The YOPP site Model Intercomparison Project (YOPPsiteMIP) phase 1: project overview and Arctic winter forecast evaluation

Jonathan J. Day[1], Gunilla Svensson[2], Barbara Casati[3], Taneil Uttal[4], Siri-Jodha Khalsa[5], Eric Bazile[6], Elena Akish[4], Niramson Azouz[6], Lara Ferrighi[7], Helmut Frank[8], Michael Gallagher[4,9], Øystein Godøy[7], Leslie M. Hartten[4,9], Laura X. Huang[3], Jareth Holt[2], Massimo Di Stefano[7], Irene Suomi[10], Zen Mariani[3], Sara Morris[4], Ewan O'Connor[10], Roberta Pirazzini[10], Teresa Remes[7], Rostislav Fadeev[11], Amy Solomon[4,9], Johanna Tjernström[12], Mikhail Tolstykh[11,13],

[1]European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
[2]Department of Meteorology and Bolin Centre for Climate Change, Stockholm University, Sweden.
[3]Meteorological Research Division, Environment and Climate Change Canada, Canada.
[4]NOAA Physical Science Laboratory, Boulder, Colorado, USA.
[5]National Snow and Ice Data Center, University of Colorado,
[6]Meteo France, Toulouse, France
[7]Norwegian Meteorological Institute, Oslo, Norway.
[8]Deutscher Wetterdienst, Offenbach, Germany
[9]Cooperative Institute for Research in Environmental Science (CIRES), University of Colorado, Boulder, Colorado, USA
[10]Finnish Meteorological Institute, Helsinki, Finland.
[11]Marchuk Institute of Numerical Mathematics Russian Academy of Sciences, Russia
[12]Swedish Meteorological and Hydrological Institute, Linköping, Sweden
[13]Hydrometeorological Research Centre of Russia, Russia

Correspondence to: Jonathan J. Day (jonathan.day@ecmwf.int)

Abstract.
Although the quality of weather forecasts in the polar regions is improving, forecast skill there still lags the lower latitudes. So far there have been relatively few efforts to evaluate processes in Numerical Weather Prediction systems using in-situ and remote sensing datasets from meteorological observatories in the terrestrial Arctic and Antarctic, compared to the mid-latitudes. Progress has been limited both by the heterogeneous nature of observatory and forecast data but also by limited availability of the parameters needed to perform process-oriented evaluation in multi-model forecast archives. The YOPP site Model Inter-comparison Project (YOPPsiteMIP) is addressing this gap by producing Merged Observatory Data Files (MODFs) and Merged Model Data Files (MMDFs), bringing together observations and forecast data at polar meteorological observatories in a format designed to facilitate process-oriented evaluation.

An evaluation of forecast performance was performed at seven Arctic sites, focussing on the first YOPP Special Observing Period in the Northern Hemisphere (SOP1), February and March 2018. It demonstrated that although the characteristics of forecast skill vary between the different sites and systems, an underestimation in boundary layer temperature variability across models, which goes hand in hand with an inability to capture cold extremes, is a common issue at several sites. Diagnostic analysis using surface fluxes suggests that this is at least partly related to insufficient thermal representation of the land-surface in the models, which all use a single layer snow model.

Deleted: nce

1 Introduction

Recent decades have seen a marked increase in human activity in the polar regions leading to an increasing societal demand for weather and environmental forecasts (Emmerson and Lahn, 2012; Goessling et al., 2016). Despite this growing need, the skill of weather forecasts in the polar regions lags that of the mid-latitudes (Jung et al., 2016; Bauer et al., 2016). This is partly the result of the relatively lower density of conventional observations in high compared to mid-latitudes (Lawrence et al., 2019), but is also related to the occurrence of meteorological situations and phenomena which are historically difficult to model such as stable boundary layers (e.g. Atlaskin and Vihma, 2012; Sandu et al., 2013; Holtslag et al., 2013), mixed-phase clouds (e.g. Pithan et al., 2014, 2016, Solomon et al., 2023), and the importance of coupling between the atmosphere and snow and ice surfaces (e.g. Day et al., 2020; Batrak and Muller, 2019; Svensson and Karlsson, 2011).

The ability of climate models to represent atmospheric processes in polar regions has recently been assessed highlighting deficiencies in near-surface and boundary layer properties (Pithan et al., 2014; Svensson and Karlsson, 2011; Karlsson and Svensson, 2013). Since many climate models are based on global weather forecasting systems, understanding the causes of forecast error after 1-2 days may help develop understanding of the sources of error in climate models (Rodwell and Palmer, 2007). Nevertheless, until recently there has been little focus on evaluating Numerical Weather Prediction (NWP) models using in-situ data from the terrestrial Arctic and Antarctic (Jung and Matsueda, 2016; Jung et al., 2016).

Deleted: 4

Recent studies, conducted as part of the World Weather Research Programme's Polar Prediction Project (PPP, Jung et al, 2016) have started to address this gap, assessing the skill of both the large scale circulation (Bauer et al., 2016) and surface weather properties (Køltzow et al., 2019). The Year of Polar Prediction (YOPP) site Model Intercomparison Project (YOPPsiteMIP) was designed to build on these earlier studies by utilising process level data from polar observatories to diagnose the causes of forecast error from a process perspective and ultimately inform model development. Although process-oriented evaluation studies focussing on polar processes are not new, those that have been done have tended to focus on one or two sites or a specific field campaign (see Day et al., 2020; Batrak and Müller, 2019; Miller et al., 2018; Tjernström et al., 2021, Kähnert et al., 2023 for some recent examples). A key aim of YOPPsiteMIP is to provide a pan-Polar perspective on forecast evaluation and process representation.

YOPPsiteMIP participants were asked to provide data in so-called Merged Data Files (MDFs) which includes both Merged Observatory Data Files (MODFs), for observatory data, and Merged Model Data Files (MMDFs), for model data. These data standards, which were developed specifically for YOPPsiteMIP, are described by Uttal et al. (2023). Using this common file format, with consistent naming and metadata, facilitates equitable and efficient comparisons between models and observations. This standardisation of the data from different observatories also aids interoperability in the sense that the same evaluation code can be applied at different sites. These MDF filetypes were developed as part of PPP, following the FAIR (Findable, Accessible, Interoperable, Reusable) data principles (Wilkinson, 2016). Details of the MDF concept and specifics of the data processing chain for producing MDFs are described in Uttal et al. (2023).

Deleted: and related Python toolkit

Deleted: and Gallagher et al., (in prep).

The observatories selected for YOPPsiteMIP represent a geographically diverse set of locations (see Mariani et al. 2024). At these sites a wide range of instruments measuring properties of the air, snow and soil are employed, extending far beyond the traditional synoptic surface and upper-air observation network, which are collected for use in the production and evaluation of NWP systems (Uttal et al., 2015). Taken together, the observations collected at these observatories offer opportunities to develop a deeper understanding of the physical processes governing the weather in the polar regions, their representation in forecast models, and how this varies from site to site. The processes and phenomena targeted in YOPPsiteMIP include boundary-layer turbulence, surface exchange (including over snow and ice) and mixed-phase clouds.

A benefit of organizing coordinated evaluation involving several NWP systems and multiple sites is that it helps clarify if the issues revealed by the analysis are model or location specific. The modelling community has organized model inter-comparisons to target various atmospheric processes relevant for Arctic conditions (e.g. Cuxart et al., 2006; Pithan et al., 2016; Tjernström et al 2005, Sedlar et al. 2020, Solomon et al., 2023) each using its own protocol for data sharing. However, the newly developed standardisation of the observational and forecast model data developed for YOPPsiteMIP is planned to be used for future MIIPs (model intercomparison and improvement projects). Converging on a standard like this will aid interoperability, making it easier for model developers to expand their evaluation to new sites or observational campaigns, but also to other models or forecasting systems.

MDFs were requested for the locations listed in Table 1 and shown in Figure 1 during the YOPP Special Observing Periods, during which the observations taken at many polar observatories (e.g. the frequency of radiosondes) was enhanced (see Lawrence et al., 2019; Bromwich et al., 2020). For the Northern Hemisphere the periods Feb–Mar 2018 and Jul–Sep 2018 were selected and named NH-SOP1 and SOP2 respectively. For the Southern Hemisphere or SH-SOP the period Nov–Feb 2018/19 was chosen. At the time of publication MMDFs have been produced and archived from seven NWP systems for these periods and all of the sites listed have MMDFs from at least one model. MODFs have been produced and archived for seven of the sites so far and it is hoped that additional MODFs will be produced in the future to fill the gaps, particularly in the Southern Hemisphere.

Deleted: M

Observatory name ***Filename***	**Latitude Longitude**	**Elevation**
Arctic land sites		
Utqiaġvik (Formerly known as Barrow, Alaska) *Utqiaġvik*	71.32°N, 156.62°W	8-20 m
Oliktok Point (Alaska) *oliktok*	70.50°N 149.89°W	2-6 m
Whitehorse (Canada) *whitehorse*	60.71°N, 135.07°W	682 m
Eureka (Canada) *eureka*	80.08°N 86.42°W	0-610 m
Iqaluit (Canada) *iqaluit*	63.74°N, 68.51°W	5-11 m
Alert (Canada) *alert*	82.49°N, 62.51°W	8-210 m
Summit (Greenland) *summit*	72.58°N, 38.48°W	3210-3250 m
Ny-Ålesund (Svalbard) (Zeppelin station) *nyalesund*	78.92°N, 11.53°E (78.9°N, 11.88°E)	0-30 m (473 m)
Sodankylä (Finland) *Sodankylä*	67.37°N, 26.63°E	198 m
Pallas (Finland) *pallas*	67.97°N, 24.12°E	305 m

Tiksi (Russia) *tiksi*	71.60°N, 128.89°E	1-30 m
Cherskii (Russia) *cherskii*	68.73°N, 161.38°E (68.51°N, 161.53°E)	8 m (16 m)
Ice Base Cape Baranova (Russia) *baranova*	79.3°N, 101.7°E	24 m

Arctic Ocean sites

SHEBA location *sheba*	165°W, 76°N	Sea level
Arctic Ocean 1 (Gakkel Ridge) *ao1*	10°E, 85°N	Sea level
Arctic Ocean 2 (North Pole) *ao2*	0°E, 90°N	Sea level
Arctic Ocean 3 (Canada Basin) *ao3*	135°W, 81°N	Sea level

Antarctic land sites

Alexander Tall Tower *alexander*	79.01°S, 170.72°E	55 m
Casey *casey*	66.28°S, 110.53°E	30 m
Davis *davis*	68.58°S, 77.97°E	
Dome C *domec*	75.08°S, 123.34°E	3233 m
Dumont d'Urville *dumont*	66.66°S, 140.01°E	0-50 m
Halley IV *halley*	75.58°S, 26.66° W	130 m
King Sejong (King George Island) *kingsejong*	62.22°S, 58.79° W	10 m

Georg von Neumayer *neumayer*	70.65°S, 8.25°W	42 m
Mawson *mawson*	67.60°S, 62.87°E	15 m
Syowa (Showa) *syowa*	69.00°S, 39.59°E	18-29 m
Jang Bogo (Terra Nova Bay) *jangbogo*	74.62°S, 164.23°E	36 m
Amundsen-Scott South Pole *southpole*	90°S, 0°E	2835 m
Byrd *byrd*	80.01°S, 119.44°W	1539 m
Rothera *rothera*	67.57°S, 68.13° W	4 m
Vostok *vostok*	78.46°S, 106.84°E	3489 m
McMurdo (Scott base) *mcmurdo*	77.85°S, 166.67°E (77.85°S, 166.76°E)	10 m (10 m)
Troll *troll*	72.01°S, 2.54°E	1275 m

Table 1: List of YOPPsiteMIP observatory locations: name, *name as used in filenames,* latitude, longitude and elevation. Where an elevation range is stated, this is because the instruments at a given observatory extend over a range of values due to variations in local topography.

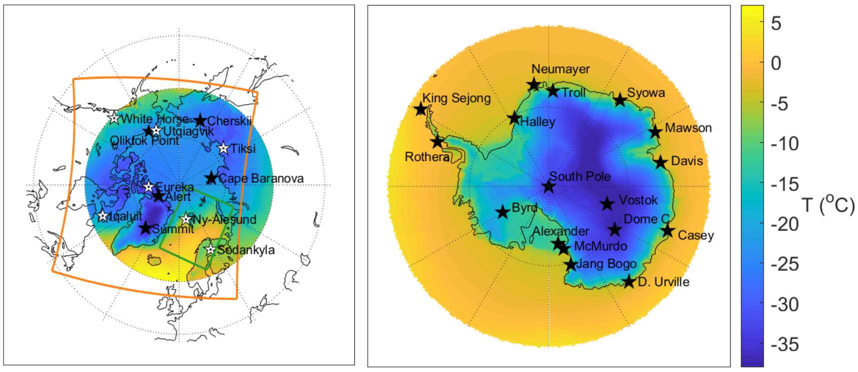


Figure 1: Maps of the ERA5 2m-temperature climatology (1990-2019) for February-March (time of NH-SOP1) for Arctic (left) and for November-February (SH-SOP) for Antarctic (right). The observatories used in YOPPsiteMIP are marked with stars. White stars indicate the sites where MODFs are currently available, which are the subject of this study; black stars indicate the sites whose MODFs are not yet complete. The orange and green boxes depict the extent of the ECCC-CAPS and AROME-Arctic domains respectively.

The purpose of this paper is two-fold: firstly, to document the first version of the YOPPsiteMIP dataset along with a basic description of the forecasting systems and their respective MMDFs that are archived at the YOPP Data Portal, hosted by the Norwegian Meteorological Institute (MET Norway). Secondly, the paper presents a multi-site evaluation of seven forecasting systems during NH-SOP1, at seven Arctic observatories that have produced MODFs. The locations are indicated by the white stars in Figure 1a and the MODFs and full details of the sites are described in Mariani et al., (2024).

The seven Arctic sites used for evaluation in this study cover both high and sub-Arctic climate zones. Tiksi, Utqiaġvik, Iqaluit, Ny-Ålesund and Eureka all sit in the Arctic tundra characterised by low vegetation. The remaining two sites Whitehorse and Sodankylä are sub-Arctic, with higher vegetation corresponding to the boreal cordillera and taiga ecozones respectively. Whitehorse, Iqaluit, Ny-Ålesund and Eureka are characterised by complex topography in the surrounding area, whereas the other sites are flatter. All the sites are in close vicinity to either frozen ocean (sea ice) or frozen inland water bodies at this time of year and the land surrounding each observatory is covered in snow throughout the period Feb-Mar 2018. A visual representation of the model grids with respect to the landscape surrounding these stations can be seen in Fig 2 of Mariani et al., (2024) in which a more detailed description of the site characteristics may be found.

2 Description of simulations, model formulation and output protocol

To date, six NWP centres have submitted forecasts from seven forecasting systems for SOP1 & SOP2, with two systems submitted for the SH-SOP (see Table 2). Four of the systems are global:

- The Integrated Forecasting System from the European Centre for Medium-Range Weather Forecasts (ECMWF-IFS; Day et al., 2023),
- The Action de Recherche Petite Echelle Grande Echelle from Meteo France (ARPEGE-MF ; Bazile and Azouz, 2023a),
- The Semi-Lagrangian, based on the absolute vorticity equation from the Hydrometeorological Research Centre of Russia (SLAV-RHMC, Tolstykh, 2023) and,
- The Icosahedral Nonhydrostatic Model from Deutscher Wetterdienst (DWD-ICON; Frank, 2023).

Three are regional:

- The Canadian Arctic Prediction System from Environment and Climate Change Canada (ECCC-CAPS; Casati, 2023)

Deleted: Morris

Deleted: in prep

Deleted: low

- and two versions of Applications of Research to Operations at Mesoscale (AROME) from Meteo France (AROME-MF; Bazile and Azouz, 2023b) and from MET Norway (AROME-Arctic; Remes, 2023).

The domain boundaries of the regional forecasting systems can be seen in Figure 1 (note that only two of the observatories are within the AROME domain). The forecasts analysed here were initialised at 00 UTC for each day of the SOPs (although 12UTC forecasts are also available on the archive for many of the systems). The forecast leadtime varies between the different systems but all forecasts are at least two days long (see Table 2 and Figs 2 & 3).

The files for some of the systems (CAPS, SLAV, ARPEGE, AROME-MF) are provided with multiple grid-points, centred on the observatory location. For others only a single grid-point was provided. Multiple grid-points centred around the observatory location were requested because many of the observatories are located in the vicinity of coasts, which leads to representativeness issues when comparing the land-based observation to model output for grid-points being partially or entirely over the ocean. In this study when there are multiple grid points we choose the closest 100% land point to the supersite location, with the exception of CAPS, for which the central grid-point within a beam of 7x7 grid-points was considered (since nearest to the observation site) and ICON which provided the single closest gridpoint to the station locaton. As a result, the evaluation utilises a 100% land gridbox at all models and locations, with the exception of ICON, which has 23% land cover at the Utqiaġvik and 73% at Ny-Ålesund, and CAPS, which has 37% land cover in Utqiagvik, 71% and 77% in Tiksi and Iqaluit, and over 90% land cover for the other sites. Comparison of the CAPS grid-points surrounding Utqiagvik with each other indicated that the evaluation would not be much influenced by the choice of gridcell (not shown) since during the Arctic winter the frozen ocean gridpoints have similar propreties to the snow-covered land surface (e.g. when analysing the surface energy budget sensitivity to radiative forcing in Section 3.4). The grid resolutions range from 2.5 km to ~30 km and the model timestep varies from 1.5 to 7.5 min (see Table 2).

The models have quite a diverse mixture of formulations for atmospheric dynamics, land surface, sub-grid scale parameterisations and initialisation/data assimilation procedures. More details about the simulations with specific models are provided below and a summary of the key model components/parameterisations used in each model is included in Table 3.

2.1 IFS-ECMWF

MMDFs for the operational forecasts with the IFS high resolution deterministic forecasts are available for the period starting Jan 2018. The initial forecasts are produced with IFS cycle 43r3 which was an atmosphere only model with persisted sea ice and anomaly SSTs. From 5 June 2018 (i.e. before SOP2) the forecasts were produced with cycle 45r1 which included dynamic sea ice and ocean fields (see Day et al., 2022 for more information). Although the model version changes the horizontal (~9km) and vertical resolution (L137) are the same in all SOPs. The data archived in the MMDFs is provided at the model timestep (7.5 min) for a single model grid point closest to the observatory. In addition to the grid point data a number of parameters (including albedo, surface temperature and surface energy fluxes) are provided on the land-surface model tiles to enable detailed evaluation of processes even at heterogeneous sites. A complete description for the two versions of the IFS can be found here: https://www.ecmwf.int/en/publications/ifs-documentation.

2.2 ARPEGE-MF

The version of ARPEGE submitted to YOPPsiteMIP was a pre-operational version based on the cy43t2_op1 operational system but coupled with the 1D sea-ice model GELATO (Bazile et al. 2020). The resolution of the model used for these simulations is the same as is used operationally at Meteo France which is variable (using a stretching factor of 2.2) with the pole (highest resolution of 7.5 km) over France for SOP1 and SOP2 and over Antarctica in SOP-SH and 105 vertical levels. The horizontal resolution is about 8-9 km over the North-Pole and timeseries have been provided for the three SOPs in the

Deleted: on
Deleted: and there are issues with
Deleted: , over land or blended
Deleted: .38
Deleted: 0.69
Deleted: .17
Deleted: While we expected land-processes in
Deleted: to be mis-represented by
Deleted: these CAPS
Deleted: and ICON
Deleted: ocean-dominated grid-points
Deleted: ,
Deleted: this revealed
Deleted:
Deleted:
Deleted: not always
Deleted: to
Deleted: be the case,
Deleted: acts as
Deleted: ,
Deleted:
Deleted:

MMDF format for the 21 YOPP observatories with an hourly output for both state variables (instantaneous) and fluxes (accumulated).

2.3 SLAV-HMRC

MMDFs were produced by the SLAV model (Tolstykh et al., 2018) for both SOP1 and SOP2 containing 7-day forecasts starting at 00 UTC. The output is available for 4 horizontal grid points surrounding selected observatories, every 15 minutes (i.e. every fourth timestep). Depending on variable, the output is instantaneous or a 15-min averaged value. Data for 13 of the Arctic observatories in Table 1 are provided. Selection of observatories is based on model resolution in latitude which is relatively low, ~16 km in Northern polar areas; also, the ao2 point is not included because the model grid does not contain the poles.

Deleted: y

2.4 ICON-DWD

MMDFs from DWD's ICON (Zängl et al., 2015) are available from February 2018 to June 2020 containing 7.5-day forecasts starting at 00 and 12 UTC for Sodankylä, Ny-Ålesund, and Utqiaġvik (Barrow). The mesh width is 13 km. Different model versions are used during this period. In February icon-nwp-2.1.02 was used followed by icon-2.3.0-nwp0 during 2018-02-14 to 2018-06-06, and from 2018-09-19 to 2018-12-05 icon-2.3.0-nwp2 was in operation. Since 2018-02-14, a new orographic data set came in operations, however, for the 3 data points provided the changes were less than 1 m in height. The sea ice analysis used in ICON, was based on the Real-Time Global SST High Resolution Analysis of NCEP until 2018-07-16. Since then it is based on the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; Donlon et al., 2012). To represent variations of subgrid scale surface characteristics ICON uses a tile approach. Since 2018-07-16 the tile values of surface fluxes, and other tile dependent variables are included in the MMDFs in addition to the grid average values. Hourly output is available based on a timestep of 120s.

Deleted: 2

2.5 CAPS-ECCC

MMDFs for ECCC-CAPS are available for the whole period from February 2018 to December 2018. Prior to the 28th of June 2018 CAPS was uncoupled and run with the GEM version 4.9.2. After the 29th of June 2018 CAPS was coupled with the Regional Ice and Ocean Prediction system (RIOPS) and run with the GEM version 4.9.4. Atmospheric Lateral Boundary Conditions (LBCs) and initial conditions (ICs) are from ECCC Global Deterministic Prediction System (GDPS). Initial surface fields are from the Canadian Land Data Assimilation System (CaLDAS). The CAPS timeseries are produced for a beam of 7 x 7 grid-points centred on each of the twelve land-based Arctic observatories listed in Table 1. Timeseries up to 48 hours leadtime are made available for the daily runs initialized at 00 UTC. The data is archived with a time frequency of 7.5 min, equivalent to five timesteps of 90 s each.

2.6 AROME-ARCTIC

MET Norway utilises the HARMONIE-AROME (HIRLAM–ALADIN Research on Mesoscale Operational NWP in Euromed–Application of Research to Operations at Mesoscale) model configuration (Bengtsson et al., 2017) for operational weather forecasting for the European Arctic with the name AROME-Arctic (Muller et al., 2017). AROME-Arctic MMDFs are based on the operational forecasts (cy40h.1) and are available for the SOP1 and SOP2 at Sodankylä and Ny-Ålesund. LBCs are derived from the ECMWF IFS-HRES described in Section 2.1. Assimilation of conventional and satellite observation with 3DVAR in the upper atmosphere, optimal interpolation of snow depth, screen level temperature and relative humidity in the surface model. Temperature tolerance in the surface assimilation scheme was increased on 15 March 2018 to better assimilate observed low temperatures. The data archived in the MMDFs are provided hourly for the single model grid-point closest to the site. Model data for the full domain in its original format are also available via thredds.met.no.

2.7 AROME-MF

The AROME -MF system from Météo-France and AROME-ARCTIC from MET Norway are both configurations of the same model system but use different parameterizations of turbulence, shallow convection, cloud microphysics and sea ice. The system used for the YOPPsiteMIP differs from the operational AROME-France configuration (Seity et al., 2011) and the version evaluated for SOP1 in Køltzow et al., (2019) in that it is coupled with the GELATO 1D sea ice model. However, the domain (see Figure 1a), horizontal and vertical grid are exactly the same as the AROME-ARCTIC operational system (see Section 2.6). The ICs and LBCs are interpolated from the global model ARPEGE-MF simulation descried above (Section 2.2). The MMDF files have been produced for Ny-Ålesund, Sodankylä and Pallas with hourly output.

2.8 Output format

For each forecast initial time and each forecasting system a single netCDF file containing all variables was archived following the MMDF format, which use the same nomenclature, metadata, and structure as the MODFs. In order to be able to assess process representation, the YOPPsiteMIP protocol requested that atmospheric fields were provided on native model vertical levels and all fields should be provided with high frequency (every 5 or 15 minutes), ideally at the frequency of the model timestep if practical to support detailed process investigations without the confounding effect of time averaging.

The actual variables archived, frequency and number of grid-points, vary from model to model. For example, ECCC provided a comprehensive set of parameters for the CAPS model focusing on precipitation and clouds microphysics to allow studies on the representation of different types of hydrometeors by the P3 scheme (Morrison and Milbrandt, 2015; Morrison et al., 2015; Milbrandt and Morrison, 2016). A full list of requested variables, along with a schema for producing the MDFs are described in a document known as the H-K Table (Hartten and Khalsa, 2022). The table is available in both human and machine-readable form (PDF and JSON, respectively). The H-K Table relies on standards and conventions commonly used in the earth sciences, including netCDF encoding with CF naming and formatting conventions and is an evolving document that is expected to evolve to fulfil the requirements of future MMDFs and MODFs. The prescribed metadata make data provenance clear and encourage proper attribution of data origin (see further information in Uttal et al., 2023).

Deleted: (Gallagher et al., in prep.)

Although we only focus on model performance during SOP1, a full set of MMDFs and MODFs was produced for both SOPs. The MODFs for Iqaluit (Huang et al., 2023a), Whitehorse (Huang et al., 2023b), Utqiaġvik (formerly known as Barrow: Akish and Morris, 2023c), Eureka (Akish and Morris, 2023a), Tiksi (Akish and Morris, 2023b), Ny-Ålesund (Holt, 2023) and Sodankylä (O'Conner 2023) are described in detail in Mariani et al., (2024) along with descriptions of the site geography. MMDFs have also been produced for the SH-SOP with the ECMWF-IFS and ARPEGE models (See Table 2), but no MODFs for the Antarctic observatories have been produced yet.

Deleted: a
Deleted: b
Deleted: c
Deleted: Morris
Deleted: 3
Deleted: y

Centre	Model-name	Global/Regional and horizontal/vertical resolution	Dynamics timestep/output frequency/forecast length	Version	Key Reference(s)	SOPs in YOPP portal

Deleted: Physics

ECMWF	IFS	Global: 9km/L137	7.5min/7.5min/3d	Cy43r3 for SOP1, Cy45r1 for SOP2 & SOP-SH	Buizza et al., (2017)	SOP1, SOP2 & SOP-SH
Meteo-France	ARPEGE-MF	Global: 7.5-25km/L105	240s/60min/4d	cy43t2_op2	Pailleux et al. (2014)	SOP1, SOP2 & SOP-SH
Meteo-France	AROME-Arctic	Regional: 2.5km/L65	50s/60min/2d	cy43t2_op2	Seity et al., (2011)	SOP1 & SOP2
ECCC	CAPS	Regional: 3km/L62	1.5min/7.5min/2d	vn1.0.0 for SOP1 & vn1.1.0 for SOP2	Milbrandt et al., (2016) Casati, et al., (2023)	SOP1 & SOP2
DWD	ICON	Global: ~13km/L90	2min/60min/7.5d	icon-nwp-2.1.02, icon-2.20-nwp0, icon-2.30-nwp0, icon-2.30.nwp2	Zängl et al., (2015) Prill et al., (2020)	SOP1 & SOP2
HMCR	SLAV	Global: ~20km/L51	3.75min/15min/3d	SLAV20 (2018)	Tolstykh et al., (2018) Tolstykh et al., (2017)	SOP1 & SOP2
MET Norway	AROME-Arctic	Regional: 2.5km/L65	50s/60/2d	HARMONIE-AROME cy40h	Müller et al. (2017) Bengtsson et al., (2017)	SOP1 & SOP2

Table 2. Summary of forecasting systems

Deleted: GELATO

Model-name	**Land-surface model**	**Surface layer/Fluxes**	**Turbulent diffusion**	**Orographic drag**	**Convection**	**Cloud microphysics**	**Radiation**	**Dynamical core**
IFS	HTESSEL: Balsamo et al., (2009)	K-diffusion with stability functions of Dyer (1974) and Högström (1988) and Holtslag and De Bruin (1988) in unstable conditions and for stable conditions	EDMF Köhler et al., (2011) in unstable conditions and K-diffusion (Louis, 1979; Sandu et al., 2013) in stable conditions	Following Lott and Miller (1997) and Baines and Palmer (1990)	mass-flux for deep, shallow and mid-level convection: Tiedtke (1993) and Bechtold et al. (2008)	double moment scheme with four categories of hydrometeor Forbes and Ahlgrimm (2014)	EcRad (Hogan and Bozzo, 2018) Is based on the Rapid Radiation Transfer Model (RRTM, Mlawer et al., 1997; Iacono et al., 2008)	Spectral/FE/H
ARPEGE	SURFEX: Masson et al., (2013)	K-diffusion with modified version of Louis (1979)	TKE: Cuxart et al., (2000) with a modified mixing length (Bazile et 2011)	Scheme described in Catry et al., (2008) following Lott and Miller (1997) for gravity wave drag, and an envelope orography approach (after Wallace et al., 1983)	Mass flux for deep convection following Bougeault (1985) and mass flux for shallow convection following Bechtold et al., (2001)	Single moment with five categories of hydrometeor (Seity et al., 2012)	RRTM	Spectral/FE/H
AROME-MF	SURFEX: Masson et al., (2013)	K-diffusion with stability function of Louis (1979)	TKE: Cuxart et al., (2000)	N/A	Deep convection is explicitly represented and shallow uses the Pergaud et al. (2009) EDMF scheme.	Single moment with six categories of hydrometeor (ICE3; Pinty and Jabouille 1998)	RRTM	Spectral/FD/NH
CAPS	ISBA: Noilhan and Planton (1989) and Bélair et al. (2003)	K-diffusion with stability functions of Delage and Girard (1992) in unstable conditions and Delage (1997) in stable conditions.	TKE with statistical representation of subgrid-scale cloudiness (MoisTKE: Bélair et al. (2005))	Lott and Miller (1997)	Deep convection from the Kain and Fritsch (1990) mass flux scheme and shallow convection from a Kuo-transient scheme (Bélair et al., 2005)	Double moment with Predicted Particle Properties (P3; Morrison and Milbrandt, 2015; Morrison et al, 2015; Milbrandt and Morrison, 2016)	Correlated-k distribution radiative transfer scheme (Li and Barker, 2005)	Gridpoint/FE (horizontal)&FD(vertical)/NH (Coté et al, 1998a,b; Girard et al, 2014)

Deleted: i

ICON	TERRA: Heise et al., (2006)	transfer-resistances approach: Baldauf et al., (2011)	TKE Baldauf et al., (2011) and Raschendorfer (2001)	Lott and Miller (1997)	mass-flux for deep, shallow and mid-level convection: Tiedtke (1993) and Bechtold et al. (2008)	Single moment scheme with four hydrometeors (Seifert, 2008)	RRTM	Grid-point/FV/NH
SLAV	ISBA 2L: Noilhan and Planton (1989) with modifications	Stability functions based on Cheng et al. (2002) with modifications leading to the absence of critical gradient Richardson number in the system.	TOUCANS (TKE+TTE) (Bašták-Ďurán et al 2014)	Scheme described in Catry et al., (2008) following Lott and Miller (1997) for gravity wave drag, and an envelope orography approach (after Wallace et al., 1983)	Mass flux following Bougeault (1982) but with modifications according to Gerard and Geleyn (2005)	Single moment scheme with four hydrometeors (Gerard et al., 2009)	Shortwave radiative transfer uses the CLIRAD model (Tarasov and Fomin, 2007) and RRTM for longwave	Grid-point/FD/H Tolstykh et al., (2017)
AROME-Arctic	SURFEX: Masson et al. (2013)	Based on Louis (1979)	HARATU: TKE together with a diagnostic length scale (Lenderink and Holtslag 2004; van Meijgaard et al. 2012)	N/A	Deep convection is explicitly represented and Shallow is represented by EDMF (Soares et al. 2004; Siebesma et al. 2007, Bentsson et al. 2017)	Single moment with five categories of hydrometeor based on Pinty and Jabouille (1998) with modifications (Müller et al 2017)	RRTM (EcRad) With modified cloud optical properties compared to AROME-MF (Bengtson et al. 2017)	Spectral/FD/NH

Deleted: Bastak-Duran

Table 3. Details of physical processes and parameterizations of the forecasting systems (see Appendix A for list of acronyms).

3 Evaluation of basic surface meteorology and vertical profiles

3.1 Evaluation/Scores

As mentioned in the introduction, the combination of MODFs and MMDFs allow detailed process-oriented diagnostics to be performed for the models. However, it is first important to assess what the errors are for standard variables such as 10m wind speed and 2m temperature. This first step is important because if they are stationary with leadtime one can simply consider a 24hr time range in the forecasts such as T+25 until T+48 (the second day of the forecast), simplifying the analysis.

The 2m temperature errors during February and March 2018 have quite different properties at each site and for each model (Fig 2). The models are typically too warm at Utqiaġvik and Tiksi and too cold at Ny-Ålesund and Whitehorse, with the sign of the bias varying between the models at Iqaluit and Eureka. At both Sodankylä and Whitehorse, which are situated at lower latitudes than the other sites, there is a distinct diurnal cycle in the bias and standard deviation that is not there at higher latitude sites. At both sites the night-time temperature bias is typically more positive than the daytime bias, indicating an underestimate of the diurnal temperature range. In the case of the CAPS and the IFS, the bias in the diurnal cycle at these observatories are representative of those seen over wider region (e.g. Casati et al., 2023 and Haiden et al., 2018).

In terms of wind speed, the forecasts all have a positive wind speed bias at Utqiaġvik and a negative bias at Iqaluit and Whitehorse (Fig 3). At Tiksi, Eureka, Sodankylä and Ny-Ålesund, the sign of the bias varies between the models. Interestingly, the largest inter-model spread and biases in wind speed is observed at the sites surrounded by the most complex orography (i.e. Iqaluit, Ny-Ålesund, Eureka and Tiksi: see Fig 2 of Mariani et al., 2024), likely due to the difficulties in representing the mesoscale flow patterns typically generated in such locations. Interestingly, there does not seem to be an obvious benefit from the increased resolution, with the AROME configurations and CAPS model actually having worse biases than the lower resolution global models at Ny-Ålesund.

Deleted: with

Deleted: Morris

Deleted: 3

Although there is some sub-daily variability with a diurnal frequency in the bias, more pronounced in wind speed bias (Figs. 2 and 3), the size of the biases does not grow dramatically with time. Thus, we consider a 24hr time range between the T+25 and T+48 forecast steps (i.e. the second day of the forecast) to be representative of the general error, simplifying the analysis.

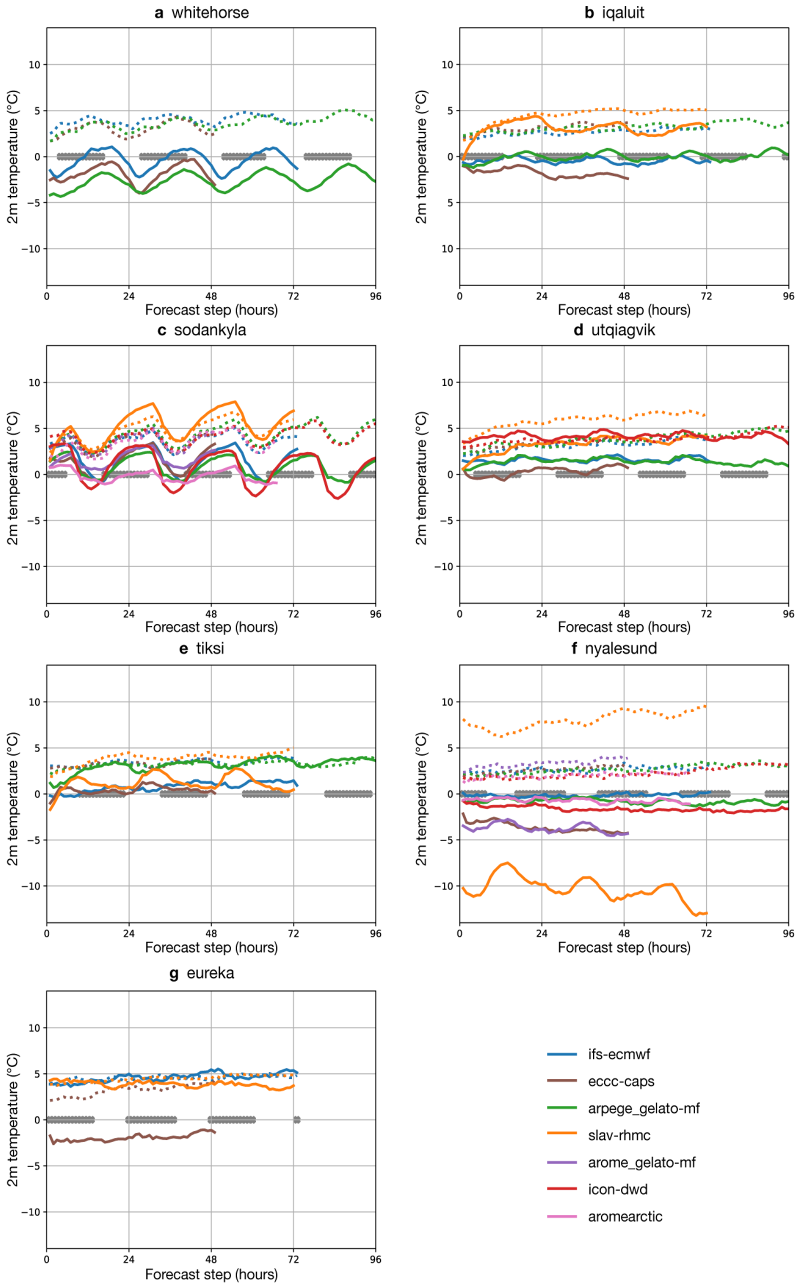


Figure 2: Mean bias (solid lines) and standard deviation (dashed lines) of the 2m temperature error (in °C) at each observatory (see Figure 1a) for forecasts initialised at 00z during SOP1, described in Table 2. Night-time periods (with mean $SW\downarrow<15Wm^{-2}$) are indicated with grey crosses along the x-axis.

Formatted: Superscript

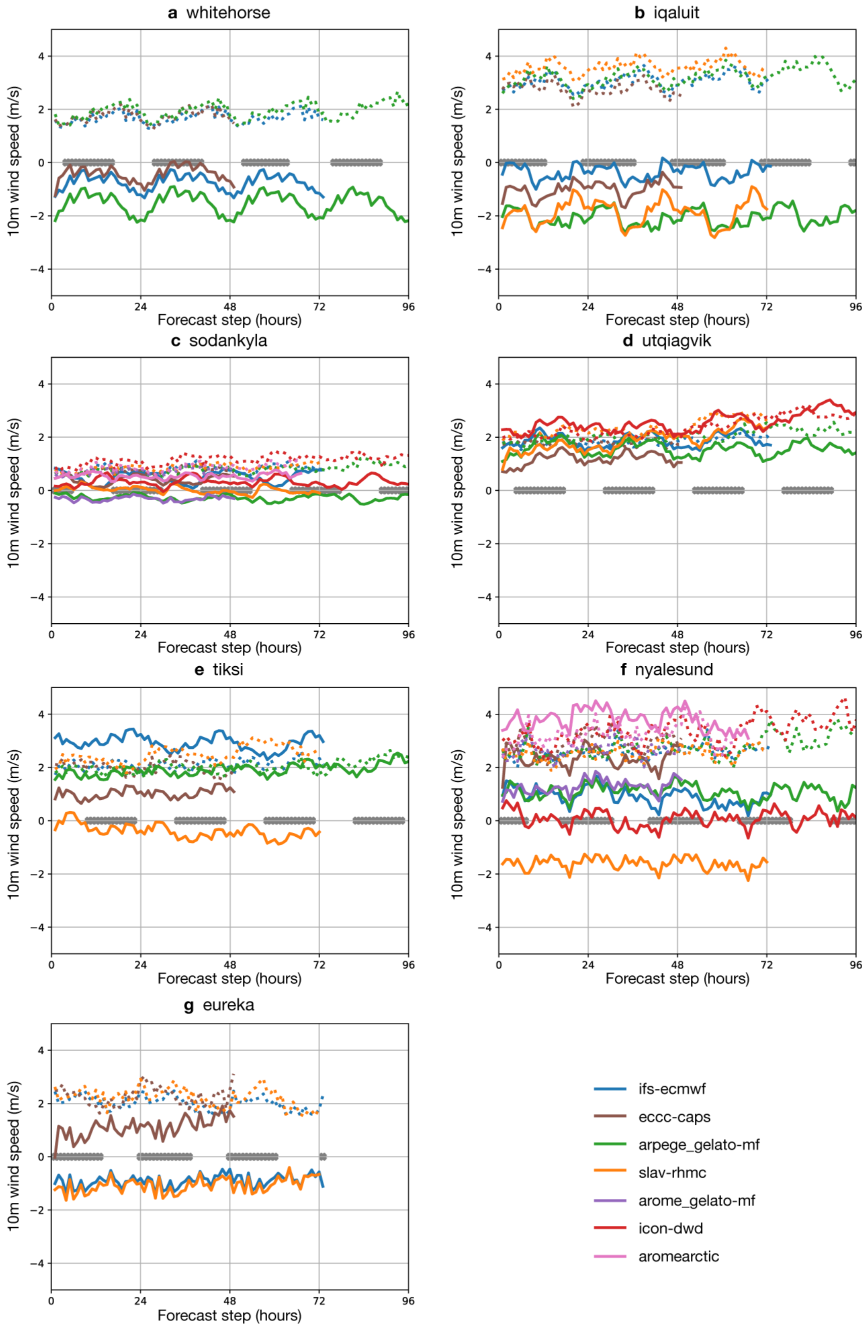


Figure 3: Mean bias (solid lines) and standard deviation (dashed lines) of the 10m wind speed error (in m s^{-1}) at each observatory for forecasts initialised at 00z during SOP1. <u>Night-time periods (with mean SW↓<15Wm^{-2}) are indicated with grey crosses along the x-axis.</u>

3.2 Vertical profiles

To gain further insights we investigate the vertical structure of the errors by comparing the model output to observations from radiosonde and tower. To do this the model and tower data were thinned to the same frequency as the radiosonde prior to calculating the median and inter-quartile range shown in Figs 4 & 5. The median temperature and specific humidity within the boundary layer is overestimated at Tiksi, Eureka, Utqiaġvik and Iqaluit (see Fig 4) and the models underestimate the strength of temperature and humidity inversions as a result. The picture is more mixed at Ny-Ålesund and Sodankylä where most models are too cold and humid, and two out of the three models are too dry at Whitehorse.

Deleted: Ny-Alesund

Deleted: all

The biases in the upper air temperatures, 2m air temperature, and the surface skin temperature tend to go hand-in-hand with each other, i.e. model with warmest/coldest surface temperature tends to have the warmest/coldest 2m and upper air temperatures. As a result, the mean 2m temperature errors seen in Fig 2 give a sense of the sign of the error in the lowest 100m, or so, of the atmosphere. This coupling between the lowest model level, the surface skin temperature and the 2m-temperature is to be expected, since the 2m-temperature is a diagnostic calculated as a function of the lowest atmospheric model layer and the surface skin temperature.

Air temperature variability in the lower boundary layer is generally underestimated by the models, except at Iqaluit (Fig 5). This generally translates to an underestimation of the 2m temperature variability at these sites. Interestingly, at Ny-Ålesund some models severely overestimate the 2m temperature variability despite underestimating the variability aloft, possibly due to the overestimation of the surface skin temperature variability. For specific humidity the observed inter-quartile-range tends to sit within the range of the models, however it is over-estimated at Eureka and underestimated at Tiksi and Whitehorse in the lower boundary layer.

Deleted: Ny-Alesund

Deleted: variance

Deleted: variance

The median of the modelled wind speed is too high in the boundary layer at Sodankylä, Utqiaġvik and Tiksi, but more mixed at other sites (Fig 4 & 5). The variability of the wind speed is within the model range, with the exception of Iqaluit, where it is underestimated. The overestimation of the wind speed at these sites is likely a contributing factor in the underestimation of the temperature and humidity inversions, since a positive bias in the wind speed will drive excessive turbulent mixing of heat and moisture inhibiting the decoupling of near-surface and upper air temperatures that occurs during periods of radiative surface cooling and low wind (Van de Weil et al., 2017). Other factors which could play a role are the radiative forcing at the surface or the response of the surface to radiative forcing. Both aspects will be addressed in the following subsection.

Deleted: variance

Deleted: Iqualuit

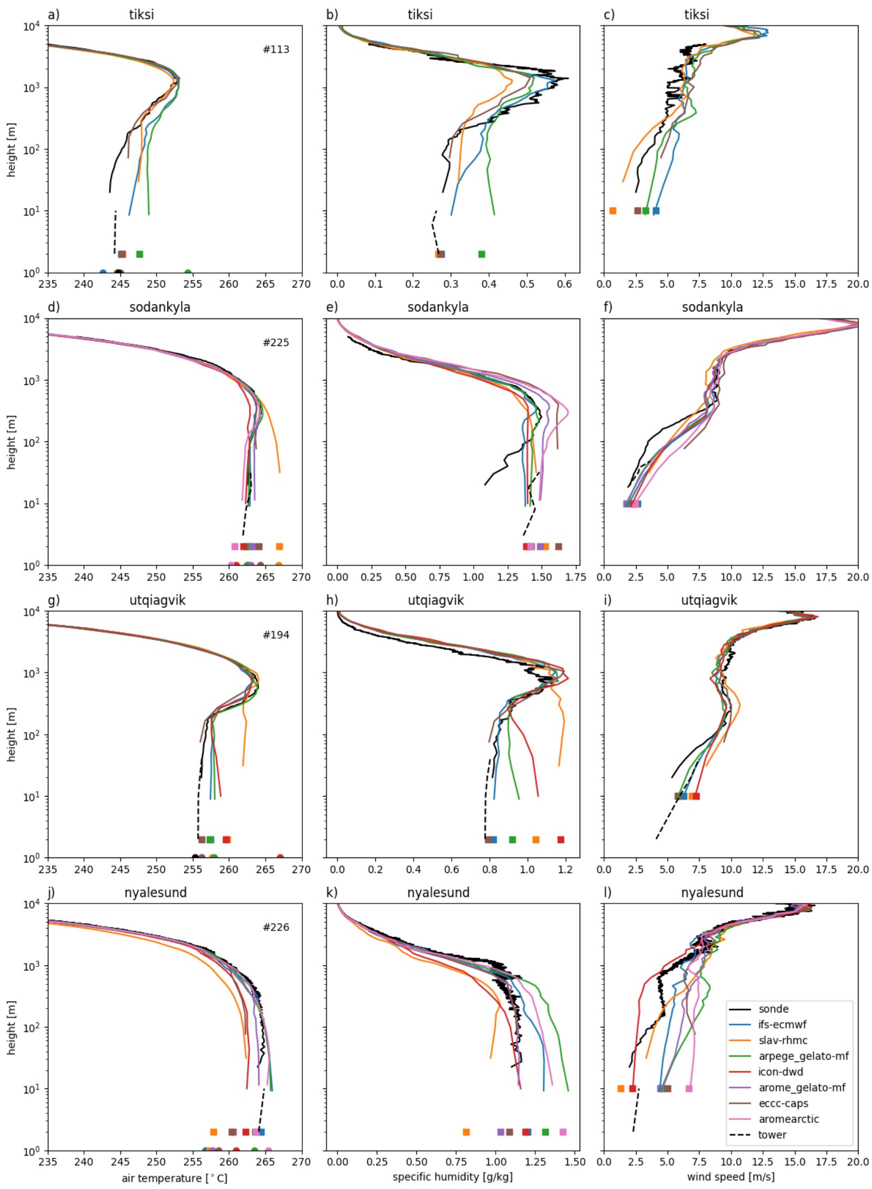


Figure 4: Median temperature (left), specific humidity (middle) and wind speed (right) from the radiosonde (black solid line), the tower (black dashed line), and the numerical models (during the second day of the forecast: colour lines). The mean surface skin temperature is indicated by a dot, 2m temperature (left), 2m specific humidity (middle) and 10m wind speed (right) are shown with a square. Note that wind speed and humidity profiles from the tower are not available in the Tiksi and Ny-Ålesund MODFs respectively. The numbers in the left hand panels correspond to the verification sample size, which was dictated by the availability of radiosonde profiles.

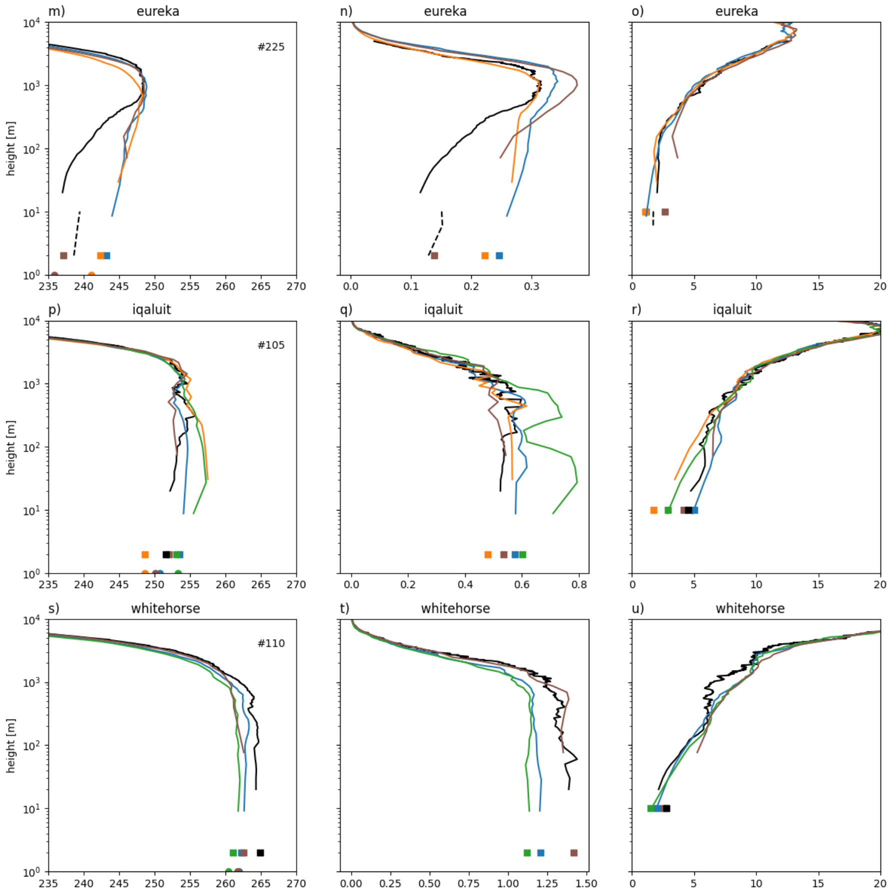


Fig 4 continued.

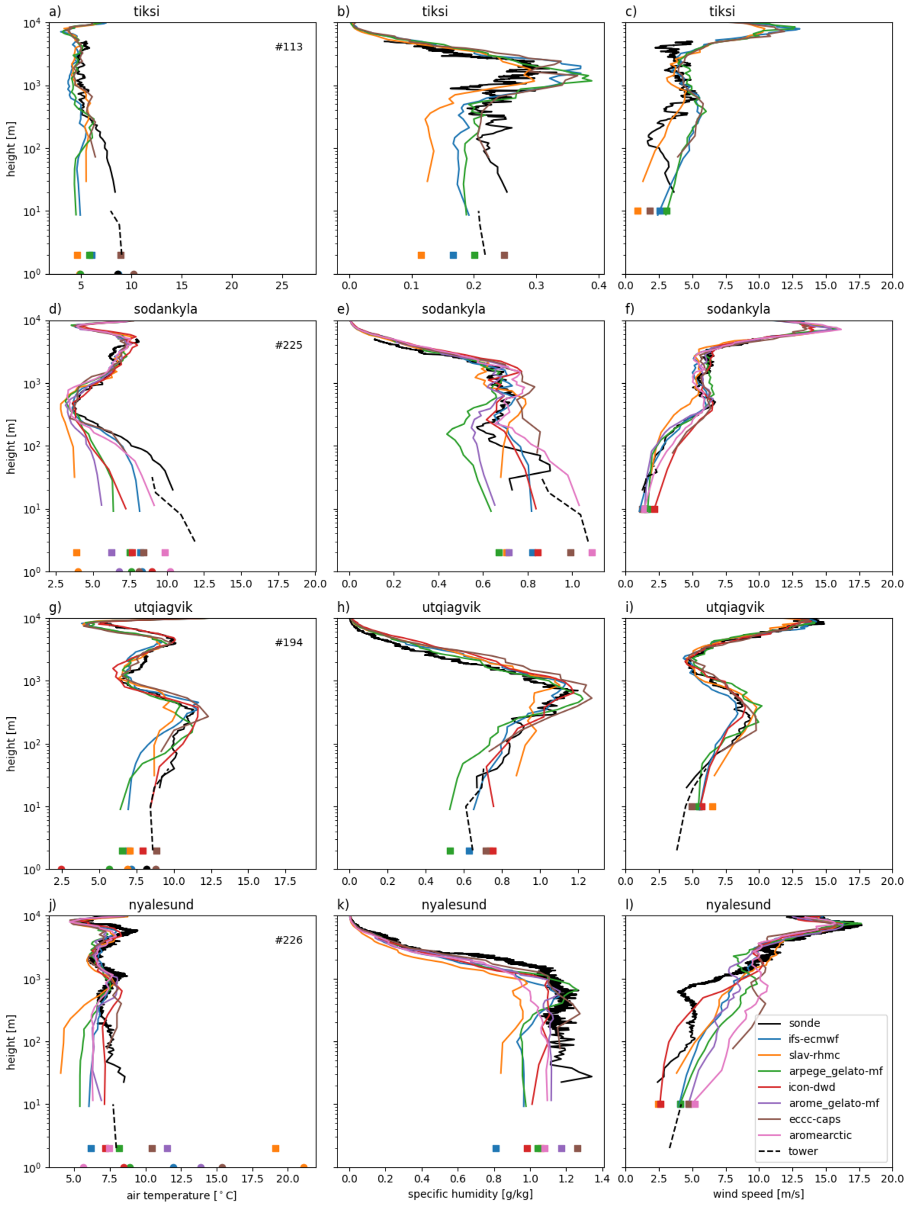


Figure 5: As Figure 4 but showing the Inter Quartile Range.

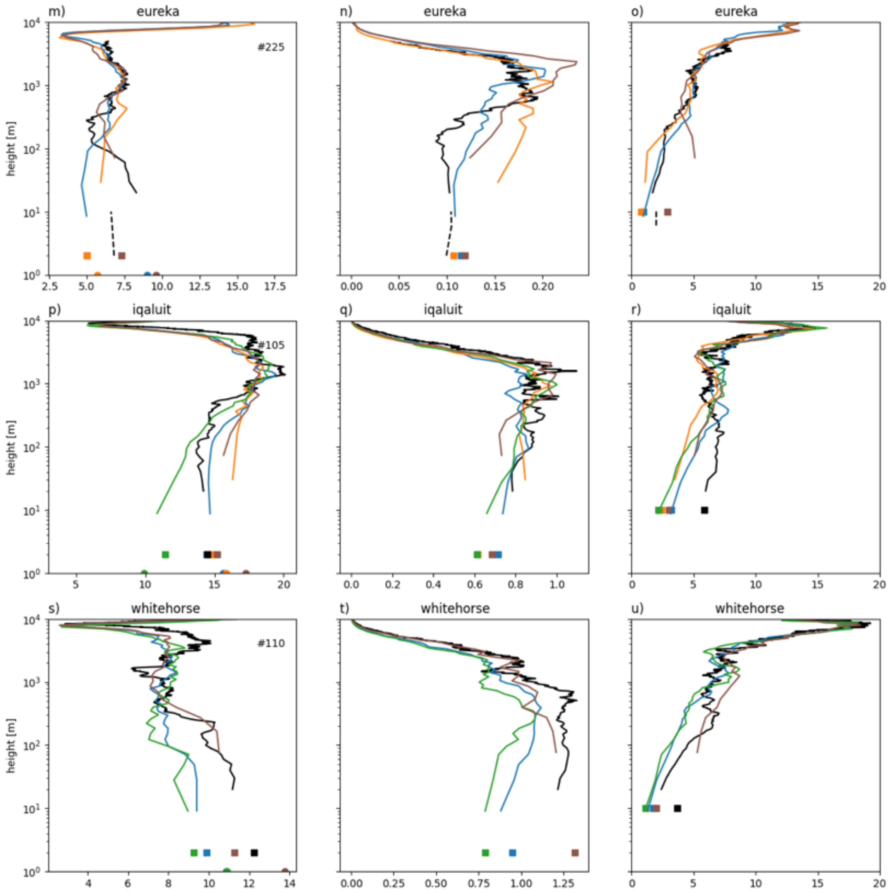


Fig 5 continued.

3.3 Links between errors in boundary-layer temperature variability and surface radiation.

Deleted: nce

In this section we investigate the role of radiative forcing in the underestimation of near-surface and boundary-layer temperature variability at Sodankylä, Utqiaġvik and Tiksi where the models underestimate the temperature variability. At these sites all upwelling and downwelling radiation components are available in the SOP1 MODFs allowing us to investigate whether the suppressed temperature variability is related to suppressed variability in the radiative forcing at the surface, a lack of sensitivity of the near-surface temperature to radiative forcing or something else.

Deleted: supressed

The box-plots shown in Fig 6a-c confirm the underestimate of near-surface-temperature Inter-Quartile Range (IQR) