https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

Open Access Earth System Science Data Discussions

Special Observing Period (SOP) Data for the Year of Polar Prediction site Model Intercomparison Project (YOPPsiteMIP)

Zen Mariani[1], Sara M. Morris[2,11], Taneil Uttal[2], Elena Akish[3,2], Robert Crawford[1], Laura Huang[1], Jonathan Day[4], Johanna Tjernström[12], Øystein Godøy[12], Lara Ferrighi[12], Leslie M. Hartten[3,2], Jareth Holt[6], Christopher J. Cox[2], Ewan O'Connor[9], Roberta Pirazzini[9], Marion Maturilli[13], Giri Prakash[10], James Mather[8], Kimberly Strong[5], Pierre Fogal[5], Vasily Kustov[7,14], Gunilla Svensson[6], Michael Gallagher[3,2], Brian Vasel[11]

[1]Meteorological Research Division, Environment and Climate Change Canada, Toronto, Canada
[2]NOAA Physical Sciences Laboratory, Boulder, CO, USA
[3]Cooperative Institute for Research in Environmental Science, University of Colorado, Boulder, Colorado, USA
[4]European Centre for Medium-Range Weather Forecasts, Reading, UK
[5]Department of Physics, University of Toronto, Toronto, Canada
[6]Department of Meteorology, Stockholm University, Sweden
[7]Arctic and Antarctic Research Institute, Air-sea interaction department, St. Petersburg, Russia
[8]Pacific Northwest National Laboratory, Richland, WA, USA
[9]Finnish Meteorological Institute, Finland
[10]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[11]NOAA Global Monitoring Laboratory, Boulder, CO, USA
[12]Norwegian Meteorological Institute, Norway
[13]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany
[14]Freelance entrepreneur, Belgrade, Serbia

Correspondence to: Zen Mariani (zen.mariani@ec.gc.ca) and Sara Morris (Sara.Morris@noaa.gov)

Abstract. The rapid changes occurring in the polar regions require an improved understanding of the processes that are driving the changes. At the same time increased human activities, such as marine navigation, resource exploitation, aviation, commercial fishing, and tourism, require reliable and relevant information. One of the primary goals of the World Meteorological Organization's Year of Polar Prediction (YOPP) Project is to improve the accuracy of numerical weather prediction (NWP) at high latitudes. During YOPP, two Canadian observatories were commissioned and equipped with new ground-based instruments for enhanced meteorological and system process observations that are considered to be "supersites" for addressing YOPP objectives, while other pre-existing supersites in Canada, the United States, Norway, Finland and Russia provided data from ongoing long-term observing programs. Data from these seven supersites were amalgamated and are being used to evaluate NWP systems from several international forecast centers and to perform meteorological process studies with the aim of improving NWP performance in the Polar Regions. In order to increase data useability and station interoperability, novel Merged Observatory Data Files (MODFs) have been created for these seven international supersites over two Special Observing Periods (February to March 2018 and July to September 2018). All observations collected at the seven supersites were compiled into this new standardized NetCDF MODF format, simplifying

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

the process of conducting pan-Arctic NWP verification and process evaluation studies. This paper describes the seven Arctic YOPP supersites, data collection and processing methods, and the novel MODF format and output files, which together comprise the observational contribution to the associated model intercomparison effort, termed YOPP supersite Model Intercomparison Project (YOPPsiteMIP). All YOPPsiteMIP MODFs are publicly accessible via the YOPP Data Portal (Whitehorse: https://doi.org/10.21343/a33e-j150, Iqaluit: https://doi.org/10.21343/yrnf-ck57, Sodankylä: https://doi.org/10.21343/m16p-pq17, Utqiaġvik: https://doi.org/10.21343/a2dx-nq55, Tiksi: https://doi.org/10.21343/5bwn-w881, Ny-Ålesund: https://doi.org/10.21343/y89m-6393, Eureka: https://doi.org/10.21343/r85j-tc61), hosted by MET Norway, with corresponding output from NWP models.

1 Introduction

In the Arctic there is a recognized lack of process-level information supplementing meteorological observations to characterize the atmosphere and the cryosphere for operational forecasting (Cassano et al., 2011; Illingworth et al., 2015; Lawrence et al., 2019). As the climate continues to change, information on weather and climate is becoming more critical in ensuring the health and safety of local communities. Unfortunately, climate models do a poor job of capturing key features of Arctic climate, such as the Arctic amplification factor, likely as a result of inaccurate representation of key physical processes, as shown by Rantanen et al. (2022). Similarly, the accuracy of weather forecasts in the Polar Regions is also lower than in mid-latitudes (Jung et al., 2016) partly due to the scattered and limited availability of observing networks (Lawrence et al., 2019). Advances in Polar prediction are expected to improve weather forecasts and climate predictions elsewhere (Jung et al., 2014 and Day et al., 2019), but understanding the causes of poor model performance in the Arctic is limited by the availability of observatory data. Data from observatories, where sometimes hundreds of parameters are measured, are needed for detailed investigations into the cause of model error, such as boundary-layer processes and turbulent exchanges (e.g., Day et al., 2023).

To address the need to improve Numerical Weather Prediction (NWP) performance in the Polar Regions, the World Meteorological Organization (WMO) launched the international Polar Prediction Project with its flagship activity, the Year of Polar Prediction (YOPP). During YOPP's core phase, from mid-2017 to mid-2019, several intensive observing periods were conducted with close coordination between the international network of polar observatories and weather forecast centers. The aim was to produce highly-concentrated sets of observed and modelled data for supporting forecast evaluation and process studies (Koltzow et al., 2019; Goessling et al., 2016; Jung et al., 2016).

One of the flagship activities of YOPP was the YOPP supersite Model Intercomparison Project (YOPPsiteMIP), an initiative to assess the performance of NWP systems at the process level by comparing with observatory data (Day et al., 2023). To achieve this, a dataset of weather forecasts was produced by various NWP centers for supersite locations. In the Arctic the

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

dataset covers two Special Observing Periods (SOPs), SOP1 (February 1 – March 31, 2018) and SOP2 (July 1 – September 30, 2018). During this period the number of routine observations (e.g. radiosonde launches, buoy deployments, etc.) were enhanced in the Arctic (doubled in the case of radiosondes), field campaigns were conducted, and enhanced observations from the designated YOPP "supersite" observatories were taken. This paper documents the efforts to compile the supersite data collected during this period as part of the YOPPsiteMIP. These supersites (Figure 1) are distributed over a diverse range of geographical locations capturing some of the diversity in the terrestrial high-latitude climate zones.

Prior to YOPP, data collection, processing, geophysical variable reporting cadences, and file output type and format were not standardized across the supersites, which are operated by different international agencies and consortiums. This lack of interoperability made performing multi-site comparisons, evaluations, and process studies difficult and time consuming, deterring potential users of supersite data (Wohner et al., 2022). In order to address this problem, the concept of standardized Merged Observatory Data Files (MODFs) was developed as part of the YOPPsiteMIP (Uttal et al., 2023). This concept is based on combining measurements from multiple international research observatories' instruments into a single NetCDF file that complies with established data management standards. Prior to MODFs, there generally existed no standardized procedures for coordinated data management at these research sites such as those that have been developed for operational datasets. Thus, the data from these sites' separate instruments were scattered between separate files with different authors, formats, metadata, post-processing techniques, physical archive locations, and requirements for usage. As such, they could not be amalgamated to provide a pan-Arctic observational dataset.

MODF files bring together observations from different earth system components in a standardized NetCDF file format to enable utilization of research-grade, process-level observations for model evaluation and parameterization development. At the same time, MODFs are compatible with and mirror Merged Model Data Files (MMDFs) that are produced by each NWP centre participating in YOPP (Day et al., 2023). Each geophysical variable observed at a site is matched to its corresponding NWP model geophysical variable using identical data format, cadence, and file structure in order to facilitate improved observation-model comparisons at the supersites (Gallagher and Tjernström, 2024). Uttal et al. (2023) provides a generalized overview for the content and data structure of MODFs, i.e., a single NetCDF data file containing measurements from multiple sources, and a series of tools to facilitate their creation. The purpose of the present work is to describe the construction and contents of MODFs for seven of the YOPP-designated Arctic supersites during SOPs 1 and 2 (hereafter, "$MODF_{ysm}$"): Whitehorse, Canada (60.71 °N, 135.07 °W, 682 m a.s.l.); Iqaluit, Canada (63.74 °N, 68.51 °W, 11 m a.s.l.); Sodankylä, Finland (67.367 °N, 26.629 °E, 179 m a.s.l.); Utqiaġvik (Barrow), Alaska (71.325 °N, 156.625 °W, 8 m a.s.l.); Tiksi, Russia (71.596 °N, 128.889 °E, 30 m a.s.l.); Ny-Ålesund, Norway (78.923 °N, 11.926 °E, 15 m a.s.l.); and Eureka, Canada (80.083 °N, 86.417 °W, 89 m a.s.l.). Table 1 provides information regarding the on-site facility location where measurements were collected and their coordinates for reference. For some sites (e.g., Sodankylä), certain geophysical variables are measured at multiple

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

locations; these are all reported in the MODF with their corresponding measurement coordinates embedded within the file so as to distinguish each measurement. These MODFs closely mirror the format used to archive the YOPPsiteMIP NWP data, in order to enable model evaluation. Final DOIs for the $MODF_{yms}$ are listed in Table 2.

Creating a standardized dataset such as MODF that contains observations from different meteorological and research agencies' sites is an extremely complex, non-trivial task. For the sake of brevity and to reduce redundancy, this paper references site- or instrument-specific publications in order to fully describe all of the aspects of the MODF dataset, including instrumentation, quality control, and processing techniques. In the case where non-trivial aspects about the MODF data arise, the data's origin, reference publications (e.g., dataset dois), and site contacts have been provided. Section 2 describes the data processing chain conducted at each supersite, including information about the site's local topography, climate, and instrumentation in order to provide site-specific context to aid the interpretation of model-observation comparisons. Section 3 describes the instrumentation and calculated variables. Section 4 describes the standardized MODF dataset file format, quality control, and post processing, which in some cases differed slightly from site-to-site. Section 5 describes the MODF data structure, attributes, and example Figures that illustrate the available dataset. Data and code availability is provided in Section 6, and concluding remarks are provided in Section 7.

2 Site Descriptions

It is important to properly contextualize and interpret the observations contained within the MODF since they come from vastly different sites. A map of the distribution of the supersites is shown in Figure 1 and local maps showing the vicinity around each supersite are found in Figure 2. For context, also shown in Figure 2, are native spatial grids of the forecast models that participated in YOPPsiteMIP. While all supersites are also designated surface synoptic observation (SYNOP) stations, the meteorological data provided in the MODFs is significantly more detailed and includes additional geophysical variables and thus is not the same as the SYNOP data. Table 3 lists the geophysical variables observed at each site that are stored in the standardized MODF format, their measurement location(s), and other attributes; the MODF featureType corresponds to the type of geophysical variable being observed at each site (they are split up into broad categories).

2.1 Whitehorse, Canada

The Whitehorse supersite was commissioned as part of the Canadian Arctic Weather Science (CAWS) project (Mariani et al., 2018; Joe et al., 2020). The supersite's instruments are installed on an elevated platform, all within a few meters of each other. Whitehorse has a population greater than 26,000 inhabitants. It is the primary gateway for air traffic for all of the Yukon Territories, parts of Alaska, and the Western Canadian Arctic. The supersite is located at the Erik Nielsen Whitehorse

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

International Airport, which is situated on a plateau ~50 m above the rest of the city. The city is located in a valley between the Yukon Ranges to its West (~1.6 km a.s.l.) and East (~1.4 km a.s.l.); this complex mountainous terrain strongly influences the weather systems that reach Whitehorse, which mostly originate from the Eastern Pacific or over Alaska. The primary surface wind direction follows the valley (NNW) and the average roughness length is estimated to be 1.0 m (Pinard et al., 2005). The soil type at and around the site is a mixture of grained alluvial and colluvial slopes and, as part of the Boreal Cordillera ecozone, the surface type is primarily Boreal Forest, including complex plateaus, mountains, valleys and Cordilleran vegetation. Whitehorse experiences cold to temperate average monthly temperatures ranging from -15 to 14 ºC and average monthly precipitation ranging from 7 to 38 mm. Since the town is in the rain shadow of the Coast Mountains, precipitation totals are relatively low year-round.

2.2 Iqaluit, Canada

Like Whitehorse, the Iqaluit supersite was commissioned as part of the CAWS project (Mariani et al., 2022). It is located ~200 m from the airport runway and all instruments are co-located to within no more than 140 m of each other on flat terrain. Co-located instrument evaluation studies were conducted for several remote sensing and upper air observations (Mariani et al., 2020, 2021), including preliminary model verification studies during the YOPP SOPs and beyond. Iqaluit has over 8,000 inhabitants and is the primary gateway for air and sea traffic for the central and Eastern Canadian Arctic. The city itself is located along the coast in a valley that runs in the NW to SE direction; thus, the primary direction of surface winds, which are frequently severe (> 15 m/s), follows this direction. The surrounding region is relatively flat Arctic tundra except for nearby hills (~300 m a.s.l.) approximately two kilometers to the NE of the supersite. The average roughness length determined from the variance of wind speed is 0.14 m (Gordon et al., 2010). The soil type at and around the site is cryosolic and the surface type is ~70% tundra and ~30% ocean within a 10 km radius of the supersite. Most storm tracks that reach Iqaluit originate over the Western Canadian Arctic or the Prairies; these storms can produce strong Easterly winds which frequently cause blowing snow that severely reduces visibility during non-summer months. Given the site's proximity to Frobisher Bay (< 600 m), the site is influenced by sea surface conditions during onshore flow (NW). Iqaluit experiences an extreme range of average monthly temperatures ranging from -28 to 8 ºC and average monthly precipitation ranging from 18 to 70 mm.

