between observations and the reanalyses are very similar, as in winter (not shown).

Extremes

Climate and weather extremes are important aspects of weather forecasting and climate research. Their representations in reanalyses are, therefore, of wide interest and are expected to improve with finer resolution (e.g.,

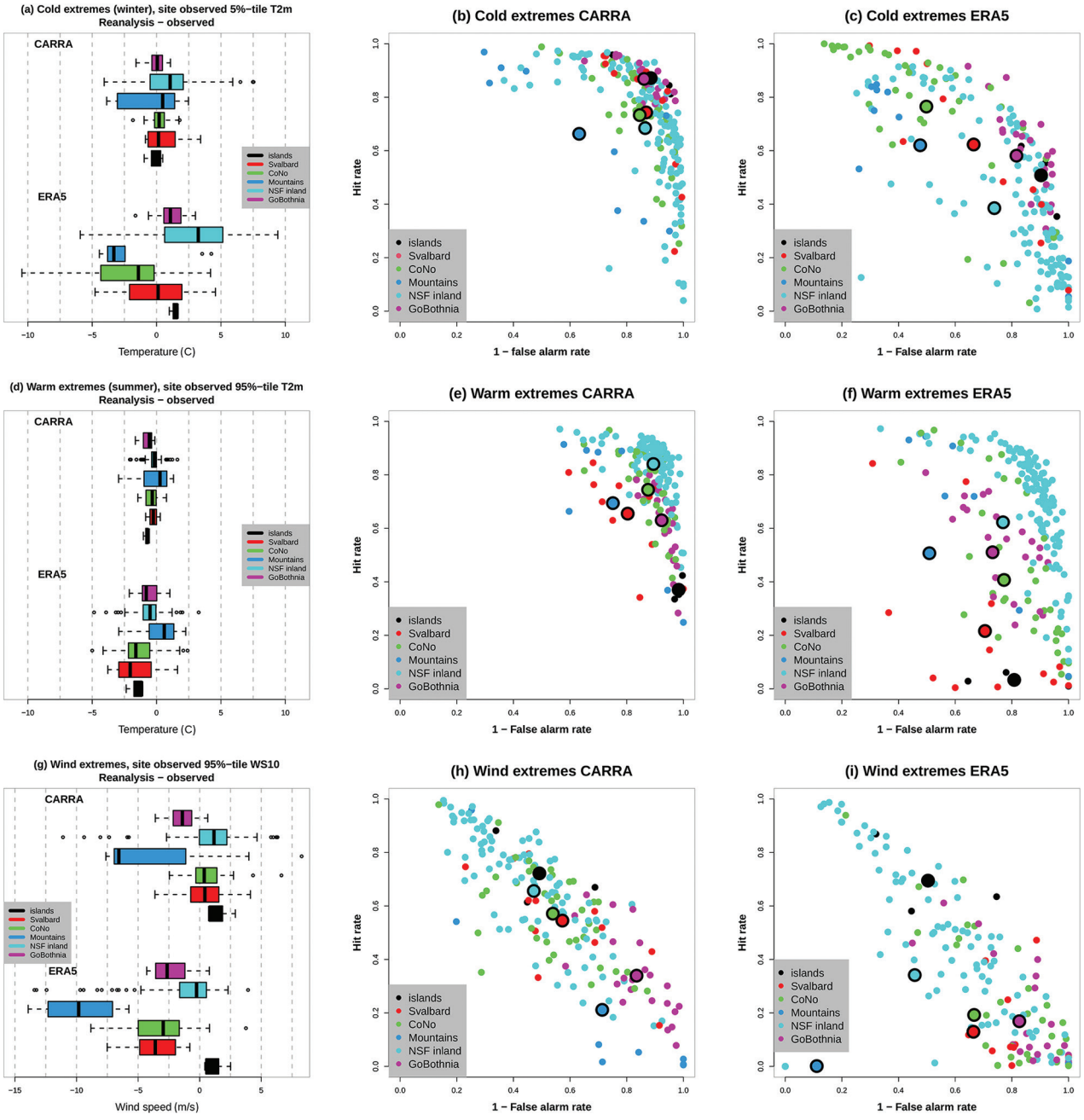

Fig. 7 Representation of (a) cold extremes (winter 5th percentile T2m), reanalysis versus observations, comparing the timing of events in observations and reanalysis, by plotting the fraction of observed events identified by the reanalysis (x axis) versus the fraction of reanalysis events that are observed (y axis) for (b) CARRA and (c) ERA5, and similarly for (d–f) warm extremes (summer 95th percentile T2m), and (g-i) high wind speed extremes (winter 95th percentile WS10).

Sheridan et al. 2020; Walsh et al. 2020; Avila-Diaz et al. 2021). There is no uniform definition of extremes (Walsh et al. 2020), but in the following, we define on the basis of observation locations with more than five years of observations: cold extremes as hourly T2m below the local observed 5th percentile during winter; warm extremes as hourly T2m above the local observed 95th percentile during summer; and wind extremes as hourly WS10 above the local observed 95th percentile during winter.

Citation: *Polar Research* 2022, *41*, 8002, http://dx.doi.org/10.33265/polar.v41.8002

Using these definitions, the thresholds calculated for observations CARRA and ERA5 at each observation site are compared to provide insight on the representation of the climatology of the extremes. In addition, the hit rate by the reanalysis is plotted against one minus the fraction of extremes in the reanalysis which are not observed (1—false alarm rate). A high value of the former indicates that the observed extremes are identified by the reanalysis. However, this can be obtained by always having extremes in the reanalysis, but a high (small) value of the latter indicates that one has a small (high) amount of false alarms in the reanalysis. A perfect representation in the reanalysis is, therefore, given by identical thresholds (percentiles) in observations and the reanalysis (Fig. 7a, d, g, for cold events, warm events and high wind events, respectively), and as a single point, indicating a 100% hit rate at the same time as no false alarms are present for CARRA (Fig. 7b, e, h) and ERA5 (Fig. 7c, f, i).

The climatology for cold extremes in CARRA agrees reasonably well with the observed 5th percentile but there is a small warm bias and higher site-to-site variability in quality for NSF inland stations and in the mountains (Fig. 7a, b). Compared to this, ERA5 shows clear weaknesses with a warm bias for islands, GoBothnia and NSF inland, and a cold bias at CoNo and in the mountains. In addition, the site-to-site variability is higher in ERA5 than in CARRA, except for that in the mountains. Also, the timing of the cold extremes in CARRA is superior to ERA5 for all regions. The fraction of the observed events identified by CARRA is ca. 0.6 or above for all regions, while, with the exception of CoNo, the same fraction is ca. 0.6 or lower in ERA5. Furthermore, the fraction of cold extremes present in the reanalyses that also are observed is well above ca. 0.8 (with the exception of in the mountains) in CARRA, while for ERA5 only GoBothnia and islands are above ca. 0.8.

The site-to-site variability in quality in the representation of warm extremes in both reanalyses is smaller than that of cold extremes (Fig. 7d–f). Both reanalyses show a small underestimation (not warm enough) in all regions, except for a small overestimation in the mountains. Even if this pattern is similar in both reanalyses, the underestimation is more pronounced in ERA5. The timing of the warm extremes is somewhat different from the cold extremes, that is, there is a tendency that a smaller part of the observed events is captured in the reanalyses, while a larger part of the events in the reanalyses are also observed. On average, the climatology in the reanalyses is better captured for warm extremes than cold extremes. However, the same is not necessarily true for the timing of the extremes.

The representation of wind extremes in the reanalyses includes some pronounced deficiencies (Fig. 7g–i). CARRA has a modest overestimation on islands, in Svalbard, in CoNo and inland (up to a couple of m/s) and an underestimation at GoBothnia (−1.5 m/s) and in the mountains (ca. −7 m/s). Compared to this, the deficiencies are larger in ERA5, with a small overestimation at islands (ca. 1 m/s) and underestimations in all other regions from very small (inland) to very large in the mountains (ca. −10 m/s). Furthermore, the timing of the events is not as good as for temperature in both reanalyses, but with some clear differences between regions and between the reanalyses. With the exception of islands, there are clear differences between CARRA and ERA5, that is, CARRA usually has a higher hit rate, but at the cost of more false alarms than ERA5. In general, regions with a positive WS10 bias in Fig. 3 (e.g., NSF inland and islands for CARRA) have a higher hit rate, but with more false alarms, while regions with a negative WS10 bias (e.g., GoBothnia and mountains in CARRA) show a lower hit rate but also less false alarms.

In this analysis, we used hourly observations and hourly output from reanalyses, that is, we neglected subhourly variations in both measurements and reanalyses, which would give more extreme values. This could have had an impact on the results, but we believe the qualitative strengths and weaknesses of the two reanalyses would remain the same. The results show that hourly extremes in T2m and WS10 on average are better represented in CARRA than ERA5. The identification of the individual extreme events (correct time and location) is still problematic, as also reported by Sheridan et al. (2020), but is improved by CARRA. This is especially seen for cold temperatures and for high wind speeds. The results further suggest that extremes for temperature (which also are used in the surface data assimilation) are better captured than the extremes for wind speed (which are not assimilated). This is a similar conclusion as made by Avila-Diaz et al. (2021) for temperature (assimilated) and precipitation (not assimilated).

Examples of high-impact events

We will now briefly investigate the representation of T2m and WS10 during some specific high-impact Arctic weather events to complement the analysis on extremes above. The chosen events are polar lows, maritime icing events and warm/melting events during winter, all of which have a potential high impact on society and safety of operations.

Polar lows. Even though there is no universal definition of polar lows, they are usually considered to be intense maritime cyclones with a horizontal scale of between 200 and1000 km and high near-surface wind speeds occurring in cold air outbreaks over high-latitude open oceans (Spengler et al. 2017; Moreno-Ibanez

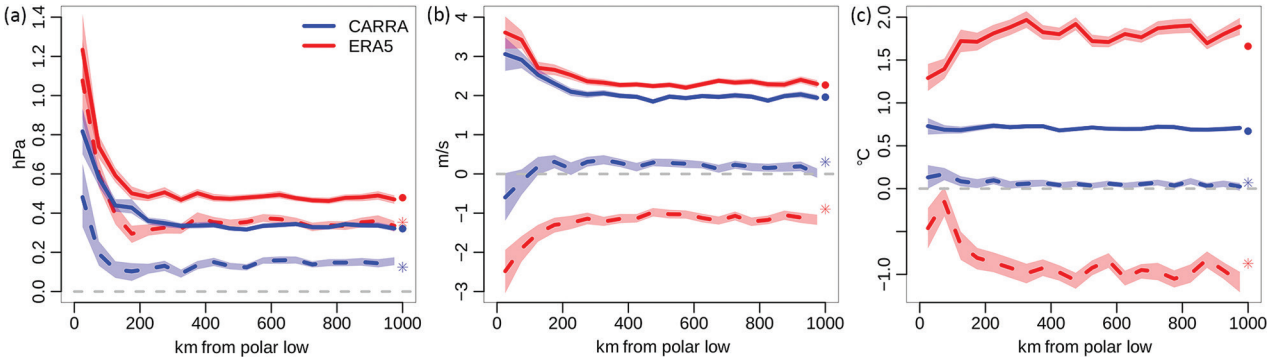

Fig. 8 Verification at the CoNo of (a) MSLP, (b) WS10 and (c) T2m during polar lows as a function of distance to the polar low centre. Biases are shown in dashed lines, MAE in solid lines, CARRA in blue and ERA5 in red. The shaded area shows the 95th percentile confidence interval estimated by bootstrapping with 1000 replicas sampled from the reanalyses and observation data with replacement. The biases (asterisks) and MAEs (circles) averaged over all conditions at CoNo from November to April are plotted for comparison to the right in each plot.

et al. 2021). With the high winds and rapid change in local weather conditions, they can pose a threat to coastal and maritime installations and activities. A widespread weakness in the representation of polar lows in existing reanalyses is the underestimation of near-surface wind speeds, but finer resolution may improve on this (Moreno-Ibanez et al. 2021). In this study, we use the polar low database of Rojo et al. (2019) and compare CoNo SYNOPs of MSLP, WS10 and T2m with CARRA and ERA5 in the proximity of polar lows to describe their representation at landfall.

The bias and MAE for CARRA and ERA5 are plotted as a function of distance to the polar low centre for 341 polar lows in the period 1998 to 2019 (Fig. 8). Both CARRA and ERA5 have a positive bias (more pronounced in ERA5) in MSLP, implying that the polar lows are not deep enough in the reanalyses. The MAE is also larger in ERA5 than in CARRA. Away from the polar low centre, the errors stabilize close to the errors typical for CoNo during winter (shown to the right in each panel). For WS10, the general underestimation of WS10 for CoNo in ERA5 (Fig. 3) is even more pronounced close to the centre of the polar lows (2–3 m/s). However, as suggested by Moreno-Ibanez et al. (2021), WS10 is much better captured in the high-resolution CARRA, with only a weak, but not significant, underestimation close to the polar lows. In the verification of T2m, CARRA shows relatively limited sensitivity to the presence of the polar lows, while ERA5 shows substantial lower errors for T2m during polar lows than on average but still larger than what is found for CARRA. This comparison shows that CARRA represents the consequences of polar lows associated with WS10 and T2m better than ERA5, which indicates that CARRA can be a useful data set to further investigate polar lows.

Coastal and maritime icing. Subfreezing temperatures interacting with waves and high wind speeds generate sea spray that may freeze on ships and maritime installations, with potentially severe consequences, including ship-capsizing and human casualties (Samuelsen 2017). The dominant weather type during maritime icing events is cold-air outbreaks from the ice-covered ocean areas (e.g., around Svalbard), while approximately 10% of the events arise from cold-air outbreaks from northern Norway and appear close to the Norwegian coastline (Samuelsen & Graversen 2019). The representation of the latter in the reanalyses is investigated in the following. To circumvent the lack of direct icing observations and to focus on the atmospheric conditions in the reanalyses, we forced a state-of-the-art ship-icing model (Samuelsen 2017) with T2m and WS10 from observations and reanalyses at a set of exposed observation sites in CoNo similar to the approach demonstrated by Køltzow et al. (2020). In addition, sea-surface temperature from ERA5 is used to force the icing model. Hence, the output of the icing model, based on observed input, is used as an estimate of observed icing and can be compared with the output of the icing model when forced with reanalyses to evaluate the combination of T2m and WS10 from the reanalyses.

The correlation of icing intensity and frequency bias, hit rate and false alarm ratio for icing/no-icing are given in Table 1. In general, CARRA fits the observed estimates of icing better than ERA5, with stronger correlation, higher hit rate and a lower false alarm ratio. Both CARRA and ERA5 underestimate the frequency of icing occurrences, but CARRA underestimates it to a lesser degree. To identify the extent to which T2m and WS10 are sources of the disagreement between observed estimate

Citation: *Polar Research* 2022, *41*, 8002, http://dx.doi.org/10.33265/polar.v41.8002

Table 1 Validation of vessel icing at exposed observation sites on the Norwegian coast. Scores are calculated after the vessel icing model is forced with T2m and WS10 from CARRA and ERA5, respectively. The first number in parentheses represents the result when only WS10 is taken from the reanalysis, that is, T2m from observations, and second number is when only T2m is taken from the reanalysis, that is, WS10 is taken from observations. The estimate of observed icing is done by forcing the icing model with observed T2m and WS10.

Verification vessel icing	Correlation	Frequency bias	Hit rate	False alarm ratio
CARRA	0.74 (0.78, 0.97)	0.75 (0.83, 0.91)	0.55 (0.61, 0.88)	0.27 (0.27 0.04)
ERA5	0.63 (0.65, 0.89)	0.48 (0.50, 1.1)	0.33 (0.38, 0.88)	0.32 (0.25, 0.20)

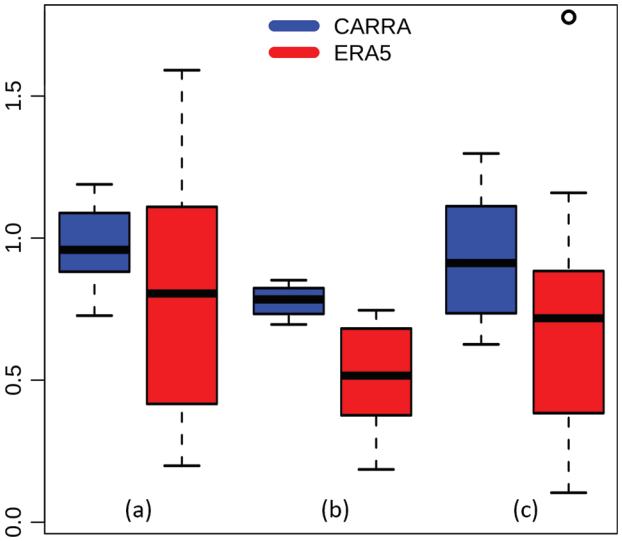

Fig. 9 Warm spell diagnostics for observation sites used in Vikhamar-Schuler et al. (2016) for CARRA and ERA5. Box plots show site-to-site variations in (a) frequency bias and (b) ETS score of above 0°C events, and (c) the sum of degree-hour above 0°C in reanalysis divided by observed. Outliers are indicated as open circles.

of icing and reanalyses, we also ran the icing model with a combination of observed T2m and WS10 from reanalyses, and the other way around. The results, given in the parentheses in Table 1, show that for the reanalyses and in particular for ERA5, the problematic issue is the representation of WS10. For example, the correlation is increased to 0.97 (CARRA) and 0.89 (ERA5) when using WS10 from observations, while it only increases to 0.78 (CARRA) and 0.65 (ERA5) while using T2m from the observations. Notice also that the frequency bias in ERA5 is 1.1 when WS10 from observations is used, reflecting on the cold bias in ERA5 in CoNo discussed earlier. In summary, the icing events are not perfectly represented in the reanalyses but are better captured in CARRA than ERA5. In addition, WS10 is the main problem in the representation of maritime icing events from reanalyses at the Norwegian coast.

Warm (melting) events. Warm (melting) events and rain-on-snow events during the Arctic winter have the potential for major consequences in places like Svalbard, with substantial impact on infrastructure, society and wildlife (e.g., Hansen et al. 2014; Serreze et al. 2015). Since this study focuses on the representation of T2m (not precipitation), we apply the indexes for warm events as used by Vikhamar-Schuler et al. (2016), that is, events with T2m above 0°C and the sum of degree-hour above 0°C during winter, to evaluate the representation of warm events.

Figure 9 shows that the frequency bias of above 0°C events is well captured in CARRA (approximately the same number of events as in the observations) averaged over all observation sites, while ERA5 on average underestimates the frequency. Furthermore, the site-to-site variability in frequency bias is less for CARRA than ERA5. The accuracy in the timing of the events is measured by the ETS, for which 1 (0) imply a perfect (no) match between the reanalysis and the observations. The ETS clearly shows that CARRA is more accurate than ERA5 in the identification of these warm events. Finally, for the sum of degree-hours, the correspondence with observations is also better for CARRA than ERA5. On average, a small underestimation is seen in CARRA, with a more pronounced underestimation and larger site-to-site variability for ERA5. Again, CARRA agrees better with the observations than ERA5 does, in line with the general verification of T2m, described above.

Representativeness errors. The difference between gridded reanalysis and point observations can be attributed to errors in the reanalysis, errors in the observations and representativeness errors. We have applied the same quality-controlled observations to evaluate both CARRA and ERA5. Therefore, the observation error part is small and similar in the two reanalyses. The differences in scores between CARRA and ERA5 can, therefore, be attributed to model errors and representativeness errors. Point observations may show substantial sub-grid variability that cannot be reproduced by the gridded reanalysis and are seen as a part of the very rapid decorrelation during the first kilometres in Fig. 4. We approximate the representativeness error based on the simple approach of Göber et al. (2008), by assuming that the mean of nearby observations represents an

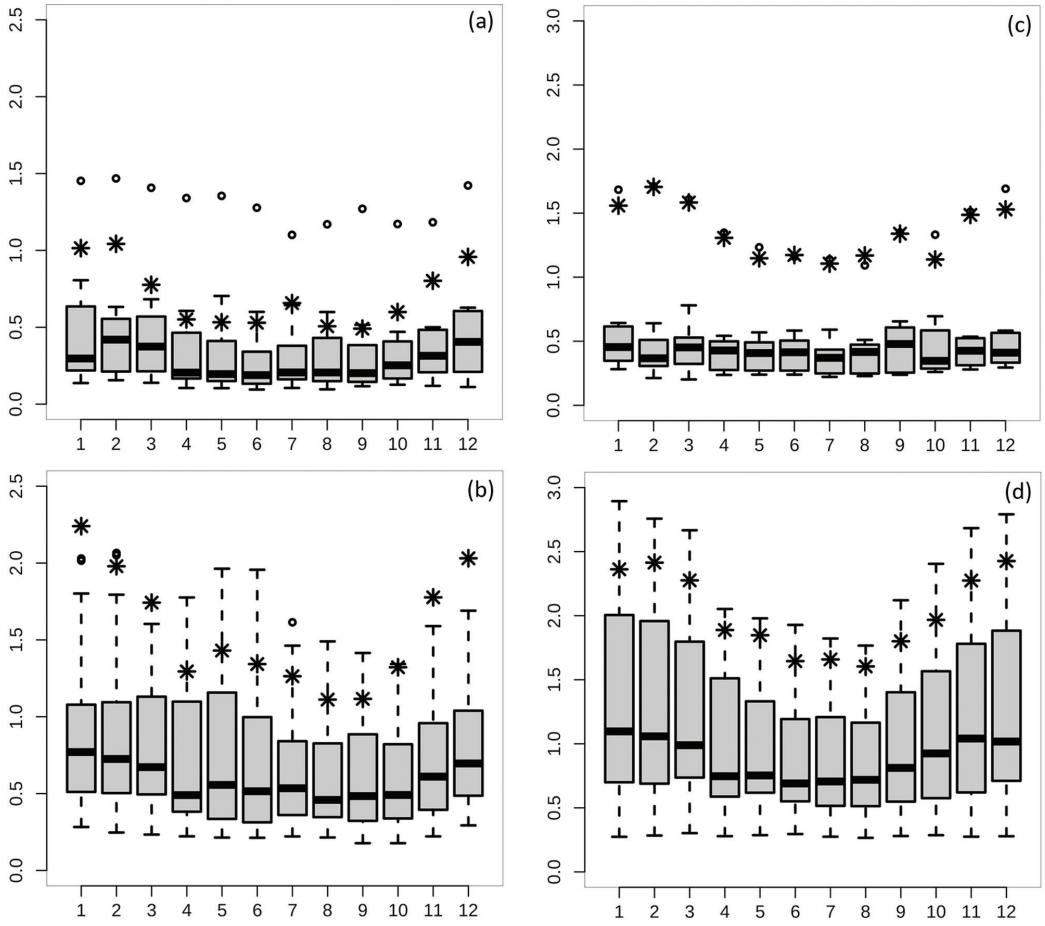

Fig. 10 Estimates of the annual cycle of representativeness errors for T2m measured by (a) MAE for 15 station pairs 0.5–3 km apart, apart representing the CARRA grid spacing of 2.5 km, and (b) 44 pairs 10–30 km apart, representing the ERA5 grid spacing of ca. 30 km, shown as box plots illustrating the variations over the different locations of observation pairs. Similar for WS10 in (c) and (d). Outliers are indicated as open circles. Asterisks in (a) and (c) represent the MAEs of CARRA for the same stations; asterisks in (b) and (d) represent the MAEs of ERA5.

approximation of the grid box mean, which we then treat as the 'perfect' grid box forecast. We then verify the perfect grid box representation against the observations, that is, the departure of the observations from the grid box mean is used to calculate MAE, but the error will not be 0 unless the observations are equal, implying that the sub-grid variability is zero. The error of the 'perfect' grid box value can then be regarded as an approximation of the representativeness error and will vary depending on the grid box size and will be different for CARRA and ERA5.

The box plots in Fig. 10 show the annual cycle of the perfect grid box MAEs calculated for 15 station pairs 0.5–3.0 km apart (to approximate representativeness errors of CARRA, stations up to 3.0 km apart are included to increase the number of observation pairs) and for 44 station pairs 10–30 km apart (to approximate representativeness errors of ERA5). We emphasize that this is a moderate approximation of the representativeness errors since the effective resolution of the models is even larger than the model grid spacing (Skamarock 2004). The T2m representativeness error for CARRA is slightly above (below) 0.5°C in winter (summer), while for ERA5 it is slightly above (below) 1.0°C in winter (summer). Therefore, the error shows not only a seasonal behaviour but also an increase with increasing grid spacing. However, since a similar annual cycle is seen in the errors of the reanalyses themselves and a smaller error is seen in CARRA than ERA5, the median T2m representativeness error is between 50 and 60% of the total MAE for most months and both reanalyses. The representativeness error for WS10 has a slightly higher absolute value than T2m, but a similar annual cycle for both grid spacings. However, the proportion of

Citation: *Polar Research* 2022, *41*, 8002, http://dx.doi.org/10.33265/polar.v41.8002

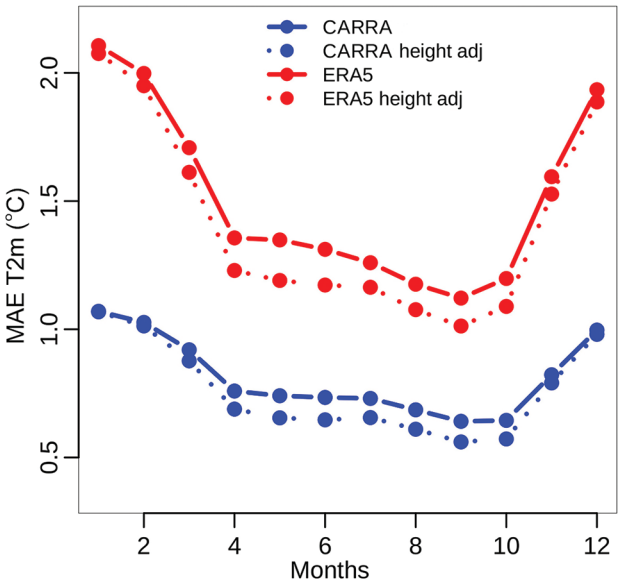

Fig. 11 Annual cycle of MAE T2m averaged over all stations, with and without adjusting for height differences between observation sites and model topography.

the total difference between the reanalyses and the observations arising from representativeness issues is different for CARRA and ERA5. For CARRA, 40–55% of the WS10 error can be attributed to representativeness issues, while values as high as 60–70% are found for ERA5.

An integrated part of the representativeness issue is the vertical height difference between the reanalyses and the point observations. This is substantially different between CARRA and ERA5, because of the different resolutions' ability to resolve the topography. In Fig. 11, a simple height correction (0.65 °C/100 m) is applied to adjust for the height difference before the MAE is calculated. During summer, this reduces the errors by 10–15% in both reanalyses. However, in winter the impact of this adjustment is negligible, most likely because of more stratified conditions, including temperature inversions, for which the simple linear adjustment is no longer valid. However, this demonstrates that height differences account for a part of the representativeness errors.

Summary and conclusions

Reanalyses provides data sets useful for climate and meteorological research. However, they are deficient in representing the true climate system for a number of reasons, such as the limited amount of observations available for assimilation, resolution of the applied NWP system and limitations in the assimilation procedure or description of physical processes in the NWP system. Therefore, an evaluation of a reanalysis is always useful to aid the utilization of the data and to stimulate developments that will reduce identified weaknesses in the next generation of reanalyses. This study investigated the added value for near-surface temperature and wind speed with the newly released CARRA data set, in comparison to the global reanalysis ERA5 in the north-east European Arctic. In both reanalyses, near-surface temperatures are used in the surface assimilation procedure and therefore constrained by the assimilation schemes, and hence, the evaluation is not done against fully independent observations. However, near-surface wind speeds are not assimilated and can be looked upon as independent observations. It is believed that CARRA will improve over ERA5 because of CARRA's (1) improved horizontal resolution (2.5-km vs. 31-km grid spacing), (2) use of more local observation data in the data assimilation process and (3) increased focus on the representation of the cold surfaces relevant for the region. The main conclusions are as follows.

The largest differences between CARRA and ERA5 are associated with complex terrain, highlighting the importance of fine model resolution to resolve surface inhomogeneity. However, the representation of sea ice also gives large differences in near-surface temperature during winter.

For near-surface temperature and wind speed, CARRA shows good agreement with observations and is substantially better than ERA5. However, the level of fidelity seems to be less strong for wind speed than for temperature in both reanalyses. Based on how near-surface temperature is used in the assimilation process while the wind speed is not, this is as expected.

CARRA adds value in terms of improved general verification statistics for all regions.

As a consequence, CARRA also adds fidelity in its representation of the assessed high-impact weather events—polar lows, maritime icing and warm spells.

CARRA generally adds value in the investigated aspects of its climatological distributions: in the representation of spatial variability and temporal variability and climatology of extremes.

There is a substantial regional difference in how well the CARRA reanalysis corresponds with observations, for example, biases in wind speed vary considerably between regions, and for near-surface temperature, the errors are smaller in regions influenced by open ocean, with prescribed sea-surface temperature derived from satellite measurements.

A clear annual cycle with larger errors in winter and smaller errors in summer is found in all regions, for both parameters and both reanalyses.

A substantial part of the difference between reanalyses and observations can be explained by representativeness errors, that is, sub-grid variability that the reanalyses could not possibly reproduce because of their given grid spacing. This representativeness error is larger in ERA5 than CARRA, but for near-surface temperature the percentage of the total error is similar in both reanalyses, while for wind speed it is slightly higher in ERA5 than in CARRA.

This study has revealed some strengths and weaknesses of CARRA and ERA5 but is far from exhaustive. We, therefore, strongly welcome more evaluation studies for other parameters with fully independent (temperature) observations. However, the presented results are very promising for the usefulness of CARRA in climate and weather research and in other areas where this information is used. The results are also very encouraging for the production of future global and regional high-resolution reanalyses.

Acknowledgements

The authors thank the entire CARRA development and production team for input and discussions on a number of topics related to this study. The work has benefitted especially from dialogue with Andras Horanyi at ECMWF. This work is a contribution to the Year of Polar Prediction, a flagship activity of the Polar Prediction Project, initiated by the World Meteorological Organization's World Weather Research Programme.

Disclosure statement

The authors report no conflict of interest.

Funding

This work has been funded by the Copernicus Climate Change Service. ECMWF implements this service on behalf of the European Commission. This study was also supported by the Norwegian Research Council Project 280573, Advanced Models and Weather Prediction in the Arctic: Enhanced Capacity from Observations and Polar Process Representations (ALERTNESS).

References

Arduini G., Balsamo G., Dutra E., Day J.J., Sandu I., Boussetta S. & Haiden T. 2019. Impact of a multi-layer snow scheme on near-surface weather forecasts. *Journal of Advances in Modeling Earth Systems 11*, 4687–4710, doi: 10.1029/2019MS001725.

Atlaskin E. & Vihma T. 2012. Evaluation of NWP results for wintertime nocturnal boundary-layer temperatures over Europe and Finland. *Quarterly Journal of the Royal Meteorological Society 138*, 1440–1451, doi: 10.1002/qj.1885.

Avila-Diaz A., Bromwich D.H., Wilson A.B., Justino F. & Wang S. 2021. Climate extremes across the North American Arctic in modern reanalyses. *Journal of Climate 34*, 2385–2410, doi: 10.1175/JCLI-D-20-0093.1.

Batrak Y. & Müller M. 2019. On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice. *Nature Communications 10*, article no. 4170, doi: 10.1038/s41467-019-11975-3.

Betts A.K., Chan D.Z. & Desjardins R.L. 2019. Near-surface biases in ERA5 over the Canadian prairies. *Frontiers in Environmental Science 7*, article no. 129, doi: 10.3389/fenvs.2019.00129.

Bromwich D.H., Wilson A.B., Bai L., Liu Z., Barlage M., Shih C.-F., Maldonado S., Hines K.M., Wang S.-H., Woollen J., Kuo B., Lin H.-C., Wee T.-K., Serreze M.C. & Walsh J.E. 2018. The Arctic System Reanalysis, version 2. *Bulletin of the American Meteorological Society 99*, 805–828, doi: 10.1175/BAMS-D-16-0215.1.

Copernicus Climate Change Service 2017. *ERA5: fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service Climate Data Store (CDS)*. Accessed on the internet at https://cds.climate.copernicus.eu/cdsapp#!/home on 18 June 2021

Delhasse A., Kittel C., Amory C., Hofer S., van As D., S. Fausto R. & Fettweis X. 2020. Brief communication: evaluation of the near-surface climate in ERA5 over the Greenland Ice Sheet. *The Cryosphere 14*, 957–965, doi: 10.5194/tc-14-957-2020.

Demchev D.M., Kulakov M.Y., Makshtas A.P., Makhotina I.A., Fil'chuk K.V. & Frolov I.E. 2020. Verification of ERA-Interim and ERA5 Reanalyses Data on surface air temperature in the Arctic. *Russian Meteorology and Hydrology 45*, 771–777, doi: 10.3103/S1068373920110035.

Göber M., Zsótér E. & Richardson D.S. 2008. Could a perfect model ever satisfy a naïve forecaster? On grid box mean versus point verification. *Meteorological Applications 15*, 359–365, doi: 10.1002/met.78.

Graham R.M., Cohen L., Ritzhaupt N., Segger B., Graversen R.G., Rinke A., Walden V.P., Granskog M.A. & Hudson S.R. 2019. Evaluation of six atmospheric reanalyses over Arctic sea ice from winter to early summer. *Journal of Climate 32*, 4121–4143, doi: 10.1175/JCLI-D-18-0643.1.

Hansen B.B., Isaksen K., Benestad R.E., Kohler J., Pedersen Å.Ø., Loe L.E., Coulson S.J., Larsen J.O. & Varpe Ø. 2014. Warmer and wetter winters: characteristics and implications of an extreme weather event in the High Arctic. *Environmental Research Letters 9*, article no. 114021, doi: 10.1088/1748-9326/9/11/114021.

Hersbach H., Bell B., Berrisford P. Hirahara S., Horányi A., Muñoz-Sabater J., Nicolas J., Peubey C., Radu R., Schepers D., Simmons A., Soci C., Abdalla S., Abellan X., Balsamo G., Bechtold P., Biavati G., Bidlot J., Bonavita M., De Chiara G., Dahlgren P., Dee D., Diamantakis M., Dragani D., Flemming J., Forbes R., Fuentes M., Geer A.,

Haimberger L., Healy S., Hogan R.J., Hólm E., Janisková M., Keeley S., Laloyaux P., Lopez P., Lupu C., Radnoti G., de Rosnay P., Rozum I., Vamborg F., Villaume S. & Thépaut J.-N. 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*, 1999–2049, doi: 10.1002/qj.3803.

Hogan R.J. & Bozzo A. 2015. Mitigating errors in surface temperature forecasts using approximate radiation updates. *Journal of Advances in Modeling Earth Systems 7*, 836–853, doi: 10.1002/2015MS000455.

Kaiser-Weiss A.K., Borsche M., Niermann D., Kaspar F., Lussana C., Isotta F.A., van den Besselaar E., van der Schrier G. & Undén P. 2019. Added value of regional reanalyses for climatological applications. *Environmental Research Communication 1*, article no. 071004, doi: 10.1088/2515-7620/ab2ec3.

Kaspar F., Niermann D., Borsche M., Fiedler S., Keller J., Potthast R., Rösch T., Spangehl T. & Tinz B. 2020. Regional atmospheric reanalysis activities at Deutscher Wetterdienst: review of evaluation results and application examples with a focus on renewable energy, *Advances in Science and Research 17*, 115–128, doi: 10.5194/asr-17-115-2020.

Keller J.D. & Wahl S. 2021. Representation of climate in reanalyses: an intercomparison for Europe and North America. *Journal of Climate 34*, 1667–1684, doi: 10.1175/JCLI-D-20-0609.1.

Køltzow M., Casati B., Bazile E., Haiden T. & Valkonen T. 2019. An NWP model intercomparison of surface weather parameters in the European Arctic during the Year of Polar Prediction Special Observing Period Northern Hemisphere 1. *Weather and Forecasting 34*, 959–983, doi: 10.1175/WAF-D-19-0003.1.

Køltzow M., Hallerstig M., Graversen R., Jonassen M. & Mayer S. 2020. *Verification metrics and diagnostics appropriate for the (maritime) Arctic. MET Report 2/2020.* Oslo: Norwegian Meteorological Institute.

Marzban C., Sandgathe S., Lyons H. & Lederer N. 2009. Three spatial verification techniques: cluster analysis, variogram, and optical flow. *Weather and Forecasting 24,* 1457–1471, doi: 10.1175/2009WAF2222261.1.

Moreno-Ibáñez M., Laprise R. & Gachon P. 2021. Recent advances in polar low research: current knowledge, challenges and future perspectives. *Tellus Series A 73*, 1–31, doi: 10.1080/16000870.2021.1890412.

Müller M., Batrak Y., Kristiansen J., Køltzow M.A., Noer G. & Korosov A. 2017. Characteristics of a convective-scale weather forecasting system for the European Arctic. *Monthly Weather Review 145*, 4771–4787, doi: 10.1175/MWR-D-17-0194.1.

Rojo M., Noer G. & Claud C. 2019. Polar Low tracks in the Norwegian Sea and the Barents Sea from 1999 until 2019. *Pangaea*, doi: 10.1594/PANGAEA.903058.

Samuelsen E.M. 2017. *Prediction of ship icing in Arctic waters—observations and modelling for application in operational weather forecasting*. PhD thesis, UiT The Arctic University of Norway.

Samuelsen E.M. & Graversen R. 2019. Weather situation during observed ship-icing events off the coast of northern Norway and the Svalbard archipelago. *Weather and Climate Extremes 24*, article no. 100200, doi: 10.1016/j.wace.2019.100200.

Sandu I., Beljaars A., Bechtold P., Mauritsen T. & Balsamo G. 2013. Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models? *Journal of Advances in Modeling Earth Systems 5*, 117–133, doi: 10.1002/jame.20013.

Serreze M.C., Crawford A.D. & Barrett A.P. 2015. Extreme daily precipitation events at Spitsbergen, an Arctic island. *International Journal of Climatology 35*, 4574–4588, doi: 10.1002/joc.4308.

Sheridan S.C., Lee C.C. & Smith E.T. 2020. A comparison between station observations and reanalysis data in the identification of extreme temperature events. *Geophysical Research Letters 47*, e2020GL088120, doi: 10.1029/2020GL088120.

Skamarock W.C. 2004. Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review 132*, 3019–3032, doi: 10.1175/MWR2830.1.

Spengler T., Claud C. & Heinemann G. 2017. Polar Low Workshop summary. *Bulletin of the American Meteorological Society 98*, ES139–ES142, doi: 10.1175/BAMS-D-16-0207.1.

Tetzner D., Thomas E. & Allen C.A. 2019. Validation of ERA5 reanalysis data in the southern Antarctic Peninsula–Ellsworth Land region, and its implications for ice core studies. *Geosciences 2019*, article no. 289, doi: 10.3390/geosciences9070289.

Vikhamar-Schuler D., Isaksen K., Haugen J.E., Tømmervik H., Luks B., Schuler T.V. & Bjerke J.W. 2016. Changes in winter warming events in the Nordic Arctic region. *Journal of Climate 29*, 6223–6244, doi: 10.1175/JCLI-D-15-0763.1.

Walsh J.E., Ballinger T.J., Euskirchen E.S., Hanna E., Mård J., Overland J.E., Tangen H. & Wihma T. 2020. Extreme weather and climate events in northern areas: a review. *Earth-Science Reviews 209*, article no. 103324, doi: 10.1016/j.earscirev.2020.103324.

Wang C., Graham R.M., Wang K., Gerland S. & Granskog M.A. 2019. Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution. *The Cryosphere 13*, 1661–1679, doi: 10.5194/tc-13-1661-2019.

Yang X., Palmason B., Sattler K., Thorsteinsson S., Amstrup B., Dahlbom M., Hansen-Sass B., Pagn Nielsen K. & Petersen G.N. 2018. IGB, the upgrade to the joint operational HARMONIE by DMI and IMO in 2018. *Aladin-Hirlam Newsletter 11*, 93–96. Accessed on the internet at http://www.umr-cnrm.fr/aladin/IMG/pdf/nl11.pdf

Yang X., Schyberg H., Palmason B., Bojarova J., Box J., Pagh Nielsen K., Amstrup B., Peralta C., Høyer J., Nielsen Englyst P., Homleid M. Køltzow M.A.Ø., Randriamampianina R., Dahlgren P., Støylen E., Valkonen T., Thorsteinsson S., Kornich H., Lindskog M. & Mankoff K. 2020. *C3S Arctic regional reanalysis—full system documentation*. Accessed on the internet at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-carra-single-levels?tab=doc on 1 November 2021