

RESEARCH ARTICLE

Comparison of selected surface level ERA5 variables against in-situ observations in the continental Arctic

Jakob Boyd Pernov  | Jules Gros-Daillon | Julia Schmale 

Extreme Environments Research
Laboratory, École Polytechnique Fédérale
de Lausanne, Sion, Switzerland

Correspondence

Jakob Boyd Pernov and Julia Schmale,
Extreme Environments Research
Laboratory, École Polytechnique Fédérale
de Lausanne, 1951 Sion, Switzerland.

Email: jakob.pernov@epfl.ch;
julia.schmale@epfl.ch

Funding information

Swiss Data Science Center, Grant/Award
Number: C20-01

Abstract

In this study, data from 17 ground-based, continental Arctic observatories are used to evaluate the performance of the European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA5) reanalysis model. Three aspects are evaluated: (i) the overall reproducibility of variables at all stations for all seasons at one-hour time resolution; (ii) the seasonal performance; and (iii) performance between different temporal resolutions (one hour to one month). Performance is evaluated based on the slope, R^2 value, and root-mean-squared error (RMSE). We focus on surface meteorological variables including 2-m air temperature (temperature), relative humidity (RH), surface pressure, wind speed, zonal and meridional wind speed components, and short-wave downward (SWD) radiation flux. The overall comparison revealed the best results for surface pressure (0.98 ± 0.02 , $R^2_{\text{mean}} \pm \text{standard deviation } [\sigma_{R2}]$), followed by temperature (0.94 ± 0.02), and SWD radiation flux (0.87 ± 0.03) while wind speed (0.49 ± 0.12), RH (0.42 ± 0.20), zonal (0.163 ± 0.15) and meridional wind speed (0.129 ± 0.17) displayed poor results. We also found that certain variables (surface pressure, wind speed, meridional, and zonal wind speed) showed no seasonal dependency while others (temperature, RH, and SWD radiation flux) performed better during certain seasons. Improved results were observed when decreasing the temporal resolution from one hour to one month for temperature, meridional and zonal wind speed, and SWD radiation flux. However, certain variables (RH and surface pressure) showed comparatively worse results for monthly resolution. Overall, ERA5 performs well in the Arctic, but caution needs to be taken with wind speed and RH, which has implications for the use of ERA5 in global climate models. Our results are useful to the scientific community as it assesses the confidence to be placed in each of the surface variables produced by ERA5.

KEYWORDS

Arctic, ERA5, meteorology, reanalysis model/in-situ observation comparison

Jakob Boyd Pernov and Jules Gros-Daillon contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

1 | INTRODUCTION

The Arctic is warming three to four times faster than the global average (Pithan & Mauritsen, 2014; Rantanen *et al.*, 2022; Serreze & Barry, 2011), and this phenomenon is termed Arctic Amplification (AA). This faster warming rate has important consequences for local communities and ecosystems (AMAP, 2022). In addition, the modeling of climate variables in the Arctic has a higher uncertainty than in other regions (Previdi *et al.*, 2020), which highlights the importance of investigating the processes related to climate change in the Arctic. To accurately model the Arctic climate, it is critical to have an accurate estimate of key meteorological conditions; however, this is a challenge due to the lack of long-term in-situ measurements (Jung *et al.*, 2016; Tjernstrom & Graversen, 2009). The use of reanalysis products is one alternative to supplement this lack of field observations. Reanalysis products are the output of weather forecasting systems that use a combination of observational data and numerical models to produce a best guess of the atmospheric state at high temporal and geospatial resolution. Reanalysis models assimilate a variety of observational sources into their forecasts such as remote sensing (satellite and aircraft) and in-situ data (ground-based weather stations, buoys, and radiosondes). Reanalysis models are run retrospectively to produce long-term records of atmospheric conditions. Thus, reanalysis models provide a consistent dataset that can be used for a wide range of applications, from climate research to supplementing in-situ observations.

The reliability and performance of reanalysis models need to be assessed. Current reanalysis models are “ensemble” models which means that they produce several versions of predictions, called members, by varying different sets of initial boundary conditions. The only indication of uncertainty that is given by the ensemble model simulations is the variance of the predictions associated with each member of the model (Hersbach *et al.*, 2020). This uncertainty can therefore be considered as related to the precision of the model, that is, the dispersion of its results. However, it does not give any information on the accuracy of the predictions, that is, whether the predictions agree with the ground truth. Therefore, an indicator of the accuracy of the model output is needed and one method of evaluating the accuracy of reanalysis models is to compare them against in-situ observations. The most common source of in-situ observations is ground-based weather stations (Lappalainen *et al.*, 2016; Petäjä *et al.*, 2020). Other sources of in-situ observations exist (e.g., drifting buoy stations, tethered balloons, radiosondes, aircraft, and cruise expeditions), but few datasets are publicly available, the available period is often very short (due to measurements being collected during dedicated campaigns), and the

measurement frequency is sporadic (Shupe *et al.*, 2022; Wendisch *et al.*, 2019). There are also fewer observations in the Arctic compared to the rest of the globe (Avila-Diaz *et al.*, 2021; Huang *et al.*, 2017; Lindsay *et al.*, 2014), which is due to the very low population density, the challenging climatic conditions, and the remoteness which makes maintaining in-situ measurements more difficult.

In the Arctic, several previous studies have compared European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA5) data against in-situ observations (Delhasse *et al.*, 2020; Demchev *et al.*, 2020; Jakobson *et al.*, 2012; Seo *et al.*, 2020; Wang *et al.*, 2019; Yu *et al.*, 2021). Table 1 provides a summary of these studies. The nature of in-situ data used as a reference for the comparison and geographic locations is diverse. For the comparison of surface variables, Delhasse *et al.* (2020) and Wang *et al.* (2019) used six years (2010–2016) of data from automatic weather stations that provided hourly values of surface variables from Greenland and the Arctic Ocean, respectively. In addition, Yu *et al.* (2021) used observations from buoy stations to compare air temperature over the Arctic Ocean. Drifting polar stations were also used by Demchev *et al.* (2020) to evaluate the air temperature over sea ice. The results of these studies will be further discussed and compared to those of the present analysis in Section 3.1. Other studies focused on the comparison of vertical measurements in the Arctic, by using radiosondes that are sent through the troposphere during measurement campaigns (Graham *et al.*, 2019; Jakobson *et al.*, 2012).

In addition to comparing ERA5 data with in-situ observational data, previous studies have also performed an intercomparison of several different reanalysis models. For instance, the Japanese 55-year Re-Analysis (JRA-55), the Modern-Era Retrospective Analysis for Research and Applications-version 2 (MERRA-2), ERA-Interim (ERA-I), and the National Center for Environmental Prediction's (NCEP) Climate Forecasting System Reanalysis-version 2 (CFSv2) were compared to ERA5. Graham *et al.* (2019) stated that the ERA5 model performs best in the Arctic for land surface variables such as temperature, relative humidity, surface pressure, and wind speed/wind direction. Avila-Diaz *et al.* (2021) showed ERA5 to be one of the best reanalysis products for the North American Arctic (although it displayed worse performance for more northerly areas compared to southerly ones) and showed a better performance for temperature compared to precipitation. Isaksen *et al.* (2022) showed ERA5 is able to better reproduce temperature trends in the Arctic compared to the Copernicus Arctic Regional ReAnalysis (CARRA). Seo *et al.* (2020) and Wang *et al.* (2019) showed that the best performance for surface short-wave radiation in the Arctic is obtained by ERA5.

TABLE 1 Overview of previous studies comparing ERA5 to in-situ observations in the Arctic, with information on the data, study period, and variables analyzed.

Reference	Data	Study period	Variables
Delhasse <i>et al.</i> (2020)	ERA5, ERA-Interim, MARS 25 PROMICE AWS stations	2010–2016	Temperature Wind speed/direction SWD & LWD
Cai <i>et al.</i> (2021)	22 models from CMIP6 ERA5	1979–2014	Temperature
Demchev <i>et al.</i> (2020)	ERA5 and ERA-Interim Ground station from Cape Baranov Automatic buoy stations	2007–2013	Temperature
Seo <i>et al.</i> (2020)	5 reanalysis models 6 ground stations from BSRN Satellite radiation estimation	No common period (at least 5 years)	SWD & LWD
Graham <i>et al.</i> (2019)	5 reanalysis models 27 radiosondes	Summer of 2017	Temperature Relative humidity Wind speed/direction SWD & LWD
Wang <i>et al.</i> (2019)	ERA5 13 automatic buoy stations	2010–2016	Temperature Snowfall Precipitation
Yu <i>et al.</i> (2021)	ERA5 and ERA-Interim 234 automatic buoy stations	2010–2020	Temperature
Wang <i>et al.</i> (2021)	ERA5 and other reanalysis models BSRN network	2006–2017	SW & LW (Downwelling and net)
Avila-Diaz <i>et al.</i> (2021)	5 reanalysis models DAYMET observations	2000–2016	Temperature Precipitation
Betts <i>et al.</i> (2019)	ERA5 and ERA-I BSRN network	1979–2011	Temperature Wind speed SWD & LWD Precipitation
Isaksen <i>et al.</i> (2022)	ERA5 and CARRA AWS	1981–2020	Temperature
Költzow <i>et al.</i> (2022)	ERA5 and CARRA SYNOP stations	1998–2020	Temperature Wind speed
Sheridan <i>et al.</i> (2020)	ERA5 and ERA-I 230 stations in North America	1979–2016	Temperature

Abbreviations: AWS, automatic weather station; BSRN, baseline surface radiation network; CARRA, Copernicus Arctic Regional ReAnalysis; CMIP6, Coupled Model Intercomparison Project Phase 6; DAYMET, Daily Surface Weather and Climatological Summaries; ERA, European Centre for Medium-Range Weather Forecasts Reanalysis; LWD, long-wave downward; MARS, Meteorological Archival and Retrieval System; PROMICE, Programme for Monitoring of the Greenland Ice Sheet; SWD, short-wave downward; SYNOP, surface synoptic observations.

Betts *et al.* (2019) highlighted the improvement of ERA5 compared to ERA-I and how biases in temperature, wind speed, short-wave/long-wave downwelling radiation are small in the sub-Arctic. Wu *et al.* (2023) compared reanalysis models for cloud cover predictions and show that ERA5 has the best performance. Sheridan *et al.* (2020) showed ERA5 was the best reanalysis at predicting extreme temperature events with cold events being identified at a higher rate than warm events. In addition, the performance of ERA5 has been evaluated in other regions of the world outside of the Arctic (Babar *et al.*, 2019; Huai *et al.*, 2021; Molina *et al.*, 2021; Tetzner *et al.*, 2019; Velikou *et al.*, 2022) – see the Supporting information Section 1 and Table S1 for an overview of these studies.

Although the above studies have partially compared the surface variables produced by ERA5, there are several gaps in the literature that the present work aims to fill. In the Arctic (and elsewhere on the globe), temperature is the surface variable most often used to evaluate the performance of ERA5, although mainly using buoys, radiosondes, and a limited number of ground-based stations (Table 1). Here we evaluate ground-based temperature measurements from all regions of the continental Arctic over a long period, which will help reveal any spatial bias regarding surface geography. Surface wind speed has been validated only on Greenland by Delhasse *et al.* (2020), and a comparison over the entire continental Arctic is therefore missing. Accurate estimation of wind speed has

implications for modeling wind-driven processes (i.e., sea spray, blowing snow, and high-latitude dust generation), planning renewable energy strategies, and the risk assessment of extreme weather events. Short-wave downward (SWD) radiation flux has been studied by Seo *et al.* (2020) and Wang *et al.* (2019) with a monthly and daily resolution, respectively; however, a comparison at a higher temporal resolution of one hour is missing. This would assess how well the small-scale time variations are captured by ERA5 as well as accurate estimations of the diurnal variability and surface energy budget for solar radiation. A comparison of RH and surface pressure against ERA5 would be one of the first comparison studies performed in the continental Arctic, thus adding insight into the performance of the RH estimation (as the estimation of RH from ERA5 data has not been tested before in the Arctic) and synoptic-scale conditions, respectively. The accurate estimation of these variables has implications for modeling AA, atmospheric circulation, feedback mechanisms, and extreme weather events. Overall, the existing literature has not yet produced a comparison of ERA5 against observations for a large temporal period, for a high temporal resolution, and with a comprehensive spatial representation of the continental Arctic, which is one of the main goals of this study.

The objectives of this study are (i) to perform a comparison between ERA5 and in-situ observations using hourly data at the surface during all seasons; (ii) to compare ERA5 with observations on a seasonal basis; and (iii) to evaluate the effects of temporal resolution using the arithmetic mean on the comparison. We used essential meteorological surface variables such as 2-m air temperature (temperature), relative humidity (RH), surface pressure, wind speed, u (zonal) and v (meridional) wind components, and SWD radiation flux. This validation is based on in-situ data from available ground-based weather stations distributed throughout the continental Arctic. The data have been averaged (arithmetic mean) to the high temporal resolution of one hour and cover a minimum period of 10 years (except for Tiksi and TAS_A which only have six years of data available). The large amount of high time resolution data used, over a long period and from geographically dispersed regions of the continental Arctic, reinforces the robustness of this analysis, which makes the comparison unparalleled compared to past works. Therefore, this study will help expand the existing knowledge on the performance of ERA5 in the Arctic, at multiple temporal resolutions and in particular seasons. This will allow the proper confidence level in essential meteorological variables over different regions of the Arctic to be evaluated, at different time resolutions, and different seasons.

2 | METHODS AND DATASETS

2.1 | ERA5 reanalysis model

ERA5 is the fifth-generation reanalysis model provided by the Copernicus Climate Change Service (C3S). The C3S is one of the six information services provided by the Copernicus Earth Observation Program. It is implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF), on behalf of the European Commission. The data have been retrieved from the C3S Climate Data Store (CDS), using the *cdsapi* Python package (v0.5.1). ERA5 is based on forecasts from the 2016 version of the Integrated Forecast System (IFS), version CY41r2. The ERA5 model replaced the ERA-Interim model which was based on the 2006 IFS version. ERA5 is available since 1940 and it is constantly extended forward up to near real-time and is based on two forecasts made each day at 0600 and 1800 UTC. A wide variety of observation sources is considered in ERA5: satellite data, in-situ measurements, radar, drifting buoys, radiosonde, and aircraft data (ECMWF, 2016). Data on single levels at hourly temporal resolution (for every third hour, e.g., 0000, 0300, 0600, 1200, etc.) and $0.5 \times 0.5^\circ$ spatial resolution were used in this study. ERA5 is one of the most widely used and best-performing reanalysis models (Avila-Diaz *et al.*, 2021; Graham *et al.*, 2019; Hersbach *et al.*, 2020; Isaksen *et al.*, 2022; Sheridan *et al.*, 2020), therefore it was selected for this study.

2.2 | Meteorological variables

2.2.1 | Temperature

This variable corresponds to the air temperature at 2 m above the land surface, hereafter referred to as “temperature”. ERA5 computes this variable by interpolating between the lowest model level and the Earth’s surface, considering the atmospheric conditions, to produce a value at 2 m height. This variable was originally expressed in Kelvin [K] and converted into degrees Celsius [$^\circ\text{C}$].

2.2.2 | Relative humidity

ERA5 does not directly provide the relative humidity at the surface. It is only available at the pressure levels of ERA5 and the lowest pressure level is 1000 hPa which approximately corresponds to 100 m. Therefore, the altitudinal discrepancy between the lowest pressure level and ground-based weather stations was deemed too large to

use RH from the lowest pressure level in this study. The RH was therefore approximated using the saturation vapor pressure over water from Equation (1):

$$\text{RH} = 100 \times \frac{\text{es}(T_d)}{\text{es}(T)}, \quad (1)$$

where es is the saturation vapor pressure over pure water, T is 2-m air temperature and T_d is 2-m dew-point temperature. The August–Roche–Magnus formula, Equation (2), (Alduchov & Eskridge, 1996) provides an approximation for saturation vapor pressure over water:

$$\text{es}(T) = 6.1064 \times \exp\left(\frac{17.625 \times T}{T + 243.04}\right). \quad (2)$$

Therefore, the RH can be approximated using Equation (3):

$$\text{RH} = 100 \times \frac{\text{es}(T_d)}{\text{es}(T)} = 100 \times \frac{\exp\left(\frac{17.625 \times T_d}{T_d + 243.04}\right)}{\exp\left(\frac{17.625 \times T}{T + 243.04}\right)}. \quad (3)$$

2.2.3 | Surface pressure

The surface pressure in ERA5 is the measure of the weight of all air contained in a vertical column above the Earth's surface in units of pascals [Pa] and was converted to hectopascals [hPa], for comparison to in-situ observations.

2.2.4 | Surface short-wave downwelling radiation flux

This variable, termed short-wave solar radiation downwards in ERA5, is the amount of short-wave downwelling solar radiation that reaches a horizontal plane at the Earth's surface, which comprises both direct and diffuse radiation. This variable is the model equivalent of what would be measured by a pyranometer with a radiation spectrum of 0.2–4 μm (Hogan, 2015). It is originally provided by ERA5 as an accumulated value in [$\text{J} \cdot \text{m}^{-2}$] but has been converted to [$\text{W} \cdot \text{m}^{-2}$] by dividing the original values by 3600 (seconds in one hour).

2.2.5 | Wind speed

ERA5 provides two wind components, both in units of $\text{m} \cdot \text{s}^{-1}$: u (easterly or zonal) and v (northerly or meridional). We use the terms zonal and meridional to describe the u and v components, respectively. These two components correspond to the air speed moving toward this direction, at a height of 10 m above the Earth's surface. The overall

wind speed (ws) is derived from Equation (4):

$$ws = \sqrt{u^2 + v^2}. \quad (4)$$

In-situ observations of wind speed and wind direction (wd) were converted into the u and v components according to Equations (5) and (6):

$$u = -ws \times \cos(wd), \quad (5)$$

$$v = -ws \times \sin(wd). \quad (6)$$

The wind direction was converted to radians from degrees before calculating the u and v components. The overall wind speed is referred to as “wind speed” while the two components (u and v) are referred to as “zonal wind speed” and “meridional wind speed”, respectively.

2.3 | Method to sample ERA5 data for specific observatories

Gridded ERA5 reanalysis data must be extracted for individual grid cells for each of the station's locations. Past studies (Delhasse *et al.*, 2020; Demchev *et al.*, 2020; Tetzner *et al.*, 2019; Wang *et al.*, 2019) have used different methods of extracting the model data for a specific location, which involved two main methods: (I) selecting the closest grid cell (in Euclidean distance) to the location of the station and (II) selecting the grid cell that contains the location of the station.

To investigate how the above-mentioned methods influenced the results of the comparison, a sensitivity analysis was performed. This analysis consisted of comparing the results using ERA5 data extracted with the two different methods against in-situ data. The difference between the average R^2 and regression slope metrics for the two extraction methods was then calculated. Table S2 shows the differences for each metric averaged (arithmetic mean) over all stations. It shows that neither of the two methods gives systematically better results. Given the resolution of the model data (0.5°), the first method (selecting the closest grid cell [in Euclidean distance] to the location of the station) was selected to extract the data from ERA5.

2.4 | In-situ measurements

2.4.1 | Selection of stations

In extreme environments such as the Arctic, meteorological stations are rare but they are a valuable source of information. In this study, in-situ measurements

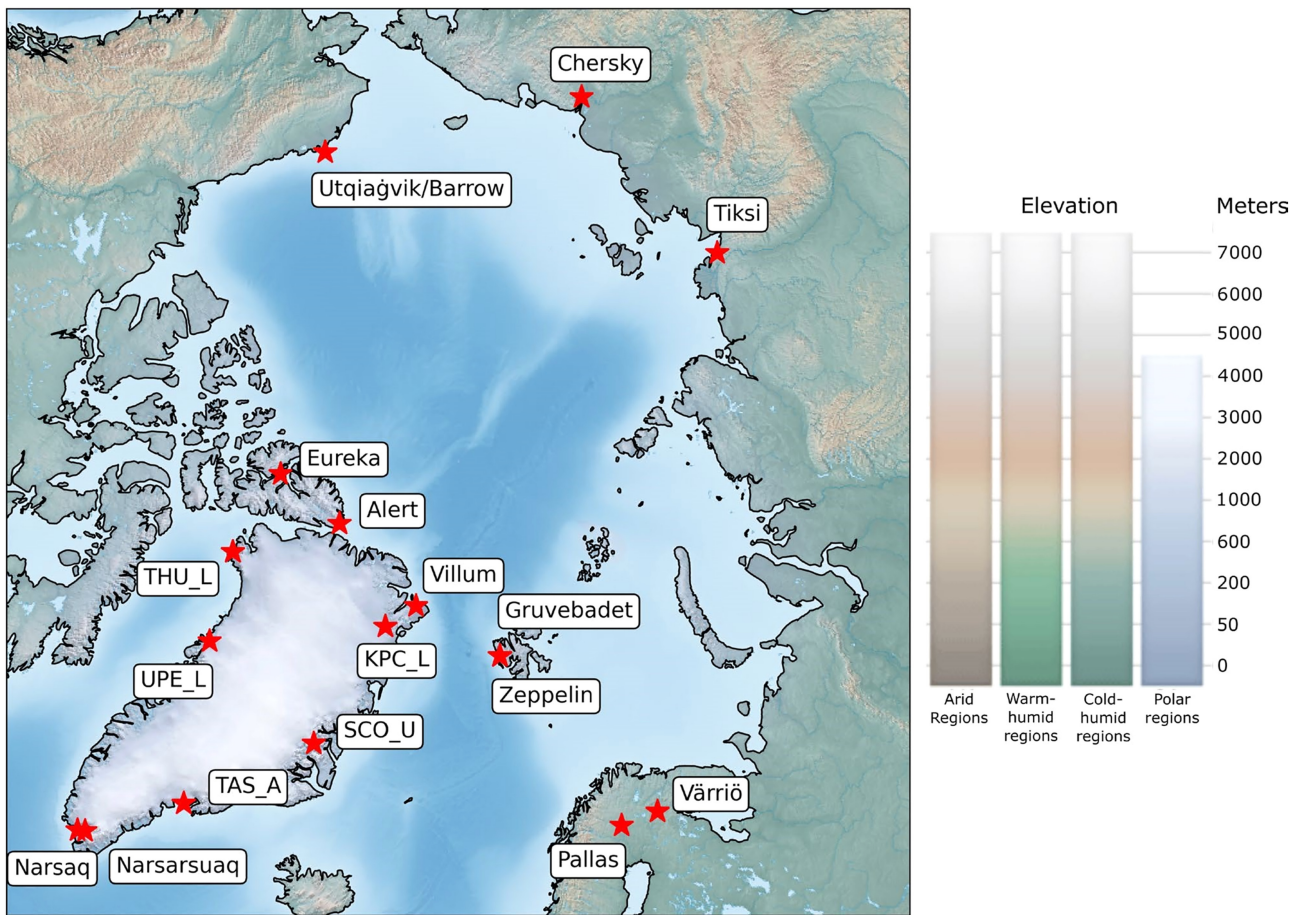


FIGURE 1 Map indicating the location of each station with a red star and the station's name specified in a box. The delineation of the land coasts and the background map come from Cross Blend Hypso and are used via the Cartopy module (v0.20.2) of Python Natural Earth (<https://scitools.org.uk/cartopy/docs/latest/>, last accessed 1 Jan 2024). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

will be used as ground truths for the comparison to ERA5. We selected ground-based meteorological stations dispersed throughout the continental Arctic to obtain as wide-ranging spatial coverage as possible. To have enough data for a reliable statistical comparison, data from stations with records covering several years were prioritized. However, there are unfortunately fewer stations in the eastern Arctic as well as in the North American region compared to the European sector. The location of each station used in this study is displayed in Figure 1. Table 2 summarizes the properties of the selected in-situ weather stations. All data were extracted from publicly available on-line repositories. The references of each dataset are available in the Appendix.

Sets of stations in proximity to each other but differentiated by another metric (distance to coast or elevation) were also selected, such as Narsaq/Narsarsuaq, Gruvebadet/Zeppelin, and Pallas/Värriö, to investigate the potential influence of these topographic differences. Narsaq/Narsarsuaq are separated by ca 44 km and are both located within a complex fjord system, they differ in terms

of elevation and distance to the coast (Table 2). Gruvebadet and Zeppelin are separated by less than 2 km but differ substantially in terms of elevation, with Gruvebadet close to sea level and Zeppelin on a nearby mountain ridge. Pallas and Värriö are both inland, separated by ca 230 km, and are at several hundred meters elevation (Table 2). Five stations of the Programme for Monitoring of the Greenland Ice Sheet (PROMICE) network, which were located on the coast and geographically dispersed around Greenland, were selected (Fausto *et al.*, 2021). Pyranometers in the PROMICE network have a wavelength spectrum of 0.31–2.8 μm .

Coastal regions around Greenland are topographically complex and are hence potentially challenging to simulate by ERA5 (Køltzow *et al.*, 2022). At the same time, Greenlandic coastal regions are undergoing a drastic change (Coulson *et al.*, 2022; Straneo *et al.*, 2022), and hence long-term climate variable information is critical. Radiation data were collected from pyranometers in the baseline surface radiation network (BSRN), which have a wavelength spectrum of 0.25–3 μm (Driemel *et al.*, 2018).

TABLE 2 Characteristics of the ground stations selected for the comparison analysis.

Station	Period	Temporal resolution	Latitude [°]	Longitude [°]	Altitude [m]	Distance to the coast [m]
Alert	01/01/2004–31/12/2020	1 min	82.47	−62.5	210	8000
Utqiagvik/Barrow	01/01/1974–31/12/2022	1 hr	71.32	−156.61	7	634
Chersky	01/01/1955–31/12/2018	1 day	68.74	161.4	8	99,000
Eureka	01/01/2007–31/12/2019	1 min	79.98	−85.95	83	0
Pallas	01/01/2010–31/12/2022	1 min	67.97	24.12	565	322,000
Kronprins Christian Land (Lower) (KPC_L)	01/01/2008–31.12/2022	1 hr	79.91	−24.08	370	130,000
Tasiilaq (Averaged) (TAS_A)	01/01/2008–31/12/2022	1 hr	65.78	−38.90	890	30,000
Thule (Lower) (THU_L)	01/01/2008–31/12/2022	1 hr	76.40	−68.27	570	23,000
Upernavik (Lower) (UPE_L)	01/01/2008–31/12/2022	1 hr	72.89	−54.30	220	4500
Scoresbysund (Upper) (SCO_U)	01/01/2008–31/12/2022	1 hr	72.39	−27.23	970	117,000
Tiksi	01/08/2010–01/10/2018	1 hr	71.60	128.88	29	650
Villum	01/01/2002–01/02/2021	1 hr	81.58	−16.64	30	2600
Grubebadet	01/10/2009–01/10/2022	30 min	78.92	11.87	67	890
Zeppelin	01/01/1998–31/12/2021	1 hr	78.91	11.89	474	2500
Värrö	01/01/2010–31/12/2021	1 hr	67.77	29.58	276	230,000
Narsarsuaq	01/01/1961–31/12/2020	3 hr	61.17	−45.42	30	1800
Narsaq	01/01/1958–31/12/2020	3 hr	60.91	−46.05	0	200

Note: For the PROMICE stations, “Lower” means that the station is near the ice sheet margin while “Average” means that the station is close to the equilibrium line altitude (Fausto *et al.*, 2021). Altitude references elevation above sea level. The distance to the coast was estimated for each station from Google Earth (<https://earth.google.com/>, last accessed 1 February 2023); while not exact this method provides an approximate distance to evaluate between stations that can be used for a relative comparison between stations.

2.4.2 | Meteorological variables

Using in-situ measurements from ground-based weather stations restricts the comparison to the surface variables, which are continuously available at high time resolution and over several years. Unfortunately, vertically resolved measurements are rarely available (often performed during campaigns) and only for short durations (Tjernstrom & Graversen, 2009). Considering the availability of in-situ data, the analysis is focused on the following essential meteorological variables: temperature, RH, surface air pressure, wind speed, meridional wind speed, zonal wind speed, and SWD radiation flux.

2.4.3 | Pre-processing steps

Several straightforward preprocessing steps were performed before analysis. The majority of the observational data has already undergone a pre-processing and quality control process from the data originators, but additional data processing steps were applied to obtain a

homogeneous dataset across all stations. The steps used in this study are presented in Supporting information Section 2, and it is important to note that values were converted to appropriate units or removed but no data were manipulated, altered, or substituted otherwise.

2.4.4 | Final data availability

After applying the pre-processing steps, the final, cleaned dataset is composed of hourly values from 17 stations, for seven variables corresponding to ca 16 million data points. Figure S1 shows the data availability (in % of data available over the considered period) for each variable and all the stations on a monthly and yearly basis. Only two stations have less than seven years of data (e.g., Tiksi and TAS_A), but were included because they cover the eastern Arctic and the southeast coast of Greenland, respectively. Not all stations have data for the seven variables and only temperature and RH are available for all the stations. Wind speed and surface pressure are measured at most stations while SWD radiation flux is measured by about half. The

left panel of Figure S1 shows the period for which data are available and the right panel of Figure S1 shows that the monthly availability, the percentage of data available in a month calculated over all available years, can have variations of about 10% depending on the month considered. This underlines the fact that the availability of data is not constant over time, but there is no specific pattern across seasons. Generally, it is important to keep in mind that the maintenance of instruments in extreme conditions such as those of the Arctic is challenging. Indeed, the instruments can suffer interruptions in the power supply, calibration problems, or deterioration due to meteorological conditions.

2.5 | Comparison methodology

The comparison between ERA5 and in-situ data was performed over the entire available period. As mentioned before, ERA5 produces hourly data but only hourly values every three hours were extracted, that is, eight data points per day. The comparison is performed by selecting the time stamps of the observation data that match the ERA5 time stamps after temporal aggregation. To perform the comparison between ERA5 and observations, it is necessary to define several metrics to quantitatively evaluate performance. The metrics used in this study are similar to those reported by previous studies comparing observations and models (Mistry *et al.*, 2022; Tetzner *et al.*, 2019; Zhao & He, 2022), namely the slope of the regression line (slope), coefficient of determination (R^2), and RMSE. The slope represents how the dependent variable changes with the independent variable and indicates if ERA5 is overestimating (slope > 1) or underestimating (slope < 1) observations. The R^2 value quantifies the agreement between ERA5 and observations with 1 indicating a perfect agreement and 0 indicating no relation between ERA5 and observations. The RMSE is the standard deviation of the residuals and represents the average distance between the regression line and the individual data points. The RMSE can range from 0 (perfect agreement) and positive infinity, with lower values indicating a better fit between ERA5 and observations – see Supporting information Section 3 for more details about each metric.

3 | RESULTS AND DISCUSSION

3.1 | Overall comparison of ERA5 against in-situ measurements

An overview of the comparison results of ERA5 and observations using the slope, R^2 , and RMSE is displayed in

Figure 2a–c, respectively. In each panel, the cells are color-coded according to the metric used, and cells with a darker color (black) indicate a better agreement between modeled values and observations. Overall, the variables with the best agreement are temperature, surface pressure, and SWD radiation flux (Figure 2). The variables with the worst agreement are wind speed including meridional and zonal wind speed components. Relative humidity shows a poor agreement with large variability depending on location. To investigate any spatial dependency of these results, Figure 3 shows the geographical location of each station, where the marker color indicates the slope and the marker size the R^2 . From Figure 3 it is evident that there is little to no spatial dependency for any particular region of the Arctic regarding the predictive power of ERA5. Therefore, the random spatial variations are probably due to local phenomena such as topographical features that are not resolved by ERA5. The following subsections focus on the results for each variable separately.

3.1.1 | Temperature

Temperature exhibits an excellent agreement between ERA5 and observations (Figure 2). The mean (standard deviation, σ) R^2 , slope, and RMSE across all stations is 0.941 (0.024), 1.012 (0.113), and 3.496°C (1.363°C), respectively. Eureka and Värriö are among the stations where the temperature is best reproduced, while stations such as THU_L or SCO_L display the worst results. While no particular spatial pattern can be discerned from Figure 3, certain stations in proximity to each other (e.g., Narsaq/Narsarsuaq, Värriö/Pallas, Villum/KPC_L) can have different results. Narsaq and Narsarsuaq are located near the coast on the southern tip of Greenland, in complex topography, but with varying distances to the inland ice, which could explain the discrepancies as the topography becomes more complex closer to the inland ice and the influence of katabatic winds from glaciers increases. Värriö and Pallas are far from the coast, surrounded by boreal forests, and are located at elevated altitudes, with Pallas at approximately twice the elevation of Värriö (Table 2), which could be responsible for the difference in performance as the difference between the actual and model elevation becomes larger with increasing altitude. For Villum and KPC_L, the combination of differences in elevation and distance to the coast might contribute to the differences in ERA5 performance. Interestingly, Villum and KPC_L have different lengths of available data, but there are no major differences between the performance of ERA5 between these two stations, indicating that record length is not affecting this comparison. Therefore, the discrepancies for these two sets of closely located stations might be explained by

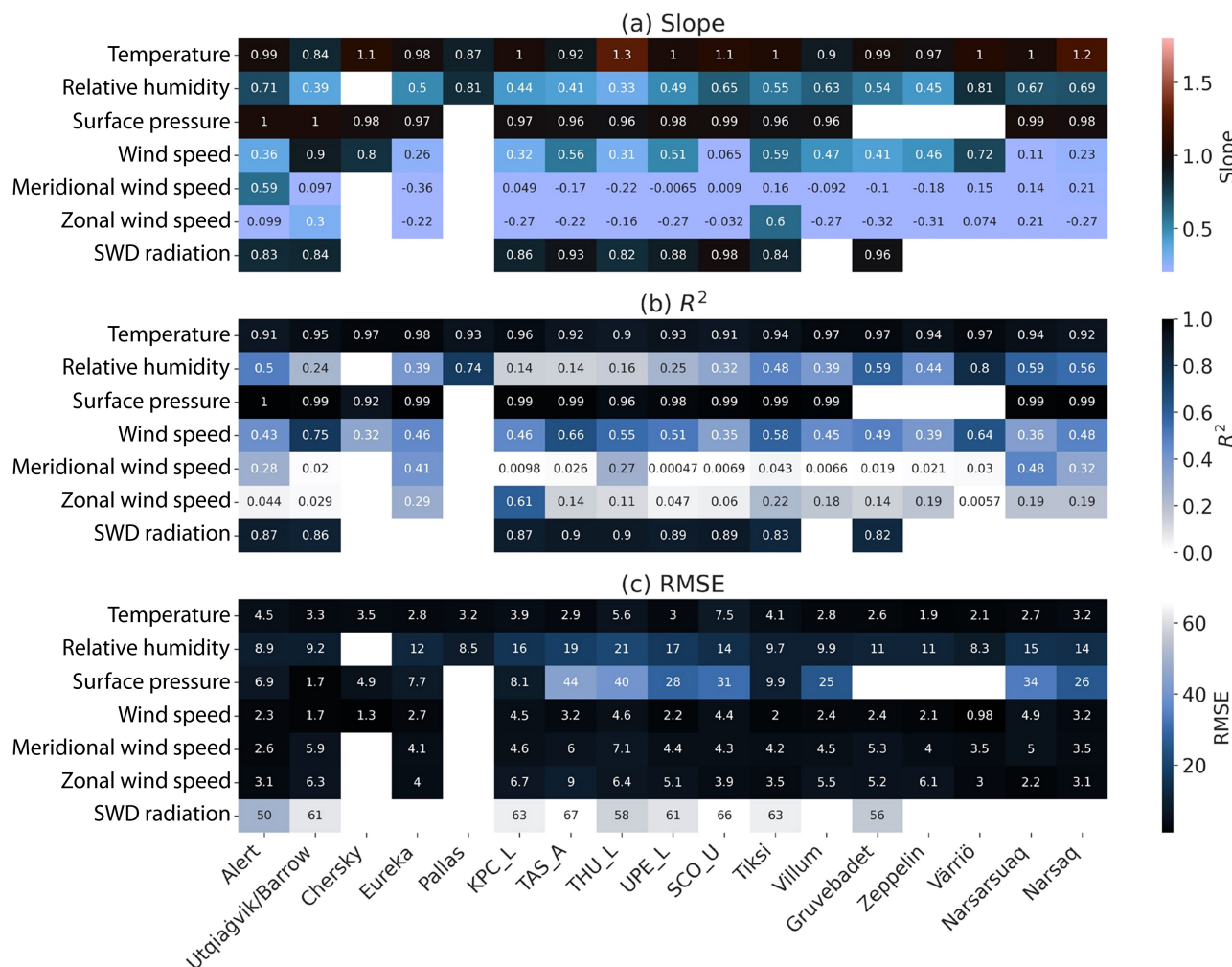


FIGURE 2 Heat map of the three metrics (a) slope, (b) R^2 , (c) root-mean-squared error (RMSE), resulting from the comparison between European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA5) data and in-situ measurements. Each variable is indicated according to rows and each station according to columns. The metrics are indicated in each cell and also represented with a color bar. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

local topographical features, inland ice, and marine influence (Narsaq/Narsarsuaq) as well as elevation differences (Värriö/Pallas) or a combination of both (Villum/KPC_L).

Stations located in complex topography likely have an altitudinal difference between the station and the corresponding ERA5 grid cell, which can lead to a larger bias (larger RMSE). Sheridan *et al.* (2020) also observed a similar pattern for extreme temperature events in North America: areas with complex terrain displayed a greater discrepancy between ERA5 and observations. These biases could be corrected by applying a dry adiabatic lapse rate, but it has been shown in a similar study in Antarctica (Tetzner *et al.*, 2019) that this correction did not improve the results, partially because of temperature inversion situations, where the dry adiabatic lapse rate is not applicable. However, Bracegirdle and Marshall (2012) observed an improvement in the biases between ERA5

and observations in Antarctica when correcting for the height difference between observations and the grid cell elevation. Koltzow *et al.* (2022) showed a height adjustment lower the mean absolute error when comparing temperature from ERA5 against observations, which had a greater effect in summer compared to winter. Given these opposing outcomes, incorporating a height-adjustment correction factor (such as an adiabatic lapse rate) should be used with caution. Other factors influencing the agreement could include topographic complexity, heat build-up, variable surface albedos, and the presence of vegetation.

The results found in this study are consistent with previous work in the Arctic. Delhasse *et al.* (2020) performed a similar analysis in Greenland for 25 stations in the PROMICE network. They used a period of six years (2010–2016) with daily temporal resolution and found a mean correlation coefficient of 0.97 between ERA5 and

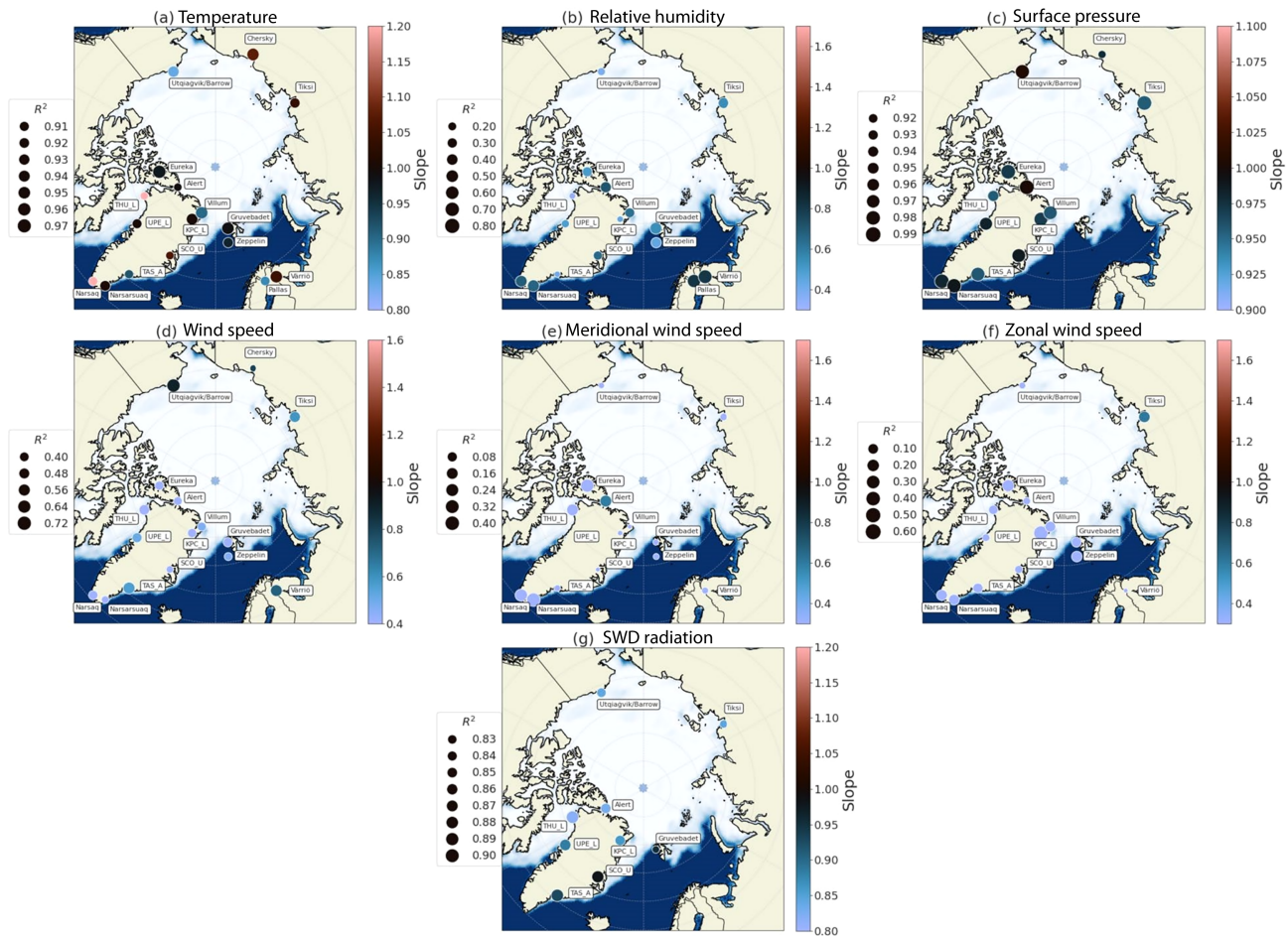


FIGURE 3 Spatial representation of the model–measurement comparison for (a) temperature; (b) relative humidity (RH); (c) surface pressure; (d) wind speed; (e) meridional wind speed; (f) zonal wind speed; and (g) short-wave downward (SWD) radiation. Each circle represents the location of the station where the measurements were collected. R^2 scores are represented using the marker size and the slopes of the regression lines using a color bar. Note that the location of Zeppelin and Narsarsuaq has been modified for clarity; these stations are located in close proximity to Gruevbadet and Narsaq, respectively. The maximum sea ice coverage for March 2020 is represented. These data are produced by the National Snow and Ice Data Center (NSIDC) (<https://nsidc.org/home>, last access: 1 February 2023). The delineation of the land coasts and the background map come from Natural Earth and are used via the Cartopy module of Python Natural Earth. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

observations. This score is identical to the one found in this study for all the stations with hourly resolution (0.97, Supporting information). When only considering the summer period, this mean correlation coefficient decreases to 0.85 (Delhasse *et al.*, 2020). Demchev *et al.* (2020) analyzed temperature over the Central Arctic Ocean and Siberian coast for the period 2007–2013. They used data from buoy observations (drifting with the sea ice) and coastal meteorological stations to compare against ERA5. Considering the buoy data and all seasons, they found correlation coefficients ranging from 0.77 to 0.97. They also compared data from ground stations in the eastern Arctic (Cape Baranov and Russian meteorological stations), which show similar results to those found in this study with a mean R^2 of 0.93 (ranging from 0.88 to 0.95) and a mean bias of 0.56°C (ranging from 0.27 to 1.07°C). These results suggest that

the performance of ERA5 for surface temperature seems to be comparatively worse over sea ice relative to observations over land, although this should be confirmed by more studies using temperature data over the Arctic Ocean. Research expeditions covering multiple seasons and large geographic areas such as the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) expedition would be a suitable source of observations for future comparisons (Herrmannsdörfer *et al.*, 2023; Shupe *et al.*, 2022; Solomon *et al.*, 2023).

Comparing ERA5 performance with regions outside of the Arctic (see Supporting information Section 1 and Table S1 for more details), it appears that ERA5 performs slightly worse for the polar regions (Huai *et al.*, 2021; Tetzner *et al.*, 2019; Velikou *et al.*, 2022; Zhu *et al.*, 2021). This difference is likely due to the difference in the

density of observations, with lower coverage of ground-based weather stations in the polar regions compared to other regions and limitations regarding the use of satellite products, which is the main source of assimilated data in ERA5 (Hersbach *et al.*, 2020) as well as the increased uncertainty associated with numerical weather prediction (NWP) modeling in the polar regions (Previdi *et al.*, 2020; Solomon *et al.*, 2023).

3.1.2 | Relative humidity

The comparison of ERA5 against observations of RH produces a wide range of results (Figures 2 and 3). Overall, the results indicate RH is underpredicted, poorly modeled, and shows a large variance. The mean (σ) slope, R^2 , and RMSE across all stations are 0.566 (0.148), 0.421 (0.205), and 12.787% (3.996%), respectively. The slopes range between 0.334 (THU_L) and 0.814 (Pallas) while the R^2 range between 0.14 (TAS_A) and 0.802 (Värriö). Figure 3 shows that there is high spatial variability with no apparent pattern. It is also important to remember that ERA5 does not provide RH directly but it is estimated by empirical equations (Equations 2 and 3), using air and dew-point temperature at 2 m. The constants used in the equation have been calibrated using standard conditions for liquid water and might not have been adapted to the extreme temperature conditions encountered in the Arctic. This could contribute to the poor agreement between model and observation.

Interestingly, the two stations with the best results, Pallas and Värriö, are located close to each other, within the boreal forest, far from the coast, and at medium elevation (Table 2). Conversely, two other stations that are very close to each other, Villum and KPC_L, have very different performances. Villum is located at sea level and near the coast, while KPC_L is located several hundred kilometers inland and at ca 370 m elevation. This suggests that the measurement of RH is subject to important local effects, such as the distance to the coast and elevation, similar to temperature (which is logical as RH is calculated using temperature). Two other stations that exhibit different evaluation metrics but are located in proximity to each other are Gruvebadet and Zeppelin (Figure 3). Gruvebadet is located near sea level and closer to the coast while Zeppelin is also near the coast but at ca 474 m elevation (Table 2). Gruvebadet has better metrics for RH compared to Zeppelin when considering the slope and R^2 although RMSE values are similar (Figure 2).

For stations located near the coast, local effects can be the melting of sea ice during the summer, which increases the water vapor content due to increased evaporation (Kopec *et al.*, 2016), or the presence of thick multiyear

ice which acts as a cap on the transfer of water vapor during winter. For stations at elevation, the estimations of Equations (2) and (3) might not capture the vertical gradients of RH in the Arctic atmosphere. Stations located in continental areas but also at elevation (e.g., Pallas and Värriö) show better performance, which indicates the distance to the coast might have a greater influence on the prediction of RH than elevation does. Accurate predictions of RH can have implications for the modeling of cloud radiative effects (Cox *et al.*, 2015), thus shortcomings in the prediction of RH at the surface and the vertical dimension (Jakobson *et al.*, 2012) must be considered when evaluating climate models which are nudged to ERA5 (or other reanalysis products).

To the authors' knowledge, there is no scientific study that validates the near-surface RH in the Arctic. This is probably due in part to the fact that this variable is not directly provided by ERA5. Graham *et al.* (2019) compared vertically resolved measurements of RH using radiosondes in the Fram Strait against ERA5 on pressure levels (a vertically resolved version of ERA5 using pressure as the vertical coordinate) and found a correlation coefficient of 0.83. In the rest of the world, it seems that there is only one study that performed the RH comparison. Zhang *et al.* (2021) compared several reanalysis products (including ERA5) against 746 meteorological stations in China for a long period (1978–2018). Zhang *et al.* (2021) used the same method as this analysis to compute RH values (Equations 2 and 3) but with monthly resolution. They found a mean correlation coefficient of 0.82 versus 0.63 in our analysis. Therefore, we can hypothesize that the estimations of Equations (2) and (3) are probably not well adapted to the extremely cold conditions in the Arctic, while it should be noted that we only used a small number of stations for comparison in this study. Testing this hypothesis in other extreme environments such as Antarctica or high-elevation sites could be beneficial for the improvement of RH estimations in the cold regions.

3.1.3 | Surface pressure

Surface pressure is the variable best reproduced by ERA5 (Figure 2). Indeed, the R^2 coefficients and the slopes of the regression are very close to 1, with mean (σ) values of 0.982 (0.021) and 0.976 (0.017), respectively. This is expected as surface pressure is directly assimilated into ERA5. This also highlights the great skill ERA5 has in capturing synoptic situations (i.e., atmospheric pressure and temperature variations). However, there are some systematic biases (represented by the RMSE) that can be high (up to 40 hPa), especially for the stations belonging to

the PROMICE network. These biases are likely related to the difference between the altitude of the station and the mean altitude of the corresponding ERA5 grid cell. For example, a very mountainous topography with abrupt elevation changes can lead to a large dispersion of the altitude among all the locations of the same grid cell, therefore introducing large biases in the model/measurement comparison. To the authors' knowledge, no ERA5 comparison studies have evaluated surface pressure in the Arctic. However, outside of the Arctic, Huai *et al.* (2021) (China) and Tetzner *et al.* (2019) (Antarctica) found mean R^2 values equal to 0.88 and 0.91, respectively, which are lower than the one found in this study (0.98, Supporting information). This is likely because in the Arctic pressure measurements from the majority of the stations have been assimilated, while both Huai *et al.* (2021) and Tetzner *et al.* (2019) used unassimilated data.

3.1.4 | Wind speed

The results of wind speed are presented first, then the individual components (meridional and zonal) are discussed. Wind speed is poorly reproduced by ERA5 with a mean (σ) R^2 value of 0.493 (0.120), which ranges from 0.351 (SCO_U) to 0.753 (Utqiagvik/Barrow). Wind speed is underestimated with a mean (σ) slope of 0.443 (0.234); SCO_U and Utqiagvik/Barrow also show the lowest and highest slopes, respectively. The mean (σ) RMSE is $2.80 \text{ m}\cdot\text{s}^{-1}$ ($1.22 \text{ m}\cdot\text{s}^{-1}$). Spatially, all of the stations on Greenland, the surrounding stations (Eureka, Alert, Gruebadet, and Zeppelin), and Tiksi exhibit poor results, while Utqiagvik/Barrow, Chersky, and Värriö exhibit comparatively better results (Figure 3). In terms of topography, Utqiagvik/Barrow, Tiksi, and Chersky are all located near sea level and in the Arctic tundra although only Utqiagvik/Barrow and Tiksi are located near the coast, while Chersky is located ca 100 km inland (Table 2). Värriö is located far from the coast, at several hundred meters elevation, and within the boreal forest. It is known that the measurement of wind at the surface is greatly influenced by topographical factors (Rotach *et al.*, 2015). For example, topographical structures such as mountain passes or narrow valleys favor a higher wind speed. The presence or absence of vegetation (i.e., forests) can also affect wind speed patterns (Wever, 2012). These topography-related meteorological phenomena could be responsible for the poor agreement for the stations around/on Greenland and the better performance of the stations located elsewhere.

Comparing the current study to previous ones performed in the Arctic, Delhasse *et al.* (2020) performed a comparison of surface wind speed observations to ERA5,

with a spatial resolution of 0.25° and at daily resolution for the 2010–2016 period for the PROMICE stations. They calculated a mean correlation coefficient of 0.85 and a mean RMSE of $2.18 \text{ m}\cdot\text{s}^{-1}$, which are consistently better than those found in the current study (mean [σ] correlation coefficient and RMSE of 0.698 [0.084] and $2.8 \text{ m}\cdot\text{s}^{-1}$ [$1.22 \text{ m}\cdot\text{s}^{-1}$], respectively). The main difference between Delhasse *et al.* (2020) and the current study is the different temporal resolutions (daily in Delhasse *et al.* (2020) vs hourly in this study), which will be discussed in Section 3.4, but also the spatial resolution used (0.25° in Delhasse *et al.*, 2020 vs 0.5° in this study). Betts *et al.* (2019) compared wind speed from ERA5 against observations in the North American Arctic from 1979 to 2011 using daily means and showed wind speed to be biased low (especially during daytime compared to nighttime) and this bias increased quasi-linearly with the magnitude of wind speed. In other regions of the world, compared to the current study, similar metric values for wind speed have been found in the literature when comparing ERA5 to observations. Molina *et al.* (2021) found Pearson correlation coefficients ranging from 0.6 to 0.85 for 245 ground stations in Europe, over the period 1979–2018. Huai *et al.* (2021) found an R^2 of 0.55 and an underestimation of wind speed for 19 AWS located in the Qilian Mountains, China. Tetzner *et al.* (2019) also found an underestimation of wind speed in Antarctica for ERA5, with a spatial resolution of 0.25° . Overall, the underestimation of wind speed by ERA5 appears to be globally ubiquitous, regardless of the spatial resolution used.

For the individual (meridional and zonal) wind components, the comparison yielded very poor results. The highest slope for meridional wind speed was found at Alert (0.59) while all other sites were below 0.21, with seven out of 15 sites displaying negative slopes. For zonal wind speed, the highest slope was found at Tiksi (0.6) while all other sites were below 0.3, with 10 out of 15 sites showing negative slopes. The RMSE values for meridional and zonal wind speed are similar to that of wind speed (Figure 2). The spatial distribution also no longer shows Utqiagvik/Barrow, Tiksi, Chersky, or Värriö having comparatively better results than other stations. Instead, the agreement is poor regardless of geographic location (Figure 3). Overall, the individual components of wind speed are not properly reproduced by ERA5. The authors are aware of the documentation for ERA5 urging caution when comparing the model output of wind speed to observations, given the nature of point measurements versus grid cell averages. This should be considered when interpreting these results. Here we have quantified the degree of caution that should be exercised when utilizing wind speed variables from ERA5.

3.2 | Short-wave downward radiation flux

The comparison revealed SWD radiation flux is generally underestimated but gives overall satisfactory results, with slopes and R^2 values all greater than 0.82 (Figure 2). The mean (σ) slope, R^2 , and RMSE for SWD radiation flux is 0.882 (0.059), 0.87 (0.029), and $60.416 \text{ W}\cdot\text{m}^{-2}$ ($5.263 \text{ W}\cdot\text{m}^{-2}$), respectively. Spatially, the High Arctic ($>70^\circ \text{ N}$) stations, Alert, THU_L, Utqiagvik/Barrow, and Tiksi, underestimate SWD radiation flux to a greater degree compared to stations located at more southerly latitudes (TAS_A and SCO_U). A notable exception to this pattern is Gruevbadet, which gives a good agreement with an R^2 and slope of 0.82 and 0.96, respectively. Furthermore, there appears to be a slight gradient regarding the comparison for the PROMICE stations, with stations in southern Greenland performing better than stations in northern Greenland (Figure 3). It should also be noted that the stations located on the Greenlandic continent belong to the PROMICE network while all other stations are in the BSRN network, which might also contribute to this spatial pattern as there are differences in the spectral range of instruments in these networks and ERA5 (see Sections 2.2 and 2.4.1).

Wang *et al.* (2021) performed a comparison of SWD radiation flux against ERA5 in the Arctic with several stations from the BSRN network, with an hourly resolution, 1° spatial resolution over the period 2006–2017. They found a mean R^2 of 0.84, which agrees with the mean R^2 produced by this study (0.87, Supporting information). Wang *et al.* (2021) and the current study used different stations in their comparisons and still produced similar results, highlighting the consistency of ERA5. Seo *et al.* (2020) compared SWD radiation flux from BSRN stations, with a monthly resolution, and for the period 2000–2018 (depending on the station). They found a mean R^2 of 0.87, which is identical to the one found for the current study (Supporting information). In addition, Delhasse *et al.* (2020) performed a comparison of SWD radiation flux for the PROMICE stations, with the daily resolution, and for the 2010–2016 period. Interestingly, Delhasse *et al.* (2020), while incorporating more stations, found a similar mean correlation coefficient (0.98) when compared to this study (0.93). The performance of SWD radiation flux in other regions of the world was also assessed. Wang *et al.* (2021) analyzed hourly values from four stations in Antarctica from the BSRN network against ERA5 and obtained a mean R^2 of 0.93. Overall, ERA5 underestimates SWD radiation flux regardless of location but gives a satisfactory agreement with observations.

3.3 | ERA5 seasonal performance

In Sections 3.1 and 3.3, we use all available data for each station in all seasons. In this section, we expand the comparison by separating data into the meteorological seasons and evaluating the performance of ERA5 on a seasonal basis. Here we decided to use the standard definition of the meteorological seasons as this facilitates comparison to previous studies and each season will have approximately the same number of data points. We refer to winter as December, January, and February; spring as March, April, and May; summer as June, July, and August; and autumn as September, October, and November. Figures 4 and 5 show the R^2 and slope values for each variable, respectively, for distributions (boxplots) and individual stations (markers). Each subpanel displays the distribution of all (All) values shown in Figure 2 as well as the distribution for each season (denoted by the first letters of each month in the respective season). Figure S2 shows the same analysis for the RMSE.

Temperature is underpredicted during winter and summer (Figure 4a), while during spring and autumn, the distribution of slopes is similar compared to using all available data. The distribution of R^2 values shows a larger variance and an overall decrease in the agreement comparatively during winter and summer (Figure 5a), while spring shows a comparatively similar distribution and autumn shows an increased variance and lower median value compared to using all available data. Interestingly, the distribution of RMSE values shows an increase during winter and a decrease during summer, compared to using all available data, while spring and autumn show a similar distribution compared to all data (Figure S2a). Previous studies have made similar observations. Delhasse *et al.* (2020) found a decrease in the mean correlation coefficient between ERA5 and observations during the summer (0.85) compared to all available data (0.97) in their study. Demchev *et al.* (2020) found mean correlation coefficients of 0.91 and 0.81 during the warm (May–October) and cold seasons (November–April), respectively, when comparing buoys and ERA5. Reanalysis products have been shown to be biased warm in the Arctic over sea ice, especially during winter (Beesley *et al.*, 2000; Graham *et al.*, 2017; Jakobson *et al.*, 2012; Tjernstrom & Graversen, 2009; Yu *et al.*, 2021), which has been tied to issues of properly predicting the stable boundary layers of the Arctic environment (Graham *et al.*, 2017; Kayser *et al.*, 2017; Tjernstrom & Graversen, 2009). Recently, bias correction models have been implemented to reduce these discrepancies with promising effects (Zampieri *et al.*, 2023). Implementing a multi-layer snow scheme into ECMWF IFS can also improve part of the warm bias

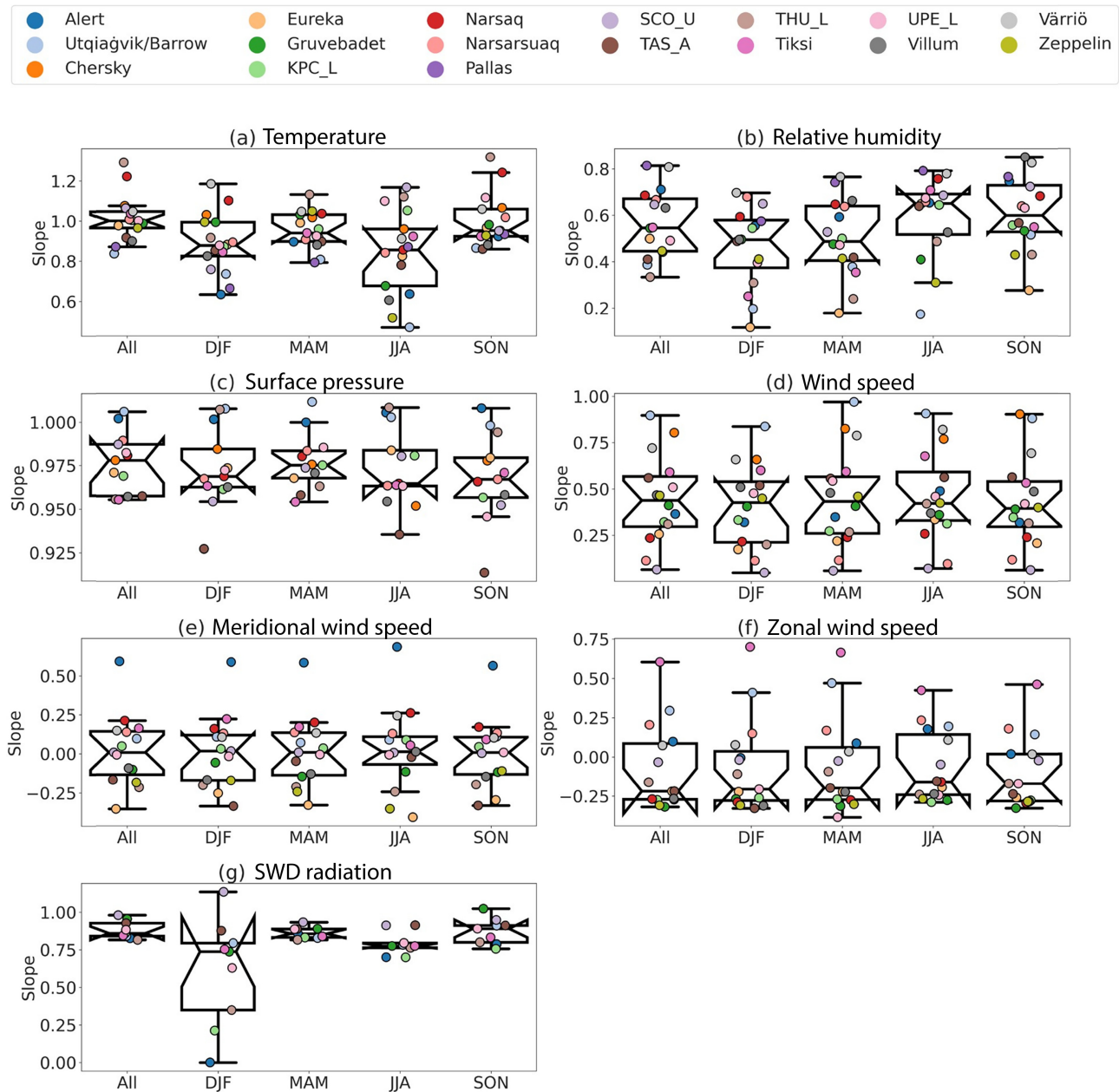


FIGURE 4 Slope values for all available data and each meteorological season for (a) temperature; (b) relative humidity; (c) surface pressure; (d) wind speed; (e) meridional wind speed; (f) zonal wind speed; and (g) short-wave downward (SWD) radiation. The markers show the individual stations while the box plot shows the distribution of these stations. The central line is the median and the two extremities of the boxes are the 25th and 75th percentiles, respectively. The whiskers extend to points that lie within 1.5 times the interquartile range. The notches on each box represent the 95% confidence level (CL) of each median. This allows us to see if the medians of two distributions are significantly different from each other if the notches do not overlap. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.2704)]

in ERA5 (Arduini *et al.*, 2019). Wang *et al.* (2019) found a warm bias when comparing ERA5 to drifting buoys over sea ice and this could possibly be due to differences in the measurement height of buoys (Vihma *et al.*, 2014), which is likely at a lower height than 2 m given the changing nature of sea ice coupled with snow accumulation during winter. In our study, we find that winter temperatures are underestimated (Figure 4a). The location of stations

used in this study (ground-based, continental stations located near the coast or inland, Table 2) could contribute to the cold bias, whereas measurement buoys located directly on sea ice (with measurement heights possibly lower than 2 m) show an overall warm bias in ERA5 (Demchev *et al.*, 2020). This discrepancy could be due to differences in the representation of the surface energy budget between sea ice and inland regions, specifically the

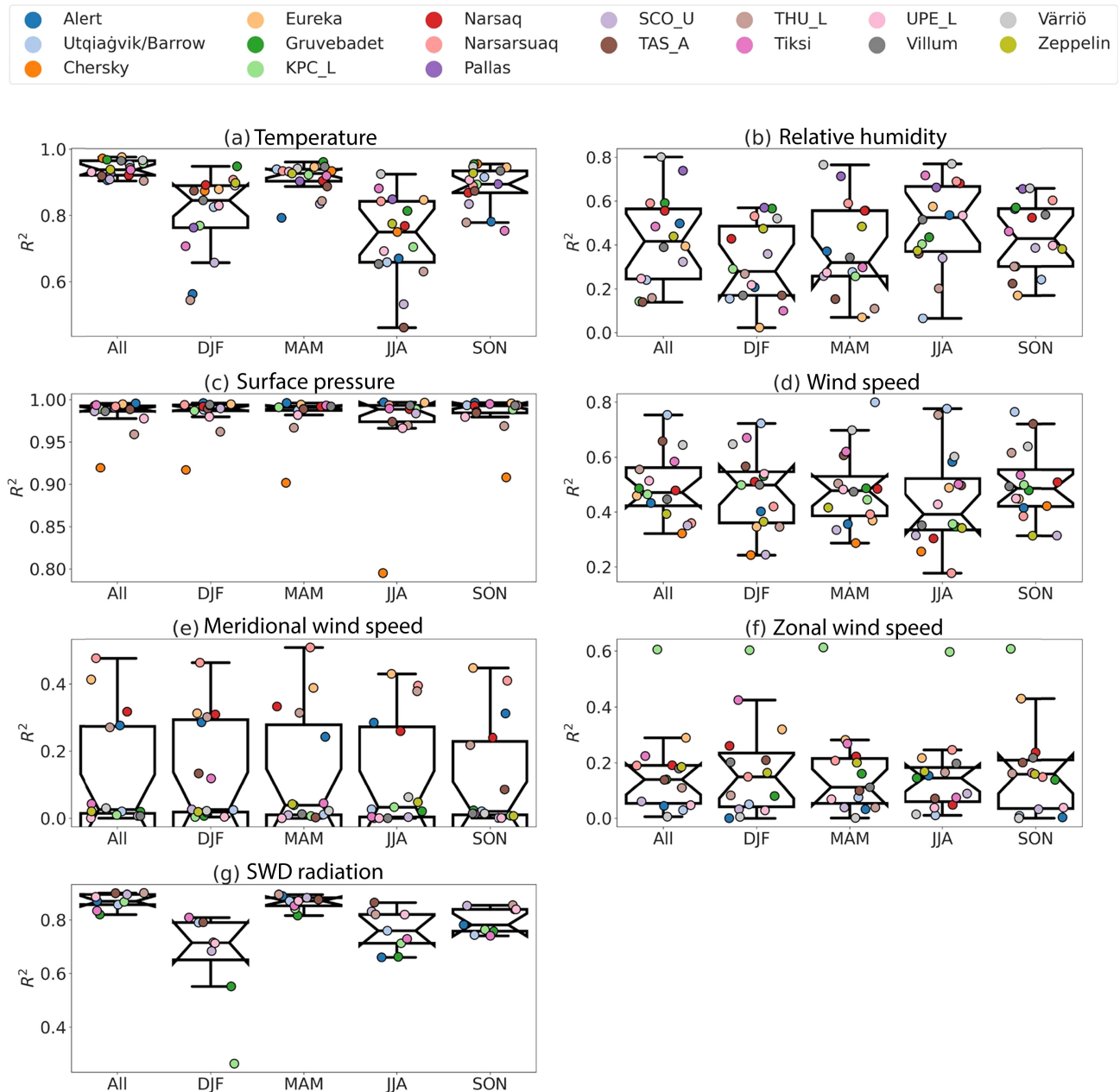


FIGURE 5 R^2 values for all available data and each meteorological season for (a) temperature; (b) relative humidity; (c) surface pressure; (d) wind speed; (e) meridional wind speed; (f) zonal wind speed; and (g) short-wave downward (SWD) radiation. The markers show the individual stations while the boxplot shows the distribution of these stations. The central line is the median and the two extremities of the boxes are the 25th and 75th percentiles, respectively. The whiskers extend to points that lie within 1.5 times the interquartile range. The notches on each box represent the 95% confidence level (CL) of each median. This allows us to see if the medians of two distributions are significantly different from each other because then the notches do not overlap. [Colour figure can be viewed at wileyonlinelibrary.com]

conductive heat flux between sea ice/snow and the atmosphere (Walden *et al.*, 2017), the missing representation of the top snow layer on sea ice and sea ice thickness (Batra & Müller, 2019), and the representation of surface inversions in the stably stratified boundary layer (Graham *et al.*, 2019). Overall, it appears that temperature is under-predicted during winter and summer and generally well reproduced by ERA5 during spring and autumn.

For RH, the slope values for winter and spring are lower compared to all available data values, while summer slopes are higher, and autumn slopes correspond to the value for all available data (Figure 4b). The distributions of R^2 values during winter and summer show the largest deviation from the comparison using all available data (Figure 5b). The distribution of median RMSE values is similar between seasons albeit with different degrees of

variance (Figure S2b). This pattern for RH is undoubtedly linked to the estimation of RH from air and dew-point temperature. The August–Roche–Magnus formula is designed for approximating the saturation vapor pressure over liquid water, which is most prevalent in the Arctic during summer and early autumn when temperatures are near or above freezing and most water is present in liquid form. During winter, when most water is in the form of ice, the comparison for RH shows the poorest results.

For surface pressure, the agreement (R^2 , slope, and RMSE values) between ERA5 and observations is largely unaffected by seasons (Figures 4c, 5c, and S2c). During summer, the variance of R^2 values is slightly increased which is likely driven by Chersky which shows a pronounced decrease in the R^2 value during this season compared to the other seasons. At other stations, similar R^2 , slope, and RMSE values are observed regardless of season (Figures 4c, 5c, and S2c).

All three wind speed variables (wind speed, meridional wind speed, and zonal wind speed) show a similar (but poor) agreement regardless of season and metric (Figures 4d–f, 5d–f, and S2d–f).

For SWD radiation, spring and autumn show a similar distribution of slopes compared to using all available data points, while summer displays a lower median slope, and winter shows a large variance of slopes, ranging from 0 to above 1 (Figure 4g). One station (SCO_U) even overpredicts SWD radiation, which is surprising since SCO_U is located above the Arctic circle (72° N, Table 2) and thus experiences polar night. During winter, the median of R^2 values is substantially smaller compared to using all available data, followed by summer and autumn (Figure 5g). During spring, a similar distribution of R^2 values is observed compared to using all available data. This poor agreement during winter is likely caused by ERA5 producing zeros for all timestamps during the polar night for stations that are above the Arctic circle while in reality observations from pyranometers often produce non-zero values for SWD radiation flux during winter, possibly either due to emissions from the surface that is captured in the short-wave range of the instruments' spectrum or due to instrumental uncertainties. The distribution of RMSE values varied across the seasons when compared to using all observations, with the lowest RMSE values during winter and the highest during summer (Figure S2g). This is likely related to the magnitude of the SWD radiation values during these seasons. Delhasse *et al.* (2020) found a mean correlation coefficient for SWD radiation flux during the summer of 0.91 compared to 0.98 when using all available data, whereas in the current study, a mean correlation coefficient of 0.87 was found during the summer and 0.93 when using all available data (Supporting information). Overall, SWD radiation flux gives acceptable results

during spring, summer, and autumn but caution should be used when using this variable during polar night at high-latitude stations.

Overall, temperature, RH, and SWD radiation flux show seasonal variations with underestimations in winter, whilst surface pressure, which is well reproduced, and wind speed variables, which are poorly reproduced, do not show any seasonal dependence.

3.4 | Effect of temporal resolution on ERA5 performance

We explored the performance of ERA5 at various temporal resolutions (one hour, six hours, one day, one month) to evaluate how different resolutions affect the agreement between ERA5 and in-situ observations. This is important, because depending on the purpose for which ERA5 data is used, a different temporal resolution might be chosen. In addition, previous studies have used varying temporal resolutions from hourly to monthly (Delhasse *et al.*, 2020; Demchev *et al.*, 2020; Graham *et al.*, 2019; Tetzner *et al.*, 2019), thus hindering a direct comparison between studies. Figures 6 and 7 show a similar analysis as Figures 5 and 6, with boxplots showing the distribution of slope and R^2 values, respectively, and markers showing the individual stations. Derivation of RH and wind speed was performed prior to any temporal averaging. We set a 50% data coverage threshold of in-situ observations for a temporal average to be included in the comparison.

For temperature, there is little change between the distribution of slope values for six hours and one day when compared to one hour, although, for a resolution of one month, it appears that ERA5 slightly overestimates temperature (Figure 6a). For the distribution of R^2 values, an increase in the overall median, that is, closer to 1, and a decrease in the variance is observed when decreasing the temporal resolution from one hour to one month (Figure 7a). A similar improvement is observed for the median of RMSE values, that decrease when decreasing the temporal resolution (Figure S3a). Delhasse *et al.* (2020) used daily resolution and found a mean correlation coefficient of 0.97 between ERA5 and observations for temperature. This is almost identical when compared to the mean correlation coefficient from the current study when using daily resolution (0.98, Supporting information). We observe that the smoothing effect from hourly to daily resolution does not seem to substantially influence the slope of the temperature predictions. Previous studies have also found an excellent agreement for temperature in ERA5 at monthly resolution. Tetzner *et al.* (2019) used monthly mean temperature from the Antarctic Peninsula

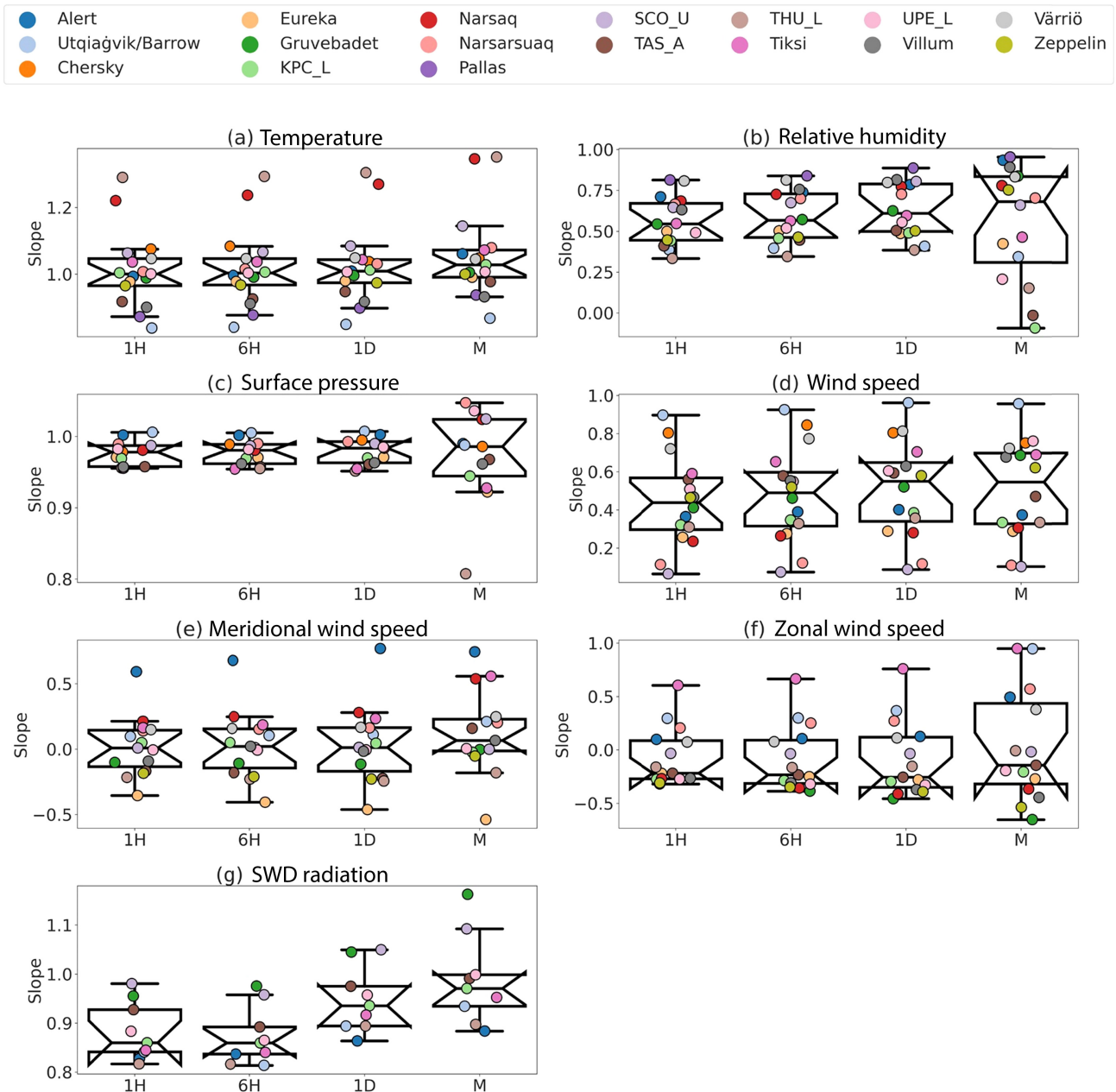


FIGURE 6 Slope values for all available data and for different time resolutions for (a) temperature; (b) relative humidity; (c) surface pressure; (d) wind speed; (e) meridional wind speed; (f) zonal wind speed; and (g) short-wave downward (SWD) radiation. The markers show the individual stations while the boxplot shows the distribution of these stations. The central line is the median and the two extremities of the boxes are the 25th and 75th percentiles, respectively. The whiskers extend to points that lie within 1.5 times the interquartile range. The notches on each box represent the 95% CL of each median. This allows us to see if the medians of two distributions are significantly different from each other if the notches do not overlap. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

and found a mean R^2 of 0.98 compared to 0.99 (Supporting information) in this study at monthly resolution. Zhu *et al.* (2021) analyzed weather stations from all over Antarctica and found correlation coefficients greater than 0.95 at monthly resolution. Temperature gives a similar median slope between ERA5 and observations for resolutions of one hour, six hours, and one day; however, for monthly resolution ERA5 slightly overestimates

temperature. The median of the R^2 values increases toward 1 and decreases in variance when increasing the resolution from one hour to one month. Overall, with decreasing time resolution from hourly to monthly, the closer the agreement (R^2) between ERA5 and observations becomes, although the slope shows a slight overestimation.

For RH, there is an improvement in the median slope values when decreasing the temporal resolution from one

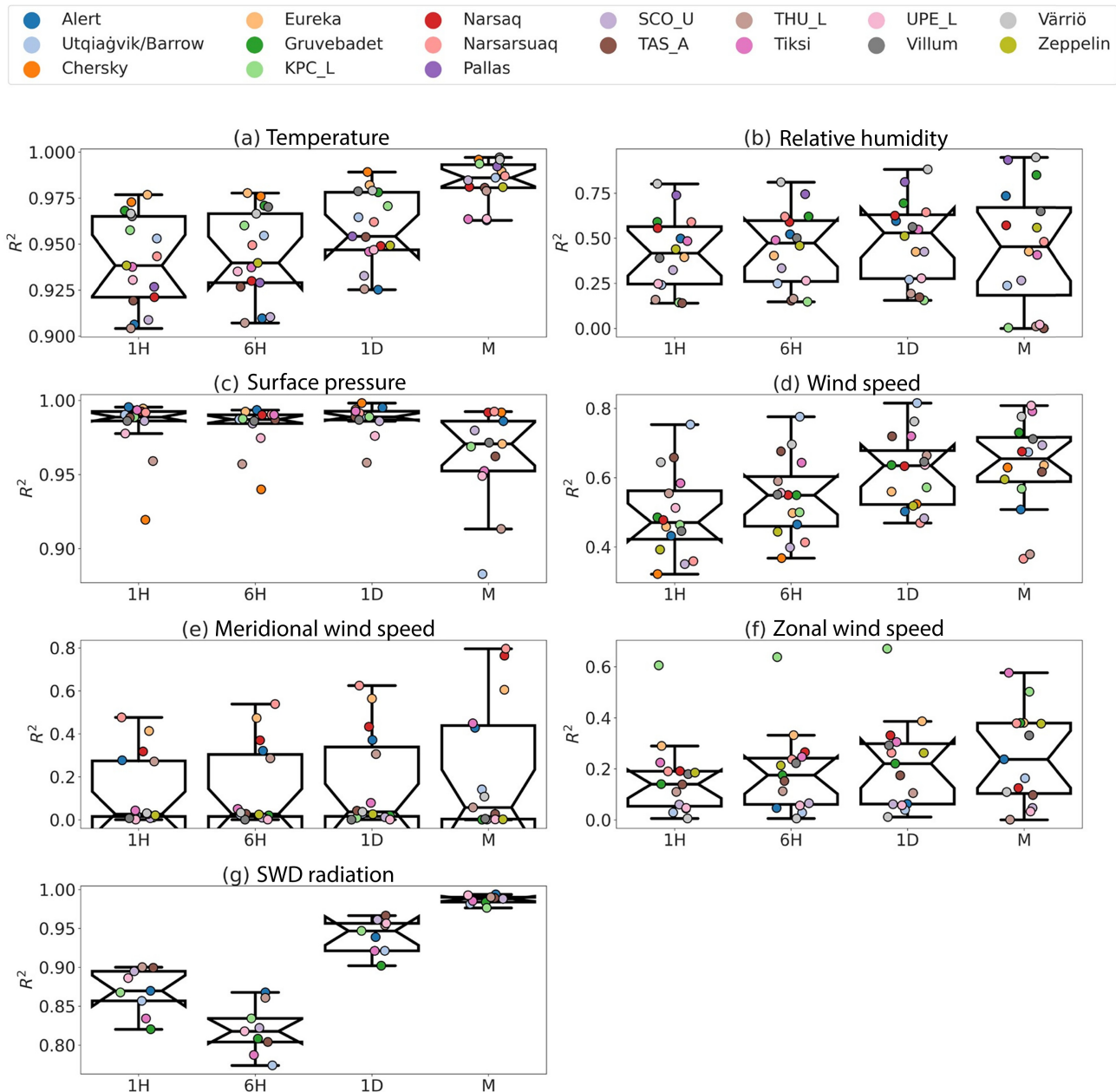


FIGURE 7 R^2 values for all available data and for different time resolutions for (a) temperature; (b) relative humidity; (c) surface pressure; (d) wind speed; (e) meridional wind speed; (f) zonal wind speed; and (g) short-wave downward (SWD) radiation. The markers show the individual stations while the boxplot shows the distribution of these stations. The central line is the median and the two extremities of the boxes are the 25th and 75th percentiles, respectively. The whiskers extend to points that lie within 1.5 times the interquartile range. The notches on each box represent the 95% confidence level (CL) of each median. This allows us to see if the medians of two distributions are significantly different from each other if the notches do not overlap. [Colour figure can be viewed at wileyonlinelibrary.com]

hour to one month when focusing on the slope values although one-month resolution shows drastically increased variance (Figure 6b). The median R^2 value improves up to one-day time resolution and then drops for one month, although is at a similar level compared to one-hour resolution (Figure 7b). With regards to the RMSE median value, it becomes smaller the longer the time integration, for example, at one hour the RMSE median is 11.3% while at one month it is 7.2% (Figure S3b). This

pattern is linked to the pattern for temperature, as expressed above since the estimation of RH uses temperature. While the best values of slope and R^2 for individual stations were found for monthly resolution, the overall best distribution of slope and R^2 values (highest quartile values and lowest variance) is produced for daily resolution.

For surface pressure, there is little difference in the distribution and medians of slope and R^2 values of one hour,

six hours, and one day; however, for a temporal resolution of one month there is a substantial increase in the variance and an overall decrease in the median R^2 value (Figures 6c and 7c, respectively). Interestingly, THU_L shows particularly worse results for the slope at monthly resolution compared to other resolutions. For R^2 , THU_L and Utqiagvik/Barrow show comparatively worse results for monthly resolution. For the distribution of RMSE values, there is little difference between different temporal resolutions, with all stations performing similarly compared to themselves regardless of resolution (Figure S3c). Overall, there is little difference for resolutions less than one day but for monthly resolution comparatively worse results are evident.

Wind speed and its directional components only show marginal improvement comparatively with coarser time resolution with respect to their median slope values, while the median R^2 improves substantially, in particular for general wind speed and the zonal component (Figures 6d–f and 7d–f, respectively). The median RMSE values improve substantially for all wind speed variables and are best for one month (Figure S3d–f). This is in line with Molina *et al.* (2021), who also found a better agreement between ERA5 and observations of wind speed when considering daily resolution over 1 or 6-h resolutions. Better agreement at lower time resolution is expected, because while observational data capture short-term gusts and orographic wind patterns, ERA5 cannot be expected to do so.

For SWD radiation, the median slope value improves with a coarser time resolution (Figure 6g). For median R^2 values, the pattern is similar with near-perfect agreement at a one-month resolution. However, there is an exception, with comparatively worse performance for six hours relative to one hour. The median RMSE values (Figure S3g) improve with coarser time resolution and feature the same exception at six hours. While hourly resolution produced satisfactory results, daily and monthly resolutions produced slopes and R^2 values are close to 1 and low RMSE values.

Overall, the lower the time resolution the better ERA5 performs for all variables except for SWD. Already decreasing the temporal resolution from one hour to one day can bring substantial improvement. These findings should be considered when utilizing ERA5 in climate models or for supplementing observations.

4 | CONCLUSIONS & IMPLICATIONS

This work compared data from the ERA5 reanalysis model with in-situ meteorological data in the continental Arctic to assess the performance of ERA5. The comparison

was performed using hourly data from 17 geographically dispersed ground-based stations in the continental Arctic, over periods ranging between six and 20 years per station. We focused on three performance aspects, that is: (i) the performance of ERA5 at one-hour time resolution across all stations and using all available data points; (ii) the seasonal performance; and (iii) the change in performance with varying time resolution (one hour to one month). The existing literature lacks a comparison for a large temporal period, and for a high temporal resolution, and with a comprehensive spatial representation across the continental Arctic, this work aims at filling this gap.

Surface pressure is the best-performing variable analyzed (0.98 ± 0.02), which is related to the fact that for the majority of the stations, observations of surface pressure are assimilated into ERA5. This also highlights the skill of ERA5 at capturing synoptic conditions. Temperature shows a very good agreement (0.94 ± 0.02), with small biases likely due to the difference between the modeled altitude and the station altitude, which is consistent with the previous literature. SWD radiation flux is fairly well modeled by ERA5 (0.87 ± 0.03) and is consistent with results from previous literature in the Arctic. The RH estimated from air and dew-point temperature from ERA5 has a rather low agreement and high variability (0.42 ± 0.20). Comparisons of RH measurements with ERA5 are rare. A comparison to another study performed in China (Huai *et al.*, 2021) showed that the results for the Arctic were substantially worse. This may indicate that the approximation used to estimate RH is not adapted to extreme Arctic temperatures and the dominant thermodynamic phase of water (mostly ice or mixed-phase in the Arctic whereas the estimation was designed for liquid water). The wind speed evaluation has quite low agreement with variable results between stations (0.49 ± 0.12). The meridional (0.129 ± 0.17) and zonal (0.163 ± 0.15) wind speeds were also not well reproduced. The performance is highly dependent on the topographic conditions of the station considered, as certain features (e.g., wind gusts, katabatic winds, canyon/fjord effects) of the local geography might not be well represented in the model grid cell, as noted by Koltzow *et al.* (2022).

When calculating the performance on a seasonal basis, temperature, RH, and SWD radiation flux showed seasonal variations whereas surface pressure and wind speed did not display any seasonal dependency. Temperature and SWD flux showed worse performance during winter and summer compared to autumn and spring, while RH showed worse performance during winter/spring compared to summer/autumn. The comparatively worse performance of ERA5 regarding temperature during winter has important implications for studies investigating AA, which occurs mainly during autumn/winter

(Serreze & Barry, 2011). While the magnitude of AA might not be affected if the temperature is estimated with similar accuracy and bias in the rest of the world as it is in the Arctic, the rate of true Arctic warming from ERA5 might be underestimated. However, Isaken *et al.* (2022) showed that ERA5 was able to accurately reproduce observed trends in temperature from Svalbard and the Barents Sea (which is an area experiencing the greatest magnitude of warming in the Arctic). The proper reproduction of temperature magnitudes and trends could have implications for the estimation of tipping points that are affected by temperature (Armstrong McKay *et al.*, 2022; Lenton, 2012). The results presented in this study and from Isaksen *et al.* (2022) show that ERA5 is able to accurately reproduce temperature values and trends, respectively. Previous studies have shown that ERA5 has a warm bias over sea ice during winter (Beesley *et al.*, 2000; Graham *et al.*, 2017; Herrmannsdörfer *et al.*, 2023; Jakobson *et al.*, 2012; Tjernstrom & Graversen, 2009; Yu *et al.*, 2021). Here we show that ERA5 has a cold bias during winter for ground-based continental stations. This could be due to different representations of the surface energy budget and thermodynamic structure of the lower atmosphere between sea ice and inland regions. Batrak and Müller (2019) showed that the warm biases in reanalysis products over the Arctic Ocean can be due to an overestimation of the conductive heat flux from the ocean to the atmosphere and is likely due to a misrepresentation of the snow layer on sea ice and sea ice thickness. Renfrew *et al.* (2021) also showed the performance of ERA5 was greater over open water compared to ice-covered regions and highlighted an overly smooth sea ice distribution. All of the stations included in this study are continental, therefore, other sources of error in ERA5 could then be the cause of the cold bias during winter (e.g., turbulent heat flux, cloud properties, and stability of the boundary layer) (Arduini *et al.*, 2022; Graham *et al.*, 2017; Graham *et al.*, 2019; Kayser *et al.*, 2017) and requires further investigation. The comparatively worse performance of ERA5 during summer (relative to other seasons) should also be considered if these data are being used to supplement meteorological observations, as most research campaigns are performed during summer/autumn. The discrepancy of ERA5 in the summer also has important implications when investigating temperatures near the freezing point of water, where a small discrepancy can influence the interpretation of the thermodynamic phase of water and thus scientific interpretation of other atmospheric phenomena.

When considering the effects of different temporal resolutions used on the comparison, the largest differences appear to be between one month and higher resolutions of less than one day. In general, there is either no change or a slight improvement between ERA5 and in-situ

observations when decreasing the temporal resolution from one hour to one day, with the best agreement when utilizing monthly resolution. This has implications for modelers when nudging climate models to ERA5 reanalysis data but also for observationalists who use ERA5 to supplement missing meteorological measurements. The investigation of regime transitions (e.g., the onset of melt and freeze-up days of sea ice) requires the use of high temporal resolution (i.e., daily) data to detect the changes in the Arctic environment. Daily averages can be used without a loss of accuracy while still maintaining a relatively high time resolution. The use of daily values will enable scientists to investigate seasonal processes that occur on timescales faster than one month, maintain a high temporal resolution, and obtain robust statistics. Very highly time-resolved gridded data can result in long computational times and storage issues in analyzing a long period and large geographic area. While the comparison evaluation improved with coarser temporal resolution, acceptable results were obtained at hourly resolution. This will allow scientists to investigate diurnal processes and feedback mechanisms at a high temporal resolution while still maintaining confidence in the reanalysis data. Extreme weather events, which are becoming more frequent in the Arctic (Fischer *et al.*, 2021; Walsh *et al.*, 2020), can occur suddenly and are transient thus making highly temporally resolved data necessary for the investigation of their causes and implications. Finally, our temporal resolution analysis will aid future studies in comparing ERA5 to in-situ observations by providing the evaluation metrics at multiple time resolutions.

Overall, this study provides an overview of the agreement between the ERA5 model and in-situ observations across the continental Arctic domain and can help guide the confidence level that can be placed in each of the surface meteorological variables from ERA5 during each season and at different temporal resolutions.

ACKNOWLEDGEMENTS

Jakob Boyd Pernov received funding from the SDSC grant (C20-01: ArcticNAP). Julia Schmale holds the Ingvar Kamprad Chair for Extreme Environments Research sponsored by Ferring Pharmaceuticals. Andrea Baccharini is acknowledged for helping produce the map in Figure 1. The CNR-ISP and its staff at the Arctic Station Dirigibile Italia are acknowledged for meteorological data from the Climate Change Tower in Ny-Ålesund. NILU (<https://www.nilu.com/>) and EBAS are acknowledged for the measurements from the Zepelin Observatory. The Finnish Meteorological Institute and the LITDB website (https://litdb.fmi.fi/suo0003_data.php) are acknowledged for measurements from Pallas. The Geological Survey of Denmark and Greenland

(GEUS), is acknowledged for measurements from the Programme for Monitoring of the Greenland Ice Sheet (PROMICE, <http://www.promice.dk>). The Danish Meteorological Institute (DMI) is acknowledged for measurements from Villum (Station Nord), Narsaq, and Narsarsuaq (Jensen, 2022). The EBAS database is acknowledged for providing data from Värriö and Zeppelin. The National Oceanic and Atmospheric Administration (NOAA) is acknowledged for measurement data from Eureka, Tiksi, Alert, Chersky, Utqiagvik/Barrow. The Baseline Surface Radiation Network (BSRN) is acknowledged for measurement data from Alert, Ny-Ålesund, Tiksi, and Utqiagvik/Barrow. Open access funding provided by École Polytechnique Fédérale de Lausanne.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data sources for each station are listed in the Appendix. The code used to analyze data and create all figures can be found here: <https://renkulab.io/projects/arcticnap/arcticnap> under the “ERA5 Obs Comparison Article Code” folder.

ORCID

Jakob Boyd Pernov  <https://orcid.org/0000-0003-1906-2589>

Julia Schmale  <https://orcid.org/0000-0002-1048-7962>

REFERENCES

- Alduchov, O.A. & Eskridge, R.E. (1996) Improved magnus form approximation of saturation vapor pressure. *Journal of Applied Meteorology and Climatology*, 35, 601–609.
- AMAP. (2022) *AMAP Arctic climate change update 2021: Key trends and impacts, Arctic monitoring and assessment programme (AMAP)*. Tromsø: Arctic Monitoring and Assessment Programme (AMAP).
- Arduini, G., Balsamo, G., Dutra, E., Day, J.J., Sandu, I., Boussetta, S. et al. (2019) Impact of a multi-layer snow scheme on near-surface weather forecasts. *Journal of Advances in Modeling Earth Systems*, 11, 4687–4710. Available from: <https://doi.org/10.1029/2019MS001725>
- Arduini, G., Keeley, S., Day, J.J., Sandu, I., Zampieri, L. & Balsamo, G. (2022) On the importance of representing snow over sea-ice for simulating the Arctic boundary layer. *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002777. Available from: <https://doi.org/10.1029/2021MS002777>
- Armstrong McKay, D.I., Staal, A., Abrams, J.F., Winkelmann, R., Sakschewski, B., Loriani, S. et al. (2022) Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science*, 377, eabn7950. Available from: <https://doi.org/10.1126/science.abn7950>
- Avila-Diaz, A., Bromwich, D.H., Wilson, A.B., Justino, F. & Wang, S.-H. (2021) Climate extremes across the North American Arctic in modern reanalyses. *Journal of Climate*, 34, 2385–2410. Available from: <https://doi.org/10.1175/JCLI-D-20-0093.1>
- Babar, B., Graverson, R. & Boström, T. (2019) Solar radiation estimation at high latitudes: assessment of the CMSAF databases, ASR and ERA5. *Solar Energy*, 182, 397–411. Available from: <https://doi.org/10.1016/j.solener.2019.02.058>
- Batrak, Y. & Müller, M. (2019) On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice. *Nature Communications*, 10, 4170. Available from: <https://doi.org/10.1038/s41467-019-11975-3>
- Beesley, J.A., Bretherton, C.S., Jakob, C., Andreas, E.L., Intrieri, J.M. & Uttal, T.A. (2000) A comparison of cloud and boundary layer variables in the ECMWF forecast model with observations at surface heat budget of the Arctic Ocean (SHEBA) ice camp. *Journal of Geophysical Research: Atmospheres*, 105, 12337–12349. Available from: <https://doi.org/10.1029/2000JD900079>
- Betts, A.K., Chan, D.Z. & Desjardins, R.L. (2019) Near-surface biases in ERA5 over the Canadian prairies. *Frontiers in Environmental Science*, 7, 129.
- Bracegirdle, T.J. & Marshall, G.J. (2012) The reliability of Antarctic tropospheric pressure and temperature in the latest global reanalyses. *Journal of Climate*, 25, 7138–7146. Available from: <https://doi.org/10.1175/JCLI-D-11-00685.1>
- Cai, Z., You, Q., Wu, F., Chen, H.W., Chen, D., & Cohen, J. (2021). Arctic warming revealed by multiple CMIP6 models: Evaluation of historical simulations and quantification of future projection uncertainties. *Journal of Climate*, 1–52. <https://doi.org/10.1175/JCLI-D-20-0791.1>
- Coulson, S., Dangendorf, S., Mitrovica, J.X., Tamisiea, M.E., Pan, L. & Sandwell, D.T. (2022) A detection of the sea level fingerprint of Greenland Ice Sheet melt. *Science*, 377, 1550–1554. Available from: <https://doi.org/10.1126/science.abe0926>
- Cox, C.J., Walden, V.P., Rowe, P.M. & Shupe, M.D. (2015) Humidity trends imply increased sensitivity to clouds in a warming Arctic. *Nature Communications*, 6, 10117. Available from: <https://doi.org/10.1038/ncomms10117>
- Delhasse, A., Kittel, C., Amory, C., Hofer, S., van As, D.S., Fausto, R. et al. (2020) Brief communication: evaluation of the near-surface climate in ERA5 over the Greenland Ice Sheet. *The Cryosphere*, 14, 957–965. Available from: <https://doi.org/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. Available from: <https://doi.org/10.3103/S1068373920110035>
- Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E. et al. (2018) Baseline surface radiation network (BSRN): structure and data description (1992–2017). *Earth System Science Data*, 10(3), 1491–1501. Available from: <https://doi.org/10.5194/essd-2018-8>
- ECMWF. (2016) IFS Documentation CY41R2–Part I: Observations, in: IFS Documentation CY41R2, ECMWF. <https://doi.org/10.21957/9phrx9eu>
- Fausto, R.S., van As, D., Mankoff, K.D., Vandecrux, B., Citterio, M., Ahlström, A.P. et al. (2021) Programme for monitoring of the Greenland Ice Sheet (PROMICE) automatic weather station data. *Earth System Science Data*, 13, 3819–3845. Available from: <https://doi.org/10.5194/essd-13-3819-2021>

- Fischer, E.M., Sippel, S. & Knutti, R. (2021) Increasing probability of record-shattering climate extremes. *Nature Climate Change*, 11, 689–695. Available from: <https://doi.org/10.1038/s41558-021-01092-9>
- Graham, R.M., Hudson, S.R. & Maturilli, M. (2019) Improved performance of ERA5 in Arctic gateway relative to four global atmospheric reanalyses. *Geophysical Research Letters*, 46, 6138–6147. Available from: <https://doi.org/10.1029/2019GL082781>
- Graham, R.M., Rinke, A., Cohen, L., Hudson, S.R., Walden, V.P., Granskog, M.A. et al. (2017) A comparison of the two Arctic atmospheric winter states observed during N-ICE2015 and SHEBA. *Journal of Geophysical Research: Atmospheres*, 122, 5716–5737. Available from: <https://doi.org/10.1002/2016JD025475>
- Herrmannsdörfer, L., Müller, M., Shupe, M.D. & Rostosky, P. (2023) Surface temperature comparison of the Arctic winter MOSAiC observations, ERA5 reanalysis, and MODIS satellite retrieval. *Elementa Science of the Anthropocene*, 11, 00085. Available from: <https://doi.org/10.1525/elementa.2022.00085>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. Available from: <https://doi.org/10.1002/qj.3803>
- Hogan, R.: *Radiation quantities in the ECMWF model and MARS*, ECMWF; 2015.
- Huai, B., Wang, J., Sun, W., Wang, Y. & Zhang, W. (2021) Evaluation of the near-surface climate of the recent global atmospheric reanalysis for Qilian Mountains, Qinghai-Tibet Plateau. *Atmospheric Research*, 250, 105401. Available from: <https://doi.org/10.1016/j.atmosres.2020.105401>
- Huang, J., Zhang, X., Zhang, Q., Lin, Y., Hao, M., Luo, Y. et al. (2017) Recently amplified arctic warming has contributed to a continual global warming trend. *Nature Climate Change*, 7, 875–879. Available from: <https://doi.org/10.1038/s41558-017-0009-5>
- Isaksen, K., Nordli, Ø., Ivanov, B., Køltzow, M.A.Ø., Aaboe, S., Gjeltten, H.M. et al. (2022) Exceptional warming over the Barents area. *Scientific Reports*, 12(1), Article 1. <https://doi.org/10.1038/s41598-022-13568-5>
- Jakobson, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H. & Jaagus, J. (2012) Validation of atmospheric reanalyses over the central Arctic Ocean. *Geophysical Research Letters*, 39. Available from: <https://doi.org/10.1029/2012GL051591>
- Jensen, C.D. (2022) *Weather observations from Greenland 1958–2022*. Danish Meteorological Institute. pp. 1–14.
- Jung, T., Gordon, N.D., Bauer, P., Bromwich, D.H., Chevallier, M., Day, J.J. et al. (2016) Advancing polar prediction capabilities on daily to seasonal time scales. *Bulletin of the American Meteorological Society*, 97, 1631–1647. Available from: <https://doi.org/10.1175/BAMS-D-14-00246.1>
- Kayser, M., Maturilli, M., Graham, R.M., Hudson, S.R., Rinke, A., Cohen, L. et al. (2017) Vertical thermodynamic structure of the troposphere during the Norwegian young sea ICE expedition (N-ICE2015). *Journal of Geophysical Research: Atmospheres*, 122, 10855–10872. Available from: <https://doi.org/10.1002/2016JD026089>
- Køltzow, M., Schyberg, H., Støylen, E. & Yang, X. (2022) Value of the Copernicus Arctic Regional Reanalysis (CARRA) in representing near-surface temperature and wind speed in the north-east European Arctic. *Polar Research*, 41. Available from: <https://doi.org/10.33265/polar.v41.8002>
- Kopec, B.G., Feng, X., Michel, F.A. & Posmentier, E.S. (2016) Influence of sea ice on Arctic precipitation. *Proceedings of the National Academy of Sciences*, 113, 46–51. Available from: <https://doi.org/10.1073/pnas.1504633113>
- Lappalainen, H.K., Kerminen, V.-M., Petäjä, T., Kurten, T., Baklanov, A., Shvidenko, A. et al. (2016) Pan-Eurasian Experiment (PEEX): towards a holistic understanding of the feedbacks and interactions in the land–atmosphere–ocean–society continuum in the northern Eurasian region. *Atmospheric Chemistry and Physics*, 16, 14421–14461. Available from: <https://doi.org/10.5194/acp-16-14421-2016>
- Lenton, T.M. (2012) Arctic climate tipping points. *Ambio*, 41, 10–22. Available from: <https://doi.org/10.1007/s13280-011-0221-x>
- Lindsay, R., Wensnahan, M., Schweiger, A. & Zhang, J. (2014) Evaluation of seven different atmospheric reanalysis products in the Arctic. *Journal of Climate*, 27, 2588–2606. Available from: <https://doi.org/10.1175/JCLI-D-13-00014.1>
- Mistry, M.N., Schneider, R., Masselot, P., Royé, D., Armstrong, B., Kysely, J. et al. (2022) Comparison of weather station and climate reanalysis data for modelling temperature-related mortality. *Scientific Reports*, 12, 5178. Available from: <https://doi.org/10.1038/s41598-022-09049-4>
- Molina, M.O., Gutiérrez, C. & Sánchez, E. (2021) Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. *International Journal of Climatology*, 41, 4864–4878. Available from: <https://doi.org/10.1002/joc.7103>
- Petäjä, T., Duplissy, E.M., Tabakova, K., Schmale, J., Altstädter, B., Ancellet, G. et al. (2020) Overview: integrative and comprehensive Understanding on Polar Environments (iCUPE)–concept and initial results. *Atmospheric Chemistry and Physics*, 20, 8551–8592. Available from: <https://doi.org/10.5194/acp-20-8551-2020>
- Pithan, F. & Mauritsen, T. (2014) Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, 7, 181–184. Available from: <https://doi.org/10.1038/ngeo2071>
- Previdi, M., Janoski, T.P., Chiodo, G., Smith, K.L. & Polvani, L.M. (2020) Arctic amplification: a rapid response to radiative forcing. *Geophysical Research Letters*, 47, e2020GL089933. Available from: <https://doi.org/10.1029/2020GL089933>
- Rantanen, M., Karpechko, A.Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K. et al. (2022) The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, 3, 168. Available from: <https://doi.org/10.1038/s43247-022-00498-3>
- Renfrew, I.A., Barrell, C., Elvidge, A.D., Brooke, J.K., Duschka, C., King, J.C. et al. (2021) An evaluation of surface meteorology and fluxes over the Iceland and Greenland seas in ERA5 reanalysis: the impact of sea ice distribution. *Quarterly Journal of the Royal Meteorological Society*, 147, 691–712. Available from: <https://doi.org/10.1002/qj.3941>
- Rotach, M.W., Gohm, A., Lang, M.N., Leukauf, D., Stiperski, I. & Wagner, J.S. (2015) On the vertical exchange of heat, mass, and momentum over complex, mountainous terrain. *Frontiers in Earth Science*, 3, 76. Available from: <https://doi.org/10.3389/feart.2015.00076>
- Seo, M., Kim, H.-C., Lee, K.-S., Seong, N.-H., Lee, E., Kim, J. et al. (2020) Characteristics of the reanalysis and satellite-based surface

- net radiation data in the Arctic. *Journal of Sensors*, 2020, 8825870. Available from: <https://doi.org/10.1155/2020/8825870>
- Serreze, M.C. & Barry, R.G. (2011) Processes and impacts of Arctic amplification: a research synthesis. *Global and Planetary Change*, 77, 85–96. Available from: <https://doi.org/10.1016/j.gloplacha.2011.03.004>
- 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. Available from: <https://doi.org/10.1029/2020GL088120>
- Shupe, M.D., Rex, M., Blomquist, B., Persson, P.O.G., Schmale, J., Uttal, T. et al. (2022) Overview of the MOSAiC expedition: atmosphere. *Elementa Science of the Anthropocene*, 10, 00060. Available from: <https://doi.org/10.1525/elementa.2021.00060>
- Solomon, A., Shupe, M.D., Svensson, G., Barton, N.P., Batrak, Y., Bazile, E. et al. (2023) The winter central Arctic surface energy budget: a model evaluation using observations from the MOSAiC campaign. *Elementa: Science of the Anthropocene*, 11, 00104. Available from: <https://doi.org/10.1525/elementa.2022.00104>
- Straneo, F., Slater, D., Bouchard, C., Cape, M., Carey, M., Ciannelli, L. et al. (2022) An interdisciplinary perspective on Greenland's changing coastal margins. *Oceanography*, 35(3/4), 106–117. Available from: <https://doi.org/10.5670/oceanog.2022.128>
- Tetzner, D., Thomas, E. & Allen, C. (2019) A validation of ERA5 reanalysis data in the Southern Antarctic Peninsula—Ellsworth land region, and its implications for ice core studies. *Geosciences*, 9, 289. Available from: <https://doi.org/10.3390/geosciences9070289>
- Tjernstrom, M. & Graversen, R.G. (2009) The vertical structure of the lower Arctic troposphere analysed from observations and the ERA-40 reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 135, 431–443. Available from: <https://doi.org/10.1002/qj.380>
- Velikou, K., Lazoglou, G., Tolika, K. & Anagnostopoulou, C. (2022) Reliability of the ERA5 in replicating mean and extreme temperatures across Europe. *Water*, 14, 543. Available from: <https://doi.org/10.3390/w14040543>
- Vihma, T., Pirazzini, R., Fer, I., Renfrew, I.A., Sedlar, J., Tjernström, M. et al. (2014) Advances in understanding and parameterization of small-scale physical processes in the marine Arctic climate system: a review. *Atmospheric Chemistry and Physics*, 14, 9403–9450. Available from: <https://doi.org/10.5194/acp-14-9403-2014>
- Walden, V.P., Hudson, S.R., Cohen, L., Murphy, S.Y. & Granskog, M.A. (2017) Atmospheric components of the surface energy budget over young sea ICE: results from the N-ICE2015 campaign. *Journal of Geophysical Research: Atmospheres*, 122, 8427–8446. Available from: <https://doi.org/10.1002/2016JD026091>
- Walsh, J.E., Ballinger, T.J., Euskirchen, E.S., Hanna, E., Mård, J., Overland, J.E. et al. (2020) Extreme weather and climate events in northern areas: a review. *Earth-Science Reviews*, 209, 103324. Available from: <https://doi.org/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. Available from: <https://doi.org/10.5194/tc-13-1661-2019>
- Wang, G., Wang, T. & Xue, H. (2021) Validation and comparison of surface shortwave and longwave radiation products over the three poles. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102538. Available from: <https://doi.org/10.1016/j.jag.2021.102538>
- Wendisch, M., Macke, A., Ehrlich, A., Lüpkes, C., Mech, M., Chechin, D. et al. (2019) The Arctic cloud puzzle: using ACLOUD/PASCAL multiplatform observations to unravel the role of clouds and aerosol particles in Arctic amplification. *Bulletin of the American Meteorological Society*, 100, 841–871. Available from: <https://doi.org/10.1175/bams-d-18-0072.1>
- Wever, N. (2012) Quantifying trends in surface roughness and the effect on surface wind speed observations. *Journal of Geophysical Research: Atmospheres*, 117. Available from: <https://doi.org/10.1029/2011JD017118>
- Wu, H., Xu, X., Luo, T., Yang, Y., Xiong, Z. & Wang, Y. (2023) Variation and comparison of cloud cover in MODIS and four reanalysis datasets of ERA-interim, ERA5, MERRA-2 and NCEP. *Atmospheric Research*, 281, 106477. Available from: <https://doi.org/10.1016/j.atmosres.2022.106477>
- Yu, Y., Xiao, W., Zhang, Z., Cheng, X., Hui, F. & Zhao, J. (2021) Evaluation of 2-m air temperature and surface temperature from ERA5 and ERA-I using buoy observations in the Arctic during 2010–2020. *Remote Sensing*, 13, 2813. Available from: <https://doi.org/10.3390/rs13142813>
- Zampieri, L., Arduini, G., Holland, M., Keeley, S.P.E., Mogensen, K., Shupe, M.D. et al. (2023) A machine learning correction model of the winter clear-sky temperature bias over the Arctic sea ice in atmospheric reanalyses. *Monthly Weather Review*, 151, 1443–1458. Available from: <https://doi.org/10.1175/MWR-D-22-0130.1>
- Zhang, J., Zhao, T., Li, Z., Li, C., Li, Z., Ying, K. et al. (2021) Evaluation of surface relative humidity in China from the CRA-40 and current reanalyses. *Advances in Atmospheric Sciences*, 38, 1958–1976. Available from: <https://doi.org/10.1007/s00376-021-0333-6>
- Zhao, P. & He, Z. (2022) A first evaluation of ERA5-land reanalysis temperature product over the Chinese Qilian Mountains. *Frontiers in Earth Science*, 10, 907730. Available from: <https://doi.org/10.3389/feart.2022.907730>
- Zhu, J., Xie, A., Qin, X., Wang, Y., Xu, B. & Wang, Y. (2021) An assessment of ERA5 reanalysis for Antarctic near-surface air temperature. *Atmosphere*, 12, 217. Available from: <https://doi.org/10.3390/atmos12020217>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pernov, J.B., Gros-Daillon, J. & Schmale, J. (2024) Comparison of selected surface level ERA5 variables against in-situ observations in the continental Arctic. *Quarterly Journal of the Royal Meteorological Society*, 1–24. Available from: <https://doi.org/10.1002/qj.4700>

APPENDIX A. IN-SITU MEASUREMENTS

Data from each ground-based weather station were retrieved via a public data repository. The data sources for each station are grouped according to the data provider.

EBAS (<https://ebas.nilu.no/>):

- Zeppelin: <https://ebas-data.nilu.no/Pages/DataSetList.aspx?key=67A339AF52AA47F1874B083D20F4AE4B>
- Värriö: <https://ebas-data.nilu.no/Pages/DataSetList.aspx?key=5DBEB601A0004EA487955B689E0C8E0F>

Italian Arctic Data Center (IADC) (<https://iadc.cnr.it/cnr/>)

- Ny-Ålesund: https://data.iadc.cnr.it/erddap/info/cct_meteo_d2/index.html

National Oceanic and Atmospheric Administration (NOAA) (<https://www.noaa.gov/>):

- Eureka: ftp://ftp1.esrl.noaa.gov/psd3/arctic/eureka/surface_properties/fluxtower/
- Alert: ftp://ftp1.esrl.noaa.gov/psd3/arctic/alert/surface_properties/fluxtower/
- Tiksi: ftp://ftp1.esrl.noaa.gov/psd3/arctic/tiksi/surface_properties/fluxtower/
- Chersky: ftp://ftp1.esrl.noaa.gov/psd3/arctic/cherskii/temperature/GHCND_Cherskii.csv

Danish Meteorological Institute (DMI) (<https://www.dmi.dk/>):

- Villum: https://www.dmi.dk/fileadmin/Rapporte/2022/DMIREp22-08_1953_2021_data1.zip

- Narsarsuaq: https://www.dmi.dk/fileadmin/Rapporte/2022/DMIREp22-08_1953_2021_data1.zip
- Narsaq: https://www.dmi.dk/fileadmin/Rapporte/2022/DMIREp22-08_1953_2021_data1.zip

PROMICE Network (<https://promice.org/>):

- Kronprins Christian Land (KPC L).
- Scoresbysund (SCO L).
- Tasiilaq (TAS A).
- Thule (THU L).
- Upernavik (UPE L).

All the data related to these stations can be accessed via the following link: <https://dataverse.geus.dk/dataset.xhtml?persistentId=doi:10.22008/FK2/8SS7EW>

Baseline Surface Radiation Network (BSRN) (<https://bsrn.awi.de/>):

- Barrow: <https://doi.pangaea.de/10.1594/PANGAEA.959215>
- Tiksi: <https://doi.pangaea.de/10.1594/PANGAEA.959213>
- Ny-Ålesund: <https://doi.pangaea.de/10.1594/PANGAEA.914927>
- Alert: <https://doi.org/10.1594/PANGAEA.932867>

Finnish Meteorological Institute (FMI) (<https://en.ilmatieteenlaitos.fi/>):

- Pallas: https://litdb.fmi.fi/suo0003_data.php