2.3 Sodankylä, Finland

The Sodankylä supersite is managed by the Arctic Space Centre of the Finnish Meteorological Institute (FMI-ARC) and is located in the Scandinavian taiga, which consists of a mix of spruces, pines and birches. The measurements at the Sodankylä supersite are distributed over seven main observational sites, each of them including several installations (48m, 24m, 20m or 16m towers, automatic weather stations (AWS), structures supporting snow and soil measurements) that cover an area of

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

approximately 1.5 km^2. The environment of the observational sites varies between dense forest, sparse forest, forest openings, and wetland, each of these environments having its own particular surface characteristics. The supersite is a calibration/validation site for numerous satellite products (such as snow water equivalent and snow extent (Luojus et al., 2021), and soil freeze-thaw (Cohen et al., 2021 and Rautiainen et al., 2016), hence the spatial distribution of the observational sites reflects the need of measuring the spatial variability of observed parameters over different spatial scales and satellite footprints (Hannula et al., 2016).

2.4 Utqiaġvik (formerly Barrow), USA

The Utqiaġvik supersite consists of observatories located ~3 km southeast from the coastline where the Beaufort and Chukchi Seas meet. The supersite is situated over tundra interspersed with thermokarst lakes having a coverage of up to 40% area (Sellmann et al., 1975). The climate in Utqiaġvik, and much of the Alaskan North Slope, is regulated by seasonal sea ice cover and the dominance of easterlies that circulate around the Beaufort High. This atmospheric pattern is punctuated by episodes of southerly advection of air masses from the north Pacific, which frequently arrive from the direction of the Bering Strait and are influential the timing of seasonal transitions of terrestrial snow cover and sea ice coverage in both autumn and spring (Cox et al., 2017).

There are two primary observatories located outside of Utqiaġvik (formerly Barrow), Alaska: The Atmospheric Radiation Measurement (ARM) North Slope of Alaska (NSA) observatory operated by the Department of Energy, and the Barrow Atmospheric Baseline Observatory facility operated by the NOAA Global Monitoring Laboratory (GML). These observatories are located 8 km east of the town of Utqiaġvik, and likely beyond the influence of a local heat island in town (Hinkel et al., 2007) and disturbance to snow cover by human activity (Stone et al., 2002). The site includes several towers and space for guest instruments. The GML Barrow Atmospheric Baseline Observatory recently built a newly furnished on-site laboratory that was completed in 2020. The site's previous facility was constructed in 1972 (https://gml.noaa.gov/obop/brw/history/index.html), and was deconstructed in 2021. The ARM NSA observatory was established in 1997 (Verlinde et al., 2016). Together, the GML and ARM observatories provide an extensive set of long-term measurements at this coastal location. Measurements include properties of aerosols, clouds, precipitation, trace gases, the atmospheric state and the surface energy balance. Radiosondes were launched three times daily during the SOPs specifically in response to a WMO YOPP organizational request.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

188 **2.5 Tiksi, Russia**

189 The original Tiksi science station was established in 1932 and at its height had 60-80 staff and families that lived onsite with
190 a school and grocery store comprising an independent community. The current Tiksi observatory, in the same location, is 7
191 km away from the town of Tiksi, Russia, in the Sakha Republic of northern Siberia and is staffed by personnel that commute
192 from the town. In collaboration with the Russian Federal Service for Hydrometeorological and Environmental Monitoring
193 (Roshydromet), a partnership was established with the National Oceanic and Atmospheric Administration (NOAA) and the
194 Finnish Meteorological Institute (FMI) in 2005 to collect climate grade meteorological, surface energy budget, greenhouse
195 gases and aerosol data (Uttal et al., 2013). The Tiksi station is a coastal site, with facilities built in a high latitude tundra regime,
196 comprising several different types of tundra land classifications including shrub (most predominant), lichen, wet/dry fen,
197 grassy, bog, water, bare and meadow (Mikola et al., 2018).
198
199 On-site, Tiksi hosts a 20-m flux tower, a clean air facility, a weather station, a Climate Reference Network (CRN) platform,
200 and a Baseline Surface Radiation Network (BSRN) platform (Ohmura et al., 1998; Driemel et al., 2018). Radiosonde data
201 were incorporated into the Integrated Global Radiosonde Archive (IGRA) and are available through NOAA's National Centers
202 for Environmental Information (NCEI) portal (Durre et al., 2018). Radiosondes had twice daily launches during the SOPs
203 specifically in response to a WMO YOPP organizational request. Meteorologically, Tiksi is located in a boundary region
204 between Atlantic and Pacific air masses. The resulting variability in atmospheric conditions with air masses originating from
205 various source regions in Russia, Northern America, Europe and Central Asia require careful attention and interpretation of
206 in-situ measurements. Tiksi is also influenced by its location at the mouth of the Lena River, the second largest river draining
207 into the Arctic Ocean and the only major Russian river underlain by permafrost which has impacts on the processes and
208 evolution of surface fluxes. Tiksi is also situated on the coast of the Laptev Sea, which is historically a region of large sea-ice
209 production.

210 **2.6 Ny-Ålesund, Norway**

211 At Ny-Ålesund Research Station in Svalbard, Norway, multi-disciplinary observations are operated by several institutions of
212 different nationalities. The settlement at 78.9°N, 11.9°E, is situated on the south coast of the Kongsfjord, which opens at the
213 west coast of Svalbard towards the Fram Strait. The fjord stretches in southeast-northwest direction from the large glacier
214 plateau to the open ocean, and is surrounded by glaciated mountains with altitudes up to 1 km. This geographical setting
215 impacts the local wind field in the lowermost kilometer, resulting in a mainly southeastern wind direction at Ny-Ålesund,
216 which is temporarily replaced by a north-westerly wind direction when large-scale synoptic wind is also coming from the
217 according direction. Only in calm conditions with wind speed < 2 m/s do katabatic winds from the glaciers south of Ny-
218 Ålesund prevail. Ny-Ålesund may be located in the high Arctic, but due to its location in a coastal environment affected by

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

the West Spitsbergen Current, the local climate is quite maritime and relatively warm. During the summer months, air temperatures above freezing and the otherwise snow-covered landscape exhibits tundra ground and the active layer soil surface of permafrost. An overview of the climate conditions and changes in Svalbard is given by the Norwegian Centre for Climate Services (NCCS, 2018), while the specific atmospheric and radiation conditions in Ny-Ålesund are described by Maturilli et al. (2019).

The Norwegian Meteorological Institute (aka MET Norway; www.met.no) is operating the standard meteorological surface and synoptic observations reported to the WMO. For the YOPP SOPs, the radiosonde launch frequency was increased from daily to 6-hourly. Radiosonde launches, four times daily, are contributed by the Alfred Wegener Institute (AWI), and carried out by the German-French AWIPEV research base that AWI jointly operates with the French Polar Institute Paul-Émile Victor (IPEV). The radiosondes and weekly ozone sondes are launched from a balloon platform about 200m west of the MET Norway weather mast. Atmospheric trace gases and cloud condensation nuclei are observed at the Zeppelin Observatory at about 474 m a.s.l. on Zeppelin Mountain south of Ny-Ålesund, operated by the Norwegian Polar Institute (NPI), the Norwegian Institute for Air Research (NILU), Stockholm University, the Japanese National Institute of Polar Research (NIPR), and others. The full complement of atmospheric measurements at Ny-Ålesund highlights the interwoven research community that contributes to making Ny-Ålesund an observational supersite. More information on the Ny-Ålesund Research Station is available at https://nyalesundresearch.no.

2.7 Eureka, Canada

The CAnadian Network for the Detection of Atmospheric Change (CANDAC) runs the Polar Environment Atmospheric Research Laboratory (PEARL) near the Environment and Climate Change Canada (ECCC) Eureka Weather Station (EWS) in Nunavut, Canada. PEARL has three facilities: the Ridge Laboratory (RL), the Zero Altitude PEARL Auxiliary Laboratory (0PAL), and the Surface and Atmospheric Flux Irradiance Extension (SAFIRE). PEARL collects a wide variety of measurements across all three facilities, and in-depth details about the site including its instrumentation, dataset validation and uncertainties, etc., can be found in Fogal et al. (2013) and at https://www.pearl-candac.ca/website/index.php/facilities. Only a subset of the available measurements collected have been included in the $MODF_{ysm}$ (Akish & Morris, 2023a) due to time constraints and processing resources.

Details of Eureka's climatology are described in Lesins et al. (2010) and water vapor climatology in Weaver et al. (2017). For the period from 1954–2007, the monthly average dry bulb air temperature minimum occurs in February at approximately -37 °C, with the maximum in July at approximately 5 °C. ECCC also publishes climate normals for Eureka at https://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?stnID=1750&autofwd=1, which for a time period

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

of 1981–2010, report a minimum monthly average temperature of -37.4 °C in February and a maximum of 6.1 °C in July. Average yearly precipitation is reported as 79.1 mm, with a yearly average snowfall of 60.3 cm and yearly average rainfall of 32.5 mm. The soils are mostly marine deposits, and the topography, apart from the stony ridges, is driven mostly by ground ice (Pollard & Bell, 1998; Pollard et al., 2015). Eureka is generally colder and drier than Utqiaġvik (Cox et al., 2012). Cloud cover over Eureka is anomalous relative to other Arctic observatories, with generally higher cloud bases, a smaller proportion of supercooled liquid, and a seasonal cycle offset from the typical pattern observed elsewhere (Shupe, 2011; Shupe et al., 2011). Ellesmere Island, where Eureka is situated, is characterized by complex topography that generates mesoscale atmospheric circulations, such as downsloping winds (e.g., Persson and Stone, 2007). The local summertime atmosphere is likely regulated also by nearby ice conditions (Persson and Stone, 2007; Tremblay et al., 2019), which vary between the northern side of the island where multiyear pack ice persists (e.g., Alert) and other coastal areas, which are generally adjacent to seasonal ice cover (e.g., Eureka). However, the general dryness of the atmosphere over Ellesmere is likely a regional anomaly related to location relative to dominant pressure patterns over the Beaufort Sea and near the pole rather than being local (Cox et al., 2012).

The observations used from the Eureka station for the $MODF_{ysm}$ (Akish & Morris, 2023a) were primarily measured at the 0PAL and SAFIRE on-site facilities. The 0PAL lab is situated at approximately 10 m a.s.l. elevation to capture measurements in the lowermost atmosphere. The SAFIRE facility is located about 5 km from the EWS, and it is located away from any structures. At SAFIRE, there is a former BSRN station, a flux tower, and additional remote sensing instrumentation. Eureka increased their twice daily radiosonde launches to four daily launches during the SOPs, specifically in response to a WMO YOPP organizational request.

3 Instrumentation and Derived Variable Calculation

Standard surface meteorological observations (winds, temperature, pressure, humidity, precipitation) were conducted by instruments of similar design, operation, and accuracy at the different sites. The MODF files have an attribute "Instrument," which specifies the exact instrument model used for each variable at each site. OTT Pluvio2 precipitation weighing gauges, which have a quoted precision of $\pm$ 0.001 mm and uncertainty < 5%, were deployed at all sites to measure precipitation with a single Alter shield configuration (no under-catchment corrections were performed; see Section 4). The reported accuracy of the Campbell Scientific probes used at some of the sites to measure soil temperature and moisture is 0.3 K and 1.5%, respectively.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

For Whitehorse and Iqaluit, a Vaisala WXT520 was used to measure wind, air temperature, pressure, and relative humidity with an accuracy of ± 0.3 m/s, ± 0.3 ºC, ± 0.5 hPa, and ± 3%, respectively. The other sites employed slightly different instruments to measure these variables; in general, their reported accuracy is similar or slightly better than the WXT520. Wind observations were collected by an RM Young Model 43408/43482/3001 at Utqiaġvik, Tiksi, and Ny-Ålesund, a Vaisala WAA25 or METEK USA-1 sonic anemometer at Sodankylä, and a Lufft Anemometer at Eureka. Temperature and relative humidity observations were collected by Vaisala HMP35D/HMP45D/HMP155/HMT337 sensors in aspirators at Utqiaġvik, Tiksi, Sodankylä, Ny-Ålesund, and Eureka; they were shielded/housed in the same way. Pressure was obtained from a Vaisala PT100/PTB110/PTB220/PTB201 at Utqiaġvik, Tiksi, Ny-Ålesund, Sodankylä, and Eureka.

For all sites, Vaisala RS92 or RS41 radiosondes were used to collect vertical profile observations from the surface up to the stratosphere. For Iqaluit and Whitehorse, however, the radiosonde manufacturer changed during SOP2 from Vaisala (RS92) to GRAW on September 12, 2018 (no impact on the data quality is anticipated). These radiosondes have a quoted uncertainty of < ± 0.5ºC, 1.0 hPa, 0.15 m/s, and 5% for temperature, pressure, wind, and relative humidity, respectively, in the lower atmosphere.

The radiation flux, cloud base height, and snowfall flux observations are the only derived variables that were explicitly calculated in the MODF (as opposed to the direct observations described in the paragraphs above). The radiation flux observations were processed using the eddy correlation and bulk method (see for instance Baldocchi, 2014). Kipp and Zonen pyranometers and pyrgeometers (e.g., CMP22/CNR4/CM11/CMA11/CGR4 models) were used at Iqaluit, Utqiaġvik, and Sodankylä, whereas an Eppley PSP pyranometer and PIR pyrgeometer was used at Tiksi, Ny-Ålesund, and Eureka. In general, these pyranometers and pyrgeometers have spectral ranges of 200 to 3600 nm (e.g., CMP22) and 4500 to 42000 nm (e.g., CGR4), respectively, a directional error < ±5 W/m², sensitivity of 5-15 µV/W/m² and an offset of < 7 W/m² (night-time for the pyranometer). All upwelling and downwelling, longwave and shortwave radiation measurements were collected at 1-minute intervals with instruments in aspirated housing units and no heating elements applied to the instruments. Additional processing and quality control methods for these observations are discussed in Section 4. Cloud-base height observations were output by the Vaisala CL51 ceilometer at most sites (where available) using a proprietary algorithm to determine the lowest cloud base height; the uncertainty of this algorithm isn't reported but the ceilometer has a reported distance accuracy of ± 5 m from the manufacturer. The snowfall flux data was derived from a Ka-band ARM Zenith Radar (KAZR) used at the ARM facility, following ARM quality control measures (Widener et al., 2012). ARM technical reports, instrument validation / evaluation, and quality control measures are linked and available within the Utqiaġvik/Barrow $MODF_{ysm}$ (Akish & Morris, 2023c).

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

For all observations, instantaneous time is reported at the instruments' raw sampling cadence in UTC. The typical temporal cadence for most observations are around 1 minute or less. No temporal interpolation or averaging was performed on the data. The only exception to this is for turbulent fluxes (the only calculated variable), where some averaging (1 to 30 minutes, depending on the variable) is implicit in the calculation of fluxes. Heights are reported as above ground level (AGL), with the exception of the soil thermistor string, which reports depths below the surface in units of cm. Note that the uncertainties provided in this Section originate from the manufacturer and often depend on the meteorological conditions (e.g., relative humidity observations are less accurate during very low temperatures); as such, the largest reported uncertainty was provided for each geophysical variable in order to provide a conservative error estimate. For more information on the instrumentation used or further details on the instrument accuracy, precision, and co-located validation studies for certain instruments, refer to the site-specific references listed in Section 2 and/or the WMO Guide to Instruments and Methods of Observation (WMO, 2021).

4 Dataset Preparation, quality control, and post-processing

Guidelines for creating MODFs were published as a table in both human-readable (PDF file) and machine-readable (JSON files) formats by Hartten and Khalsa (2022). This "H-K Table" adopts the standards and conventions commonly used in the earth sciences, including NetCDF encoding with Climate and Forecast (CF) Conventions and following CMPI6 naming, as agreed upon by the YOPP community (Uttal et al., 2023). This H-K standard facilitates the creation of MODFs using current requirements and the creator's software of choice, with the MODF toolkits providing tools to assist the user in creating MODFs (Section 6). For the present work, we used H-K Table version 1.3 to guide the criteria for the generation and standardization of naming conventions, units, and global/variable attribute metadata. Observational datasets were collated and formatted for each of the seven supersites into a set of NetCDF files in accordance with the table's criteria. The native variable name is saved as an attribute in the MODFs and as previously discussed, no resampling was performed to harmonize different time stepping (the instrument's instantaneous raw sampling frequency is reported, usually about minutely). Acceptance of data into the $MODF_{ysm}$ was generally determined by the variable list described in the table. The processing script is openly available and described in Section 6.

Radiosonde (timeSeriesProfileSonde variables) data in the MODF were binned into 5-meter intervals (10 m for Iqaluit and Whitehorse) of geopotential height and all measurements within each bin were averaged. In the case of 5-meter intervals, this most often results in 0, 1, or 2 measurements in each bin: 8%, 82%, 9% respectively in SOP1 and 6%, 80%, 13% in SOP2. In both SOP1 and SOP2 at least 99.9% of the measurements have 2 or fewer measurements, but a given bin can have up to 14 measurements. The number of measurements per bin has been included in the dataset to filter for these situations, as have the

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

actual time and height of each measurement (though also averaged within each bin). For surface precipitation observations, no corrections for solid precipitation under-catchment were performed (the dataset is raw in the MODF); where appropriate, users are recommended to process under-catchment corrections via Kochendorfer et al. (2020).

The principal goal of the present phase of the MODF concept is to standardize data organization, metadata, and interoperability. While data quality assurance and measurement operation procedures remain in the purview of the contributing stations, considerable effort was undertaken to ensure MODF production followed a transparent, consistent, and standardized data processing chain. This includes efforts to standardize post-processing and filtering techniques (e.g., quality control methods) as much as possible for the same geophysical variable across the different sites. This consistent processing chain is another unique feature of the MODF dataset as it enforces a level of consistency across vastly different observation sites that normally follow their agencies' own data production procedures and methods. As identified in the below subsections, there are some cases where site-specific data processing could not be avoided; data should be used cautiously and with due consideration to each supersite's processing techniques and quality control (QC) methods for the $MODF_{ysm}$.

4. 1 Whitehorse and Iqaluit, Canada

All geophysical variables observed at the Iqaluit and Whitehorse sites were processed in the same manner and included in the $MODF_{ysm}$ (Huang et al., 2023b; 2023a). For most geophysical variables, limited QC was performed on the raw dataset with the intention to remove obvious outliers only. Surface variables were checked against climatology ranges and the rate of change thresholds, which were based on hourly criteria. A very small number (<5%) of observations were flagged by the QC algorithm. The radiation flux observations should be treated with caution since they typically require additional QC processing prior to analysis; no additional QC was performed on these observations to account for potential frost or snow deposition on the sensors, for instance. No additional QC was performed on the cloud base height data, which was processed by the Vaisala software. Vaisala also processed the raw data feed from the radiosonde observations, which was obtained at 2 s resolution; no additional QC was performed. When no data was available (due to the instrument being down, loss of power at the site, or because it was flagged by the QC algorithm), a missing value (-9999.0) was reported in the $MODF_{ysm}$ (Huang et al., 2023a; 2023b) and is notated via the missing_value attribute associated with each variable. Mariani et al. (2020, 2021) provides instrument validation studies and more detailed information on the quality control processing routines for the remote sensing and upper air observations.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

370 **4. 2 Sodankylä, Finland**

371 The Sodankylä observations included in the $MODF_{ysm}$ (O'Connor, 2023) are automatically uploaded every day to the FMI
372 open access web site https://litdb.fmi.fi/ where the data are organized on the basis of platforms and stations. Before being
373 uploaded to the web page, the data undergo an automatic quality check to remove outliers. In the current $MODF_{ysm}$ version
374 (O'Connor, 2023), no further quality check was applied to the data, implying that errors from several sources (such as
375 snow/frost deposition on radiation and temperature sensors or absorption of solar radiation by unsheltered temperature sensors)
376 are occasionally included. In a future version of the $MODF_{ysm}$, a deeper quality check will be applied to some of the variables
377 included in the current $MODF_{ysm}$ (O'Connor, 2023). This quality check is based on the comparison among the same variables
378 measured at different sites, on visual inspection and, in the case of global radiation, on the comparison with radiative transfer
379 model calculations. This processing will enable the identification of the shortwave data affected by the shadows casted by the
380 vegetation, of errors caused by frost formation on the domes of pyranometers, and of the error in unshaded thermometers
381 caused by the absorption of solar radiation. As in the case of the Eureka observatory, the radiosonde data in the MODF was
382 ingested and processed by IGRA and is available through NOAA's NCEI portal (Durre et al., 2018).

383 **4. 3 Utqiaġvik (formerly barrow), USA**

384 The Utqiaġvik/Barrow data within the $MODF_{ysm}$ (Akish & Morris, 2023c) originated from both Atmospheric Radiation
385 Measurement (DOE/ARM) and the Global Monitoring Laboratory (NOAA GML) datasets, with GML proving datasets for
386 ozone, snow thickness, skin temperature, soil temperature profile. Value added products were generated and disseminated to
387 the users using the ARM Data Discovery interface. Both the ARM and GML datasets were ingested into a single $MODF_{ysm}$
388 with variable attribution detailing how each variable and data set was quality controlled, processed and accessed. The surface
389 ozone data was collected in 1-minute intervals and was manually quality controlled and submitted to NCEI.
390
391 The measurements collected by the ARM facility were processed, QC analyzed, and archived at the ARM Data Center archive.
392 The radiation measurements were QC'd and processed following Long & Shi (2008). Heat fluxes were processed and QC'd
393 via Eddy correlation corrections including stability correction, Webb-Pearman correction, frequency correction, sensor
394 separation correction, filtering correction, line-averaging correction, and volume-averaging correction (Cook et al. 2008,
395 Fuehrer and Friehe 2002). Bulk corrections were also employed and utilized ARM data from the radiation, ground, met, and
396 tower. Radiosonde data were ingested and processed by NOAA's NCEI and was processed through IGRA, following their
397 standards (Durre et al., 2018) and is available through NOAA's NCEI portal. The IGRA 2 QA system processed the sonde
398 data, which is based largely on the QA procedures in the IGRA 1 system (Durre et al. 2006; Durre et al. 2008). Like the IGRA
399 1 system, it consists of a deliberate sequence of specialized algorithms, each of which makes a binary decision on the quality
400 of a value, level, or sounding; either the data item passes the check and remains available, or it is identified as erroneous and

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

thus set to missing. For all other observations' QC, a first level automated QC was established by climatology ranges in the same way as for Whitehorse and Iqaluit. A second level of QC was performed whereby data was reviewed by instrument mentors and assessed by the site scientist/data quality office.

4. 4 Tiksi, Russia and Eureka, Canada

Data collection and processing techniques for Eureka are the same as for the Tiksi site. The long-term Eureka and Tiksi datasets (flux tower and radiation) are hosted by the NOAA Physical Sciences Laboratory (PSL), in collaboration with ECCC (Eureka site only), and Roshydromet (Tiksi site only). All meteorological measurements within the $MODF_{ysm}$ (Akish & Morris, 2023b), i.e., air temperature, skin temperature, soil temperature, snow thickness, pressure, relative humidity, wind speed and direction, were manually quality controlled first via an automated QC established by climatological ranges in the same way as for Whitehorse et al. Following this, a manual/visual inspection was performed. This included removing non-physical values and outliers, after confirming that they were either biased, incorrect, or collected during site maintenance periods. The radiation measurements were validated and processed using the Long QCrad method (Long & Shi, 2008) and improved correction of the infrared loss in diffuse shortwave measurements (Younkin & Long, 2004), and again, were visually inspected. The radiosonde dataset was processed through IGRA's processing techniques and is based on the QC procedures in the IGRA 1 system (Durre et al., 2006; Durre et al., 2008). If data was not available for any of the collected measurements across any of the variables, due to the instrument being down, loss of power at the site, or because it was flagged by the QC algorithm, a missing value (-9999) was reported in the $MODF_{ysm}$ (Akish & Morris, 2023b).

4. 5 Ny-Ålesund, Norway

The meteorological measurements used for the $MODF_{ysm}$ (Holt, 2023) are taken from the AWIPEV weather mast (Driemel et al., 2018; Maturilli, 2020b). Except for precipitation, all other data used in the $MODF_{ysm}$ for Ny-Ålesund originated from the following data sets: Maturilli, 2020a, 2020b, 2020c, 2022. The precipitation data reported in the $MODF_{ysm}$ are the direct instrument output and no quality checks were applied; as such this data should be treated with caution (Holt, 2023). The Ny-Ålesund observations included in the $MODF_{ysm}$ are a subset of those regularly uploaded in the PANGAEA data repository (www.pangaea.de). Before being uploaded, all data undergo an automatic quality check established by climatological ranges in the same way as for Whitehorse et al. Following this, additional manual/visual inspection was performed accounting for e.g., physical plausibility. Surface radiation data were validated and have undergone all quality checks of BSRN before archiving (Maturilli, 2020a). Automated QC was performed on the radiosonde data, established by climatological norms; a second level of data was reviewed by the instrument mentor before storing the data at the PANGAEA repository.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

5 MODF Data Structure

The data inside a MODF comprises of all the observations listed in Table 3 for a given observation site. The data itself follows the same standardized format and structure for all observations and sites and is stored into a single NetCDF file using CF conventions. NetCDF file formatting was chosen to best accommodate the high-level of metadata detail required for merging such large quantities of individual measurements together, particularly given the need to be as transparent as possible when reporting instrument-specific details for each observation. All $MODF_{ysm}$ measurements provided in the data files maintained their native time cadence (typically on the order of minutely) with no averaging undertaken, and details of the collection and processing techniques can be found in the variable attributes within the files. Each DOI in Table 2 contains four (e.g., Whitehorse) or six (e.g., Utqiaġvik) files, depending on whether the site had timeSeriesProfile observations on a tower/mast. The filename convention for each MODF is as follows: site name + "obs" + MODF_featureType + start_date + end_date.nc.

Guidelines for creating inventories of variable and attribute information (metadata) necessary for the MODF file attributes were published in spreadsheet format by Morris and Akish (2022). This "A-M Template" uses variable content criteria from the H-K Table to generate a metadata matrix of attribute and variable information for each of the measurements contained within the MODFs. The template has individual tabs for each of the corresponding CF metadata featureTypes (i.e., timeSeries and timeSeriesProfile) of the MODF NetCDF files, as well as one tab for the Global Attributes of the MODFs. The CF Conventions can be found here: https://cfconventions.org/cf-conventions/cf-conventions.html. The attributes within the template are mandatory when applicable, and serve as a guideline for MODF creators. The A-M Template is machine-readable and can be ingested into MODF software to create the final output.

The file content is well-illustrated in Table 3; other details of the $MODF_{ysm}$ format and structure are outlined in Uttal et al. (2023). MODFs can contain featureTypes such as timeSeries and timeSeriesProfile, which refer to time series having one and two data dimensions, respectively. In cases where data subcategories exist, featureType modifications can be depicted in the file name, for example timeSeriesProfileSonde exist for the $MODF_{ysm}$. Currently, more than one featureType can be used within an individual MODF file, but all subscribe to the same formatting structure and nomenclature. To generate an MODF, creators would first visit the H-K Table to determine the variables that will be included in their MODF, and then they should utilize the A-M Template to fill in the needed attribute and variables information requested by existing MODF software. Once the A-M Template has been completed, then users can ingest the template into their MODF software to create the final MODF outputs. For the $MODF_{ysm}$, individual toolkits were developed by MODF makers for each YOPP supersite. Python code was developed for Whitehorse, Iqaluit and Ny-Alesund, and MATLAB code for Utqiagvik, Tiksi, Eureka and Soldankyla (see Section 6). After the generation of the $MODF_{ysm}$ outputs, the files were run through an MODF checker that identifies the various inconsistencies or issues with the files before their upload to the MET Norway data portal. The $MODF_{ysm}$ checker developed for the YOPPsiteMIP files is part of a larger toolkit being designed to continue the creation of MODFs.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

As an example of the uniformity of the observations (in terms of data format, post-processing, temporal cadence, etc.) contained within each supersite's $MODF_{ysm}$ and their excellent data coverage during the two YOPP SOPs, Figures 3 and 4 provide the surface downwelling longwave radiation and near-surface temperature observations from each supersite's $MODF_{ysm}$ during SOP1, respectively and Figures 5 and 6 show the same except for SOP2. The MET Norway data portal and MODF maker toolkit (Sect. 6) also provides plotting tools that work with any MODF or MMDF and can produce similar figures automatically. Periods of interest can be quickly identified by users and analyzed for further investigation and/or comparison with their corresponding MMDFs. MODFs significantly simplify the process of analyzing observations from multiple sites and multiple instruments, as analyses and Figures can be produced for each site using a single code that works for any observed geophysical variable and (if desired) their corresponding NWP model output in the MMDF. In contrast, without MODFs a user would have to contact each meteorological agency individually, find each sites' data repository, obtain data access privileges, find the files they need from multiple instruments, reprocess and reformat multiple uniquely-formatted datasets and file types, then develop several different codes (e.g., readers) specific to each instruments' dataset to ingest the multi-variate datasets and plot them.

The $MODF_{ysm}$ at Sodankylä are unique in that their measurements are collected across a series of sub-sites in the area; therefore, it is important to describe here the possible methods for extracting the data for specific locations, or for co-located measurements. The Sodankylä station comprises at least 25 distinct locations, the precise number of which is given by the dimension 'site_id' inside the MODF data file. Each distinct location is given a unique index key in the variable 'subsite_name', with these indices also identifying the 'lat', 'lon' and 'soil_type' for each location. The corresponding FMI names for each location are identified in the attribute 'flag_meanings' for the variable 'subsite_name' via their indices; for example, the index value of 16 pointing to IOA003_spot_8, which is one of the automatic weather stations located in the Intensive Observations Area (IOA). There may be multiple locations providing the same measurement. However, not all locations provide the same set of measurements, and to keep the MODF compact, each measurement variable has the location dimension truncated to include only locations which measure that variable; i.e., the location dimension for the measurement variables is 'nsubsites_X', where X is the number of locations making the particular measurement. This set of locations is accessed through the indices given in the attribute 'subsite_name' for the measurement variable, which corresponds to the key given in the 'subsite_name' variable; i.e., a subsite_name attribute of "1, 3, 10" means that these measurements were made at the locations identified by their indices, from which their locations (latitudes and longitudes) and soil_type can also be determined.

This method permits diverse options of collecting measurements for particular uses. All measurements, for example, at one location can be obtained by identifying the appropriate 'subsite_name' index inside the MODF data file, iterating through the 'subsite_name' attribute of each variable to see if it contains the selected index, and, if so, selecting the column or slice of data

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

for the data that matches the location of the index (i.e. if subsite_name = 10 and the subsite_name attribute for a timeSeries variable is "1, 3, 10", the measurement timeSeries for the requested location is in the third column, the next variable may have a subsite_name attribute of "1, 3, 5, 6, 10" and the measurement timeSeries for the requested location is in the fifth column). The user could also select a specific area of interest and identify all measurements made within this region as follows: select the indices for the locations within a specified latitude and longitude range, then iterate through the 'subsite_name' attribute of each variable to see if it contains the selected indices and return the columns or slices that match them.

Note that each supersite conducts additional observations not listed in table 3 that will be included in upcoming updates to the $MODF_{ysm}$ with the intent to eventually incorporate all observations into the $MODF_{ysm}$ for each supersite. This process of developing and appending to MODFs can be extended to other sites and/or research programs that wish to create MODFs of their observations. Given the standardized nature of the MODFs, reading and analyzing datasets from any of the YOPP supersites is simplified. Quick-look plotting tools have been developed via the MET Norway YOPP data portal and the MODF maker toolkit (Sect. 6), which enable near-instantaneous plotting of the observations contained within the $MODF_{ysm}$.

6 Data and Code Availability

The $MODF_{ysm}$ for each supersite are available via the MET Norway YOPP Data Portal (https://yopp.met.no/) where they are indexed through FAIR compliant discovery metadata and can be directly accessed at: https://thredds.met.no/thredds/catalog/alertness/YOPP_supersite/obs/catalog.html (Whitehorse: https://doi.org/10.21343/a33e-j150, Iqaluit: https://doi.org/10.21343/yrnf-ck57, Sodankylä: https://doi.org/10.21343/m16p-pq17, Utqiaġvik: https://doi.org/10.21343/a2dx-nq55, Tiksi: https://doi.org/10.21343/5bwn-w881, Ny-Ålesund: https://doi.org/10.21343/y89m-6393, Eureka: https://doi.org/10.21343/r85j-tc61).

Proper data citation ensures appropriate credits to authors of both input data sources and merged $MODF_{ysm}$ datasets. Data from each station has been assigned a DOI. The variable attributes of the merged data products contain information about the source datastreams and their DOIs, to more clearly establish data provenance in a traceable manner. When using data from the $MODF_{yms}$, it is expected that the user references the $MODF_{ysm}$ DOI, and any subsidiary variable DOIs when available. Assigning citations for merged data streams such as the $MODF_{ysm}$ is a challenging and still evolving concept. For example, the US Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Program uses a combination of DOI and citation structure for continuous data streams, as outlined in Prakash et al. (2016). They recommend when registering DOIs for derived and higher-order data, source DOIs in the metadata of the newly created DOI should be added and linked when possible.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

The source code used to produce the $MODF_{ysm}$ for each supersite (and MODFs in general) are available via gitlab: https://gitlab.com/mdf-makers/mdf-toolkit. This MODF toolkit is openly available for anyone interested in developing their own MODF file or generating quick-look plots of the data contents inside the MODFs. The toolkit is regularly updated as the MODF community grows and new geophysical variables and/or functions are added. Additional site-specific python and MATLAB codes that were used to prepare the observation data files for MODF ingestion are available upon request (e.g., contact the site principle investigator).

7 Concluding Remarks

The enhanced ground-based observations conducted at both Poles during the YOPP fill significant and identified gaps in our current meteorological observation capabilities for the Polar Regions. YOPPsiteMIP MODFs ($MODF_{ysm}$) have been published for seven of the YOPP Arctic supersites, whereby all geophysical variables are stored in an identical, standardized format in a single NetCDF file following CF conventions, fulfilling a key objective of the program to perform single- or multi-variate model-observation comparisons. These MODFs archive data in a manner as similar as possible to corresponding MMDF (see Uttal et al., 2023) that contain high-resolution forecast variables from a single NWP model at and around a supersite (Figure 2). Thus, combined, MODFs and MMDFs greatly simplify integration of these complex datasets, enabling further scientific study as demonstrated in the recent publications using the latest $MODF_{ysm}$ and $MMDF_{ysm}$ (Day et al., 2023).

Standardized geophysical variable nomenclature, cadences, metadata, basic QC, and file structure were employed to create these files. MODFs provide the first standardized files for archiving all the different ground-based observation supersite observations, containing a multitude of geophysical variables observed by (at times) different instruments. This amalgamation of different sites' observations into a standardized, user-friendly MODF format enables easier analysis of the MODF dataset, inter-site comparisons, and detailed NWP model validation, evaluation, intercomparisons, and process-based diagnostic studies that are currently underway (see Figures 3 to 6 as an example). The further adoption, creation, and use of MODFs outside of YOPP is encouraged; a suite of tools and documentation is openly available via Gitlab (see Sect. 6) for other site managers, researchers, and users to develop and create their own site-specific MODFs outside of YOPP or to analyze an observation sites' dataset.

The YOPP $MODF_{ysm}$ discussed here provide novel access to datasets of enhanced meteorological observations collected at several supersites across the Arctic. The MODF concept is not limited for use in polar regions and could be exported elsewhere. Seven YOPP-designated supersites in the Arctic developed and published $MODF_{ysm}$ covering both SOP periods (February –

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

March 2018 and July – September 2018), including Iqaluit, Whitehorse, and Eureka in Canada, Utqiaġvik in the United States, Tiksi in Russia, Sodankylä in Finland, and Ny-Ålesund in Norway. Additional geophysical variables observed at each of these seven supersites will be included in a future update of their $MODF_{ysm}$, with the goal of having 100% of a site's observations available. Observations at most of these sites continue today beyond YOPP and are available for subsequent analyses, in some cases using updated MODFs generated in near-real time. $MODF_{ysm}$ for the other YOPP sites, including ship-based platforms and supersites in the Antarctic, will be made available in the future to complete the YOPP dataset. The $MODF_{ysm}$ described here directly ties to process-oriented verification studies aiming to improve NWP predictions at the Poles by contributing and enabling NWP inter-comparisons.

Author contributions

SM, ZM, and TU wrote the first draft of the manuscript. SM and ZM conducted scientific analyses and created tables and figures with JD and JT. All authors managed data archiving, creation of the $MODF_{ysm}$, and publication to the MET Norway YOPP Data Portal. All authors contributed to the writing and the editing of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Disclaimer

Use of specific instrument manufacturers/models and suppliers mentioned in the manuscript and/or used at the supersites is not a commercial endorsement of their products.

Acknowledgements

This is a contribution to the Year of Polar Prediction (YOPP), a flagship activity of the Polar Prediction Project (PPP), initiated by the World Weather Research Programme (WWRP) of the World Meteorological Organisation (WMO). We acknowledge the WMO WWRP for its role in coordinating this international research activity. This study was supported by NOAA's Global Ocean Monitoring and Observing Program through the Arctic Research Program (FundRef:

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

https://doi./org/10.13039/100018302). Special thanks to the station technicians and operators at the supersites for deploying instruments, maintenance, and technical services. In particular, thank you to the radiosonde operators for providing extra daily sonde launches during the two SOP periods. Thank you to Jenn Glaser for her contract work in creating the station graphic in Figure 1, and to Kyrie Newby and Kalvin Jesse for updating the Google Earth images in Figure 2. JD was supported by the European Union funded INTERACTIII project (Grant Agreement: 871120). AK and LMH were supported in part by NOAA cooperative agreements NA17OAR4320101 and NA22OAR4320151. Potions of the $MODF_{ysm}$ data were obtained from the Atmospheric Radiation Measurement (ARM) user facility, a U.S. Department of Energy (DOE) office of science user facility managed by the biological and environmental research program. Thank you to MET Norway for hosting the YOPP data portal. All data products are produced by their respective institutions and are available via the YOPP data portal (https://yopp.met.no) and directly at: https://thredds.met.no/thredds/catalog/alertness/YOPP_supersite/obs/catalog.html.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

References

Akish, E., & Morris, S.: MODF for Eureka, Canada, during YOPP SOP1 and SOP2, Norwegian Meteorological Institute, https://doi.org/10.21343/R85J-TC61, 2023a.

Akish, E., & Morris, S.: MODF for Tiksi, Russia, during YOPP SOP1 and SOP2, Norwegian Meteorological Institute, https://doi.org/10.21343/5BWN-W881, 2023b.

Akish, E., & Morris, S.: MODF for Utqiagvik, Alaska, during YOPP SOP1 and SOP2, Norwegian Meteorological Institute, https://doi.org/10.21343/A2DX-NQ55, 2023c.

Baldocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere – the state and future of the eddy covariance method. Global Change Biology (2014)20, 3600–3609, https://doi.org/10.1111/gcb.12649, 2014.

Cassano, J. J., Higgins, M. E., and Seefeldt, M. W.: Performance of the Weather Research and Forecasting Model for Month-Long Pan-Arctic Simulations, Monthly Weather Review, 139, 3469-3488, 10.1175/mwr-d-10-05065.1, 2011.

Cohen, J., Rautiainen, K., Lemmetyinen, J., Smolander, T., Vehvilainen, J., and Pulliainen, J.: Sentinel-1 based soil freeze/thaw estimation in boreal forest environments, Remote Sens Environ, 254, ARTN 112267 10.1016/j.rse.2020.112267, 2021.

Cox, C.J., Stone, R.S., Douglas, D.C., Stanitski, D.M., Divoky, G.J., Dutton, E.S., Sweeney, C., George, J.C., and Longenecker, D.U.: Drivers and Environmental Responses to the Changing Annual Snow Cycle of Northern Alaska, B Am Meteorol Soc, 98, 2559-2577, https://doi.org/10.1175/BAMS-D-16-0201.1, 2017.

Cox, C.J., Walden, V.P., and Rowe, P.M.: A Comparison of the atmospheric conditions at Eureka, Canada, and Barrow, Alaska (2006-2008), J Geophys Research, 117, https://doi.org/10.1029/2011JD017164, 2012.

Day, J.J., Sandu, I., Magnusson, L., Rodwell, M.J., Lawrence, H., Bormann, N., and Jung, T.: Increased Arctic influence on the midlatitude flow during Scandinavian Blocking episodes, Q.J.R. Meteorol. Soc., 725, 3846-3862, https://doi.org/10.1002/qj.3673, 2019.

Day, J., Svensson, G., Casati, B., Uttal, T., Khalsa, S.J., Bazile, E., Akish, E., Azouz, N., Ferrighi, L., Frank, H., Gallagher, M., Godoy, Ø., Hartten, L., Huang, L., Holt, J., Di Stefano, M., Mariani, Z., Morris, S., O'Connor, E., Pirazzini, R., Remes, T., Fadeev, R., Solomon, A., Tjerström, J., and Tolstykh, M.: The YOPP site Model Intercomparison Project (YOPPsiteMIP) phase 1: project overview and Arctic winter forecast evaluation, 2023, *submitted to Geoscientific Model Development (GMD) August 25, 2023 – under review*.

Driemel A, Augustine JA, Behrens K, Colle S, Cox C, Cuevas-Agulló E, Denn FM, Duprat T, Fukuda M, Grobe H, Haeffelin M, Hyett N, Ijima O, Kallis A, Knap W, Kustov V, Long CN, Longenecker D, Lupi A, Maturilli M, Mimouni M, Ntsangwane L, Ogihara H, Olano X, Olefs M, Omori M, Passamani L, Pereira EB, Schmithüsen H, Schumacher S, Sieger R, Tamlyn J, Vogt R, Vuilleumier L, Xia X, O A, König-Langlo G. Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017) Earth Syst Sci Data, 10, 1491–1501, 2018.

Durre, I., Menne, M. J., and Vose, R. S.: Strategies for evaluating quality assurance procedures, J Appl Meteorol Clim, 47, 1785-1791, 10.1175/2007jamc1706.1, 2008.

Durre, I., Vose, R. S., and Wuertz, D. B.: Overview of the Integrated Global Radiosonde Archive, J Climate, 19, 53-68, Doi 10.1175/Jcli3594.1, 2006.

Durre, I., Yin, X., Vose, R. S., Applequist, S., and Arnfield, J.: Enhancing the Data Coverage in the Integrated Global Radiosonde Archive. J. Atmos. Oceanic Technol., 35, 1753–1770, https://doi.org/10.1175/JTECH-D-17-0223.1, 2018.

Fogal, P. F., LeBlanc, L. M., and Drummond, J. R.: The Polar Environment Atmospheric Research Laboratory (PEARL): Sounding the Atmosphere at 80 degrees North, Arctic, 66, 377-386, 2013.

Gallagher & Tjernström: Accelerating research in weather prediction and model improvement with new free community open source software tools. To be submitted, 2024.

Goessling, H. F., Jung, T., Klebe, S., Baeseman, J., Bauer, P., Chen, P., Chevallier, M., Dole, R., Gordon, N., Ruti, P., Bradley, A., Bromwich, D. H., Casati, B., Chechin, D., Day, J. J., Massonnet, F., Mills, B., Renfrew, I., Smith, G., and Tatusko, R.: Paving the Way for the Year of Polar Prediction, B Am Meteorol Soc, 97, Es85-Es88, 10.1175/Bams-D-15-00270.1, 2016.

Gordon, M., Biswas, S., Taylor, P. A., Hanesiak, J., Albarran-Melzer, M., and Fargey, S.: Measurements of Drifting and Blowing Snow at Iqaluit, Nunavut, Canada during the STAR Project, Atmos Ocean, 48, 81-100, 10.3137/Ao1105.2010, 2010.

Hannula, H. R., Lemmetyinen, J., Kontu, A., Derksen, C., and Pulliainen, J.: Spatial and temporal variation of bulk snow properties in northern boreal and tundra environments based on extensive field measurements, Geosci Instrum Meth, 5, 347-363, 10.5194/gi-5-347-2016, 2016.

Hartten, L. M. and Khalsa, S. J. S.: The H-K Variable SchemaTable developed for the YOPPsiteMIP (1.2), Zenodo, https://doi.org/10.5281/zenodo.6463464, 2022.

Hinkel, K.M. and Nelson, F.E.: Anthropogenic heat island at Barrow, Alaska, during winter: 2001-2005, J Geophys Research, 112, https://doi.org/10.1029/2006JD007837, 2007.

Holt, J.: Merged Observatory Data File (MODF) for Ny Alesund, Norwegian Meteorological Institute, https://doi.org/10.21343/Y89M-6393, 2023.

Huang, L., Mariani, Z., & Crawford, R.: MODF for Erik Nielsen Airport, Whitehorse, Canada during YOPP SOP1 and SOP2, Norwegian Meteorological Institute, https://doi.org/10.21343/A33E-J150, 2023a.

Huang, L., Mariani, Z., & Crawford, R.: MODF for Iqaluit Airport, Iqaluit, Nunavut, Canada during YOPP SOP1 and SOP2, Norwegian Meteorological Institute, https://doi.org/10.21343/YRNF-CK57, 2023b.

Illingworth, A. J., Cimini, D., Gaffard, C., Haeffelin, M., Lehmann, V., Lohnert, U., O'Connor, E. J., and Ruffieux, D.: Exploiting Existing Ground-Based Remote Sensing Networks to Improve High-Resolution Weather Forecasts, B Am Meteorol Soc, 96, 2107-2125, 10.1175/Bams-D-13-00283.1, 2015.

Joe, P., Melo, S., Burrows, W. R., Casati, B., Crawford, R. W., Deghan, A., Gascon, G., Mariani, Z., Milbrandt, J., and Strawbridge, K.: The Canadian Arctic Weather Science Project Introduction to the Iqaluit Site, B Am Meteorol Soc, 101, E109-E128, 10.1175/Bams-D-18-0291.1, 2020.

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H. F., Holland, M., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P., Perovich, D., Reid, P., Renfrew, I. A., Smith, G., Svensson, G., Tolstykh, M., and Yang, Q. H.: Advancing Polar Prediction Capabilities on Daily to Seasonal Time Scales, B Am Meteorol Soc, 97, 1631-+, 10.1175/Bams-D-14-00246.1, 2016.

Kochendorfer, J., M. Earle, D. Hodyss, A. Reverdin, Y-A. Roulet, R. Nitu, R. Rasmussen, S. Landolt, S. Buisan, and T. Laine: Undercatch Adjustments for Tipping-Bucket Gauge Measurements of Solid Precipitation. J. Hydrometeor., 21, 1193–1205, https://doi.org/10.1175/JHM-D-19-0256.1, 2020.

Koltzow, M., Casati, B., Bazile, E., Haiden, T., and Valkonen, T.: An NWP Model Intercomparison of Surface Weather Parameters in the European Arctic during the Year of Polar Prediction Special Observing Period Northern Hemisphere 1, Weather Forecast, 34, 959-983, 10.1175/Waf-D-19-0003.1, 2019.

Lawrence, H., Bormann, N., Sandu, I., Day, J., Farnan, J., and Bauer, P.: Use and impact of Arctic observations in the ECMWF Numerical Weather Prediction system, Q J Roy Meteor Soc, 145, 3432-3454, 10.1002/qj.3628, 2019.

Lesins, G., Duck, T. J., and Drummond, J. R.: Climate trends at Eureka in the Canadian high arctic, Atmos Ocean, 48, 59-80, 10.3137/AO1103.2010, 2010.

Long, C. N. and Shi, Y.: An Automated Quality Assessment and Control Algorithm for Surface Radiation Measurements, Open Atmospheric Science Journal, 23-37, 10.2174/1874282300802010023, 2008.

Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Mortimer, C., Derksen, C., Mudryk, L., Moisander, M., Hiltunen, M., Smolander, T., Ikonen, J., Cohen, J., Salminen, M., Norberg, J., Veijola, K., and Venalainen, P.: GlobSnow v3.0 Northern Hemisphere snow water equivalent dataset, Sci Data, 8, ARTN 163, 10.1038/s41597-021-00939-2, 2021.

Mariani, Z., Crawford, R., Casati, B., and Lemay, F.: A Multi-Year Evaluation of Doppler Lidar Wind-Profile Observations in the Arctic, Remote Sens-Basel, 12, ARTN 323, 10.3390/rs12020323, 2020.

Mariani, Z.; Hicks-Jalali, S.; Strawbridge, K.; Gwozdecky, J.; Crawford, R.W.; Casati, B.; Lemay, F.; Lehtinen, R.; Tuominen, P. (2021). Evaluation of Arctic Water Vapor Profile Observations from a Differential Absorption Lidar. Remote Sens. 2021, 13, 551. https://doi.org/10.3390/rs13040551, 2021.

Mariani, Z., Dehghan, A., Gascon, G., Joe, P., Hudak, D., Strawbridge, K., and Corriveau, J.: Multi-Instrument Observations of Prolonged Stratified Wind Layers at Iqaluit, Nunavut, Geophys Res Lett, 45, 1654-1660, 10.1002/2017gl076907, 2018.

Mariani, Z., Hicks-Jalali, S., Strawbridge, K., Gwozdecky, J., Crawford, R. W., Casati, B., Lemay, F., Lehtinen, R., and Tuominen, P.: Evaluation of Arctic Water Vapor Profile Observations from a Differential Absorption Lidar, Remote Sens-Basel, 13, ARTN 551, 10.3390/rs13040551, 2021.

Mariani, Z., Huang, G., Crawford, R., Blanchet, J. P., Hicks-Jalali, S., Mekis, E., Pelletier, P., Rodriguez, P., and Strawbridge, K.: Enhanced automated meteorological observations at the Canadian Arctic weather science (CAWS) supersites, Earth System Science Data, 2022.

Maturilli, M.: Basic and other measurements of radiation at station Ny-Ålesund (2006-05 et seq). Alfred Wegener Institute - Research Unit Potsdam, PANGAEA, https://doi.org/10.1594/PANGAEA.914927, 2020a.

Maturilli, M.: Continuous meteorological observations at station Ny-Ålesund (2011-08 et seq). Alfred Wegener Institute - Research Unit Potsdam, PANGAEA, https://doi.org/10.1594/PANGAEA.914979, 2020b.

Maturilli, M.: High resolution radiosonde measurements from station Ny-Ålesund (2017-04 et seq). Alfred Wegener Institute - Research Unit Potsdam, PANGAEA, https://doi.org/10.1594/PANGAEA.914973, 2020c.

Maturilli, M.: Ceilometer cloud base height from station Ny-Ålesund (2017-08 et seq). Alfred Wegener Institute - Research Unit Potsdam, PANGAEA, https://doi.org/10.1594/PANGAEA.942331, 2022.

Maturilli, M., Hanssen-Bauer, I., Neuber, R., Rex, M., and Edvardsen, K.: The Atmosphere above Ny-Ålesund – Climate and global warming, ozone and surface UV radiation / Hop, H. and Wiencke, C. (editors), Advances in Polar Ecology, The Ecosystem of Kongsfjorden, Svalbard, Springer, ISBN: 978-3-319-46423-7, doi:10.1007/978-3-319-46425-1_2, 2019.

Mikola, J., Virtanen, T., Linkosalmi, M., Vaha, E., Nyman, J., Postanogova, O., Rasanen, A., Kotze, D. J., Laurila, T., Juutinen, S., Kondratyev, V., and Aurela, M.: Spatial variation and linkages of soil and vegetation in the Siberian Arctic tundra - coupling field observations with remote sensing data, Biogeosciences, 15, 2781-2801, 10.5194/bg-15-2781-2018, 2018.

Morris, S. M. and Akish, E.: A-M Variable & Attribute Template Table developed for the YOPPsiteMIP (1.2), Zenodo, https://doi.org/10.5281/zenodo.6974550, 2022.

NCCS: Climate in Svalbard 2100 – a knowledge base for climate adaptation, 2018.

O'Connor, E.: Merged observation data file for Sodankyla, Norwegian Meteorological Institute, https://doi.org/10.21343/M16P-PQ17, 2023.

Ohmura, A., Dutton, E.G., Forgan, B., Frohlich, C., Gilgen, H., Hegner, H., Heimo, A., Konig-Langlo, G., McArther, B., Muller, G., Philipona, R., Pinker, R., Whitlock, C.H., Dehne, K., and Wild, M.: Baseline Surface Radiation Network (BSRN/WCRP): New Precision Radiometry for Climate Research, B Am Meteorol Soc, 79, 2115-2136, https://doi.org/10.1175/1520-0477(1998)079<2115:BSRNBW>2.0.CO;2, 1998.

Persson, O. and Stone, R.: Evidence of forcing of Arctic regional climates by mesoscale processes, AMS Symposium on Connection Between Mesoscale Processes and Climate Variability, San Antonio, Texas, 15-16 January 2007, 2.6, https://ams.confex.com/ams/87ANNUAL/techprogram/paper_119015.htm, 2007.

Pinard, J. D. J. P., Benoit, R., and Yu, W.: A WEST wind climate simulation of the mountainous Yukon, Atmos Ocean, 43, 259-282, DOI 10.3137/ao.430306, 2005.

Pollard, W. H. and Bell, T.: Massive Ice Formation in the Eureka Sound Lowlands: A Landscape Model, PERMAFROST - Seventh International Conference, Yellowknife, Canada, Collection Nordicana, 1998.

Pollard, W. H., Ward, M. A., and Becker, M. S.: The Eureka Sound lowlands: an ice-rich permafrost landscape in transition, Dept. of Geography, McGill University, 2015.

Prakash, G., Shrestha, B., Younkin, K., Jundt, R., Martin, M., and Elliott, J.: Data Always Getting Bigger—A Scalable DOI Architecture for Big and Expanding Scientific Data, 1, 11, 2016.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvarinen, O., Ruosteenoja, K., Vihma, T., and Laaksonen, A.: The Arctic has warmed nearly four times faster than the globe since 1979, Commun Earth Environ, 3, ARTN 168 10.1038/s43247-022-00498-3, 2022.

Rautiainen, K., Parkkinen, T., Lemmetyinen, J., Schwank, M., Wiesmann, A., Ikonen, J., Derksen, C., Davydov, S., Davydova, A., Boike, J., Langer, M., Drusch, M., and Pulliainen, J.: SMOS prototype algorithm for detecting autumn soil freezing, Remote Sens Environ, 180, 346-360, 10.1016/j.rse.2016.01.012, 2016.

Sellmann, P.V., Brown, J., Lewellen, R., McKim, H.L., Merry, C.J.: The classification and geomorphic implications of thaw lakes on the Arctic coastal plain, Alaska. Cold Regions Research and Engineering Laboratory (CRREL); CRREL-No. 344, https://hdl.handle.net/11681/5852, 1975.

Shupe, M.D.: Clouds at Arctic Atmospheric Observatories. Part II: Thermodynamic Phase Characteristics, J Appl Meteorol Clim, 50, 645-661, https://doi.org/10.1175/2010JAMC2468.1, 2011.

Shupe, M.D., Walden, V.P., Eloranta, E., Uttal, T., Campbell, J.R., Starkweather, S.M., and Shiobara, M.: Clouds at Arctic Atmospheric Observatories. Part I: Occurrence and Macrophysical Properties, J Appl Meteorol Clim, 50, 626-644, https://doi.org/10.1175/2010JAMC2467.1, 2011.

Stone, R.S., Dutton, E.G., Harris, J.M., and Longenecker, D.: Earlier spring snowmelt in northern Alaska as an indicator of climate change, J Geophys Research, 107, https://doi.org/10.1029/2000JD000286, 2002.

Tremblay, S., Picard, J.-C., Bachelder, J. O.,Lutsch, E., Strong, K., Fogal, P., Leaitch, W. R., Sharma, S., Kolonjari, F., Cox, C. J., Chang, R. Y.-W. and Hayes, P. L.: Characterization of aerosol growth events over Ellesmere Island during the summers of 2015 and 2016, Atmos. Chem. Phys., 19, 5589-5604, 10.5194/acp-19-5589-2019.

Uttal, T., Makshtas, A. and Laurila, T.: The Tiksi International Hydrometeorological Observatory - An Arctic Members Partnership, WMO Bulletin Vol 62 (1) – 2013, 2013.

Uttal, T., Hartten, L.M., Khalsa, S.J., Casati, B., Svensson, G., Day, J., Gallagher, M., Holt, J., Akish, E., Morris, S., O'Connor, E., Pirazzini, R., Huang, L., Crawford, R., Mariani, Z., Godoy, Ø., Tjernström, J.A.K., Prakesh, G., Hickmon, N., Maturilli, M., and Cox, C.: Merged Observatory Data Files (MODFs): An Integrated Research Data Product Supporting Process Oriented Investigations and Diagnostics, 2023, *submitted to Model Intercomparison and Improvement Projects (MIIPs) for the polar regions and beyond (GMD/ESSD inter-journal SI) October 17, 2023 – under review*.

Verlinde, J., Zak, B. D., Shupe, M. D., Ivey, M. D., and Stamnes, K.: The ARM North Slope of Alaska (NSA) Sites, Meteor Mon, 57, 10.1175/Amsmonographs-D-15-0023.1, 2016.

Weaver, D., Strong, K., Schneider, M., Rowe, P. M., Sioris, C., Walker, K. A., Mariani, Z., Uttal, T., McElroy, C. T., Vömel, H., Spassiani, A., and Drummond, J. R.: Intercomparison of atmospheric water vapour measurements at a Canadian High Arctic site, Atmos. Meas. Tech., 10, 2851–2880, https://doi.org/10.5194/amt-10-2851-2017, 2017.

Widener, K., Bharadwaj, N., and Johnson, K.: Ka-Band ARM Zenith Radar (KAZR) Instrument Handbook, United States Department of Energy (USDOE), https://doi.org/10.2172/1035855, 2012.

WMO: Guide to Meteorological Instruments and Methods of Observation. WMO-No.8, Geneva, Switzerland, ISBN: 978-92-63-10008-5, https://library.wmo.int/idurl/4/68662, 2021.

Wohner, C., Peterseil, J., and Klug, H.: Designing and implementing a data model for describing environmental monitoring and research sites, Ecol Inform, 70, ARTN 101708 10.1016/j.ecoinf.2022.101708, 2022.

Younkin, K. and Long, C.: Improved Correction of IR Loss in Diffuse Shortwave Measurements: An ARM Value-Added Product, PNNL; Richland, WA, United States, Medium: ED, 10.2172/1020732, 2003.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

Open Access Earth System Science Data Discussions

	Facility Name	Coordinates	Measured Variables
Whitehorse	Whitehorse	N60.71, W135.07	All
Iqaluit	Iqaluit	N63.74, W68.51	All
Sodankylä	Operative Sounding Station Area; Automatric Weather Station (LUOxxxx)	N67.366618 – N67.367220, E26.628253 - E26.63144	Pressure, Visibility
	CO2 Flux Mast Area (VUOxxxx)	N67.361883, E26.643003 - E26.64323	Total precipitation of water, all wind, vertical velocity, temperature, dew-point temperature, relative humidity, snow thickness, all radiation, cloud base height
	Intensive Observation Area (IOAxxxx)	N67.361654 - N67.361950, E26.633190 - E26.634191	Temperature, relative humidity, snow thickness, snowfall flux, snow water equivalent, all short-wave radiation, soil temperature profile, soil moisture, snow temperature
	Lichen Fence (JAKxxxx)	N67.36710 - N67.36716, E26.634740 - E26.63513	All radiation
	Micrometeorological Mast Area (METxxxx)	N67.361711 - N67.36216, E26.63726 - E26.65117	All wind, temperature, vertical velocity, relative humidity, snow thickness, all radiation, all heat fluxes, friction velocity, soil temperature profile, soil moisture, snow temperature
	Peatland Area (SUOxxxx)	N67.361903 - N67.36707, E26.633802 - E26.654067	Temperature, dew-point temperature, relative humidity, snow thickness, all short-wave radiation, soil temperature profile, soil moisture, snow temperature
Utqiaġvik	ARM Facility	N71.19228, W156.3654	All except ozone concentration, snow thickness, and soil temperature profile
	GML Barrow Atmospheric Baseline Observatory	N71.3230, W156.6114	Ozone concentration, snow thickness, and soil temperature profile
Tiksi	Baseline Surface Radiation Network (BSRN)	N71.5862, E128.9188	All radiation observations
	Fluxtower	N71.595, E128.882	All except radiation observations
Ny-Ålesund	Baseline Surface Radiation Network (BSRN)	N78.92278, E11.92725	All radiation observations, pressure, cloud base height
	AWIPEV Met.Tower	N78.92226, E11.92667	All wind, temperature, relative humidity, specific humidity
	Balloon Launch Facility	N78.92301, E11.92271	All timeSeriesProfileSonde observations
Eureka	Baseline Surface Radiation Network (BSRN)	N79.989, W85.9404	All radiation observations

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

	Fluxtower	N80.083, W86.417	Pressure, all wind, temperature, relative humidity, snow thickness, ground heat flux, soil temperature profile
	Sonde Launch	N79.9833, W85.9333	All timeSeriesProfileSonde observations

Table 1. List of facility coordinates for locations where $MODF_{ysm}$ measurements were collected at each of the supersite locations. The variables (listed in Table 3) that are measured at each location are listed. In some cases, the same variable is measured at multiple locations for a single site; these observations and their corresponding coordinates are embedded within the MODF.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

847

	DOI	Title	Citation
Whitehorse	https://doi.org/10.21343/a33e-j150	MODF for Erik Nielsen Airport, Whitehorse, Canada during YOPP SOP1 and SOP2	Huang et al., 2023a
Iqaluit	https://doi.org/10.21343/yrnf-ck57	MODF for Iqaluit Airport, Iqaluit, Nunavut, Canada during YOPP SOP1 and SOP2	Huang et al., 2023b
Sodankylä	https://doi.org/10.21343/m16p-pq17	Merged observation data file for Sodankylä	O'Connor, 2023
Utqiaġvik	https://doi.org/10.21343/a2dx-nq55	MODF for Utqiaġvik, Alaska, during YOPP SOP1 and SOP2	Akish & Morris, 2023c
Tiksi	https://doi.org/10.21343/5bwn-w881	MODF for Tiksi, Russia, during YOPP SOP1 and SOP2	Akish & Morris, 2023b
Ny-Ålesund	https://doi.org/10.21343/y89m-6393	Merged Observatory Data File (MODF) for Ny Ålesund	Holt, 2023
Eureka	https://doi.org/10.21343/r85j-tc61	MODF for Eureka, Canada, during YOPP SOP1 and SOP2	Akish & Morris, 2023a

848
849 **Table 2.** List of final DOIs for each of the supersite's $MODF_{ysm}$.

850

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

851

MODF featureType	Measured Variables	Whitehorse *lat: 60.71 N lon: 135.07 W*	Iqaluit *lat: 63.74 N lon: 68.51 W*	Sodankylä *lat: 67.367 N lon: 26.629 E*	Utqiaġvik *lat: 71.325 N lon: 156.625 W*	Tiksi *lat: 71.596 N lon: 128.889 E*	Ny-Ålesund *lat: 78.923 N lon: 11.926 E*	Eureka *lat: 80.083 N lon: 86.417 W*
timeSeries Variables	Pressure (Pa)	surface	surface	surface, mean sea-level	surface	surface	surface	surface
	Total precipitation of water in all phases per unit area (kg m-2 s-1)	surface	surface	surface			surface	
	Eastward Wind (m s-1)	surface	near-surface	near-surface	near-surface (2m)	near-surface (4m)	near-surface (10m)	near-surface (6m)
	Northward Wind (m s-1)	surface	near-surface	near-surface	near-surface (2m)	near-surface (4m)	near-surface (10m)	near-surface (6m)
	*Wind gust (m s-1)			near-surface (10m)				
	Vertical velocity (m s-1)			near surface (2 m)				
	Temperature (K)	near-surface (2m)	near-surface (2m)	skin, near-surface (2m)	skin, near-surface (2m)	skin, near-surface (2m)	near-surface (2m)	skin, near-surface (2m)
	Dew-point Temperature (K)	near-surface (2m)	near-surface (2m)	near-surface (2m)	near-surface (2m)			
	Relative Humidity (1 or %)	near-surface (2m)	near-surface (2m)	near-surface (2m)	near-surface (2m)	near-surface (2m)	near-surface (2m)	near-surface (2m)
	Specific Humidity (1 or kg kg-1)						near-surface (2m)	
	Ozone Concentration in Air (mole fraction)				surface			
	Snow thickness (m)		surface	surface	surface	surface		surface
	Snowfall Flux (kg m-1 s-2)				surface			
	Snow water equivalent (kg m-2)				surface			
	Upward Short-wave Radiation (W m-2)		surface	surface	surface	surface	surface	surface
	Downward Short-wave Radiation (W m-2)		surface	surface	surface	surface	surface	surface
	Upward Long-wave Radiation (W m-2)		surface	surface	surface	surface	surface	
	Downward Long-wave Radiation (W m-2)		surface	surface	surface	surface	surface	surface
	Net Short-wave Radiation at the Surface (W m-2)			surface				
	*Horizontal East-facing Long-wave Radiation (W m-2)		surface					
	*Horizontal West-facing Long-wave Radiation (W m-2)		surface					
	*Horizontal South-facing Long-wave Radiation (W m-2)		surface					
	*Horizontal North-facing Long-wave Radiation (W m-2)		surface					
	**Turbulent Latent Heat Flux (W m-2)			surface (EC)	surface (EC, bulk)			
	**Turbulent Sensible Heat Flux (W m-2)			surface (EC)	surface (EC, bulk)			
	**Turbulent time-average eastward stress (Pa)			surface (EC)	surface			
	**Turbulent time-average northward stress (Pa)				surface			
	*Friction Velocity (m s-1)			surface (EC)				
	Cloud Base Height (m)	ground-based remote sensing	ground-based remote sensing	ground-based remote sensing			ground-based remote sensing	

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

Open Access Earth System Science Data Discussions

	Ground Heat Flux (W m-2)			near-surface	near-surface	near-surface		near-surface
	Visibility (m)			near-surface				
timeSeriesProfile Variables	Atmospheric pressure (Pa)		near-surface (2m, 10m)					
	Total precipitation of water in all phases per unit area (kg m-2 s-1)		near-surface (2m, 10m)					
	Eastward Wind (m s-1)		near-surface (2m, 10m)	near-surface (18m, 32m, 38m, 48m)	near-surface (2m, 10m, 20m, 40m)		near-surface (2m, 10m)	near-surface (6m, 11m)
	Northward Wind (m s-1)		near-surface (2m, 10m)	near-surface (18m, 32m, 38m, 48m)	near-surface (2m, 10m, 20m, 40m)		near-surface (2m, 10m)	near-surface (6m, 11m)
	Temperature (K)		near-surface (2m, 10m)	near-surface (3m, 8m, 18m, 32m, 48m)	near-surface (2m, 10m, 20m, 40m)	near-surface (2m, 6m, 10m)	near-surface (2m, 10m)	near-surface (2m, 6m, 10m)
	Dew-point Temperature (K)				near-surface (2m, 10m, 20m, 40m)			
	Relative Humidity (1 or %)		near-surface (2m, 10m)	near-surface (3m, 8m, 18m, 32m, 48m)	near-surface (2m, 10m, 20m, 40m)	near-surface (2m, 6m, 10m)		near-surface (2m, 6m, 10m)
	Soil Temperature Profile (K)			sub-surface (5cm, 30cm)	sub-surface (5cm, 10cm, 15cm, 20cm, 25cm, 30cm, 45cm, 70cm, 95cm, 120cm)	sub-surface (5cm, 10cm, 15cm, 20cm, 25cm, 30cm, 45cm, 70cm, 95cm, 120cm)		sub-surface (5cm, 10cm, 15cm, 20cm, 25cm, 30cm, 45cm, 70cm, 95cm, 120cm)
	Soil Moisture (kg m-2)			sub-surface (5cm, 30cm)				
	Snow Temperature (K)			near-surface (10cm, 20cm, 30cm, 40cm, 50cm, 60cm, 70cm, 80cm, 90cm, 100cm, 110cm)				
timeSeriesProfileSonde Variables	Atmospheric pressure (Pa)	radiosonde	radiosonde		radiosonde		radiosonde	
	Eastward Wind (m s-1)	radiosonde	radiosonde		radiosonde		radiosonde	
	Northward Wind (m s-1)	radiosonde	radiosonde		radiosonde		radiosonde	
	Temperature (K)	radiosonde	radiosonde		radiosonde		radiosonde	
	Dew-point Temperature (K)	radiosonde	radiosonde		radiosonde		radiosonde	
	Specific Humidity (1 or kg kg-1)						Radiosonde	
	Relative Humidity (1 or %)	radiosonde	radiosonde		radiosonde		radiosonde	
**** Denotes a variable NOT included in the H-K Table*** ***** Denotes a calculated variable (not a direct observation)***								

Table 3. List of the geophysical variables currently included in each supersite's MODF. Note that this table only includes variables currently in the existing $MODF_{ysm}$, and does not indicate the complete list of variables that are observed at each site. An asterisk (*) denotes a variable not included in the H-K table and a double asterisk (**) denotes a calculated variable.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

Open Access Earth System Science Data Discussions

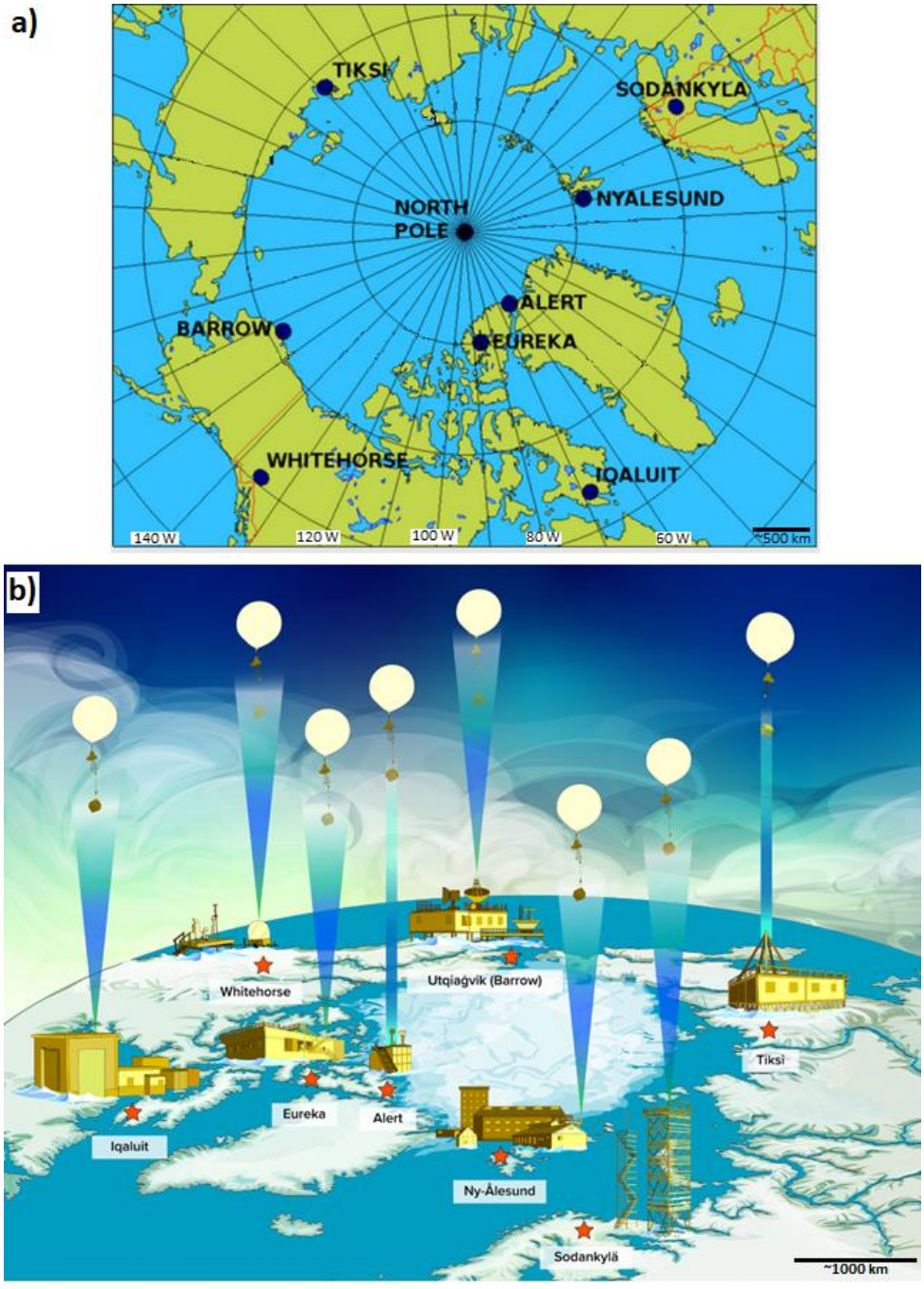

Figure 1. a) Locations of the MODF$_{ysm}$ YOPP supersites (Antarctic supersites not shown). (b) Infographic depicting iconic building(s) at each supersite. The infographic is roughly centred around the North Pole (centre). All locations shown have generated a MODF$_{ysm}$, with the exception of Alert (in progress).

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

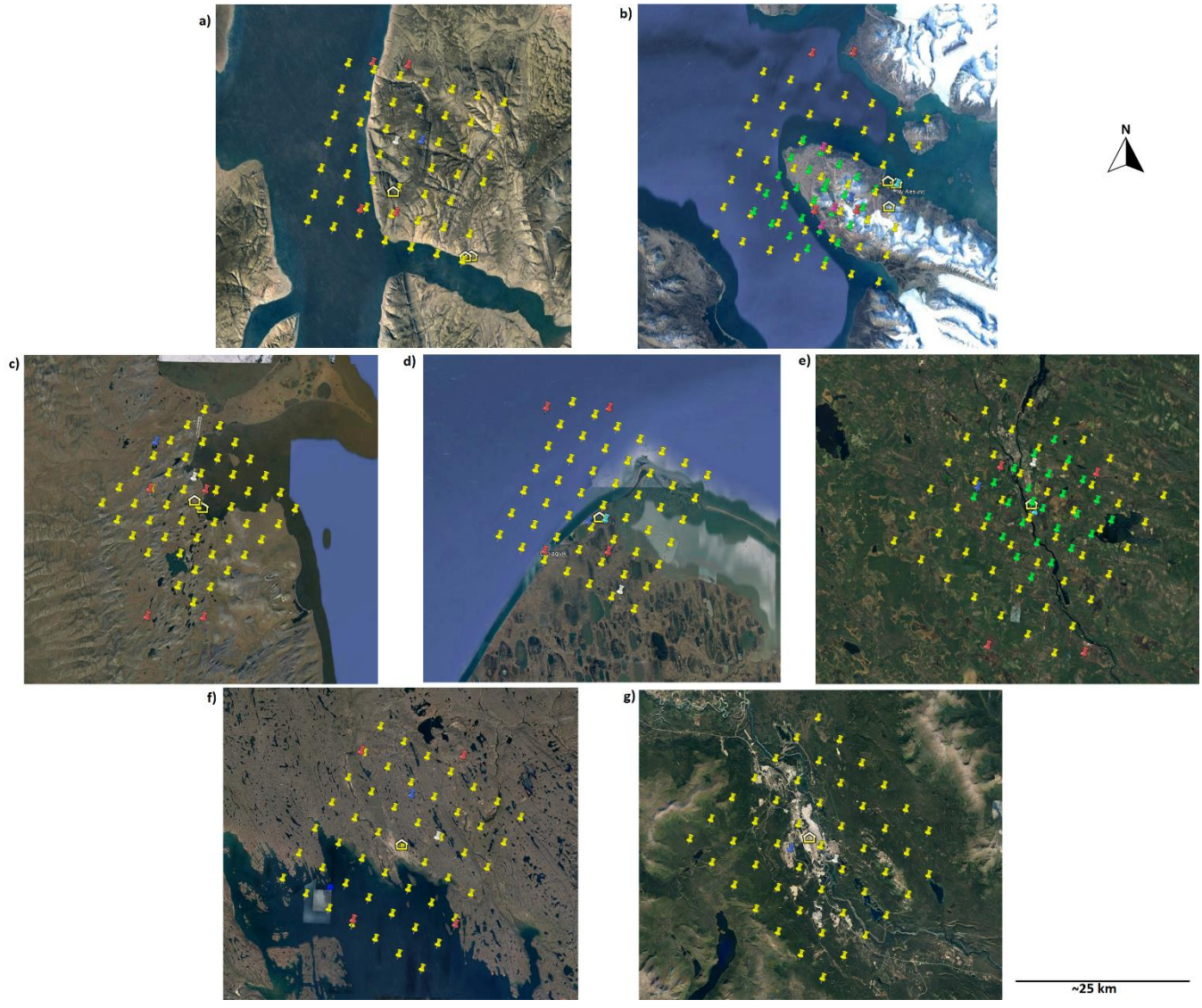

Figure 2. Model grid points at and around each supersite (a) Eureka, (b) Ny-Ålesund, (c) Tiksi, (d) Utqiaġvik, (e) Sodankylä, (f) Iqaluit, and (g) Whitehorse, displayed through the Google Earth web-platform: *Image Landsat / Copernicus, Image ©2023 Maxar Technologies*. Sites are organized from highest latitude (Eureka) to lowest (Whitehorse). Yellow building icons represent the location of the facility on-site which contains all co-located instruments. Similarly, icons for the AROME-Arctic model grid are indicated by a green pin, ARPEGE pins are in white, DWD-ICON pins are light blue, ECCC-CAPS pins are yellow, ECMWF-IFS pins are dark blue, and SL-AV pins are in red. All images are north-aligned, nadir view.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

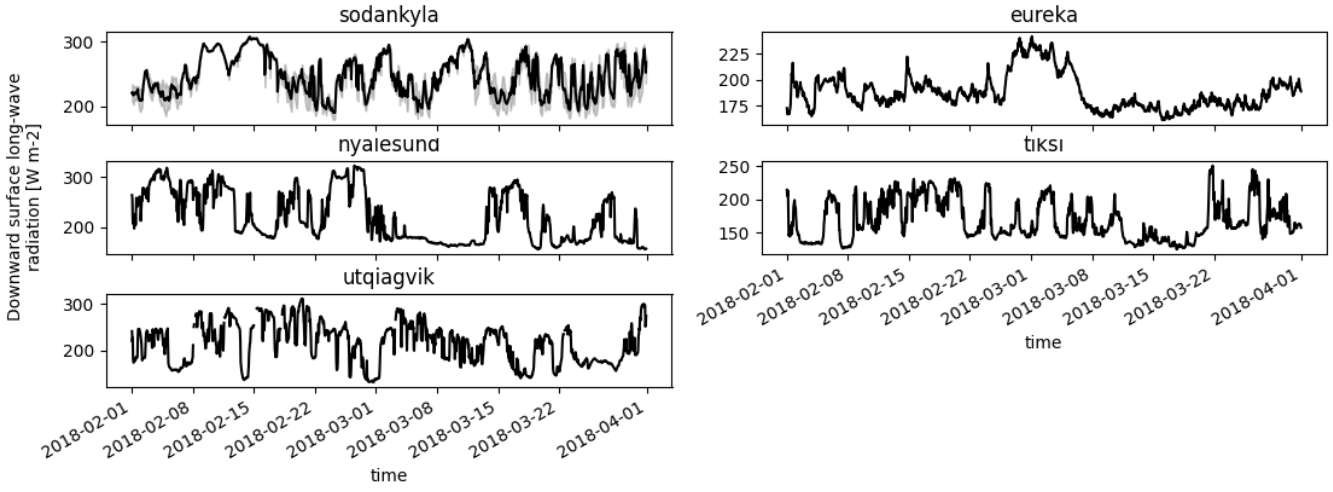

Figure 3. Observations of downward surface long-wave radiation ("rlds") conducted during SOP1 at each supersite. Observations from Whitehorse and Iqaluit were not available during SOP1. Sodankylä conducts multiple observations of rlds; the mean (black line) and spread in observed rlds (grey) are shown.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

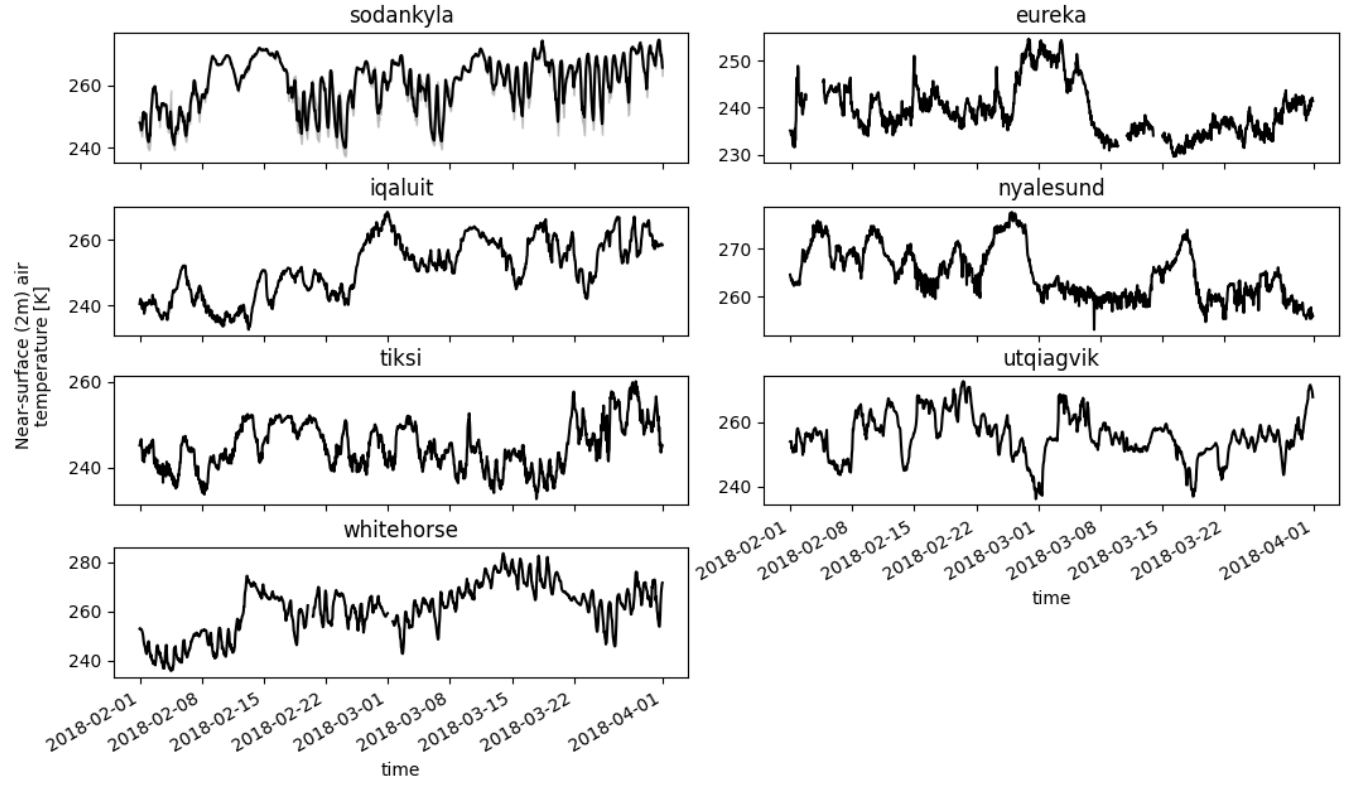

Figure 4. Similar to Figure 3, except for observations of near-surface (2 m) air temperature ("tas") conducted at each supersite during SOP1.

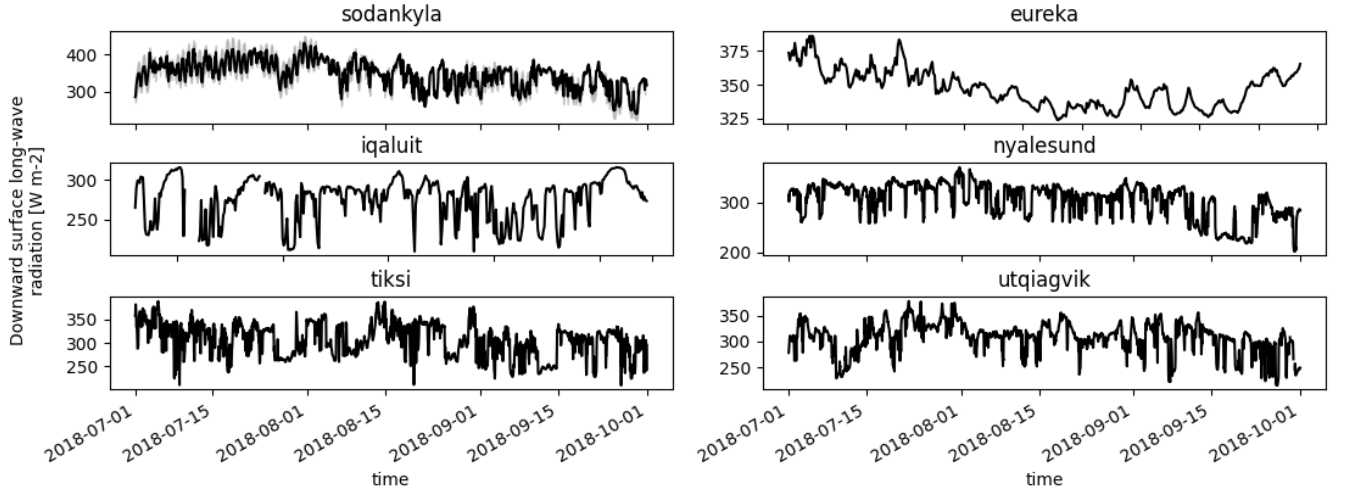

Figure 5. Similar to Figure 3, except for observations of downward surface long-wave radiation ("rlds") conducted during SOP2 at each supersite. Observations from Whitehorse were not available during SOP2.

https://doi.org/10.5194/essd-2023-497
Preprint. Discussion started: 23 January 2024

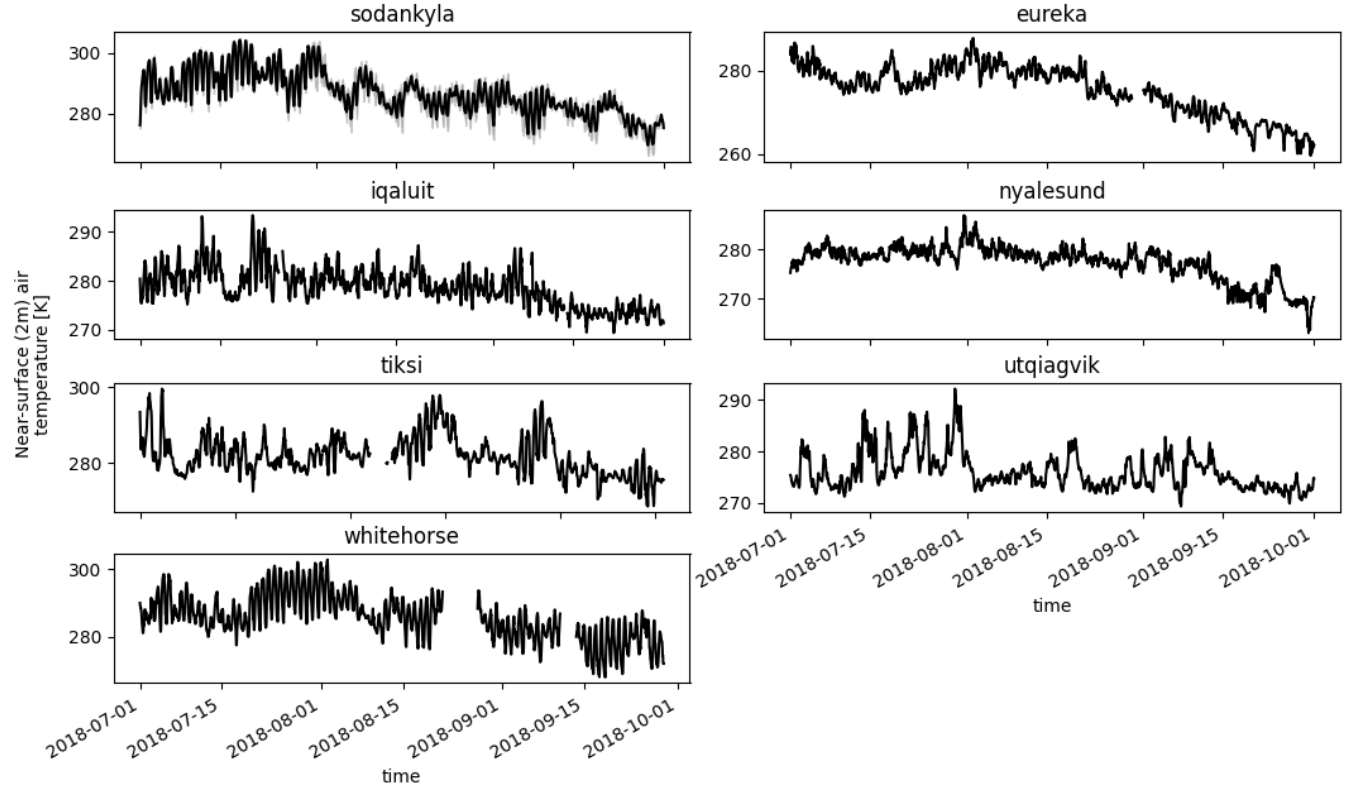

Figure 6. Similar to Figure 3, except for observations of near-surface (2 m) air temperature ("tas") conducted at each supersite during SOP2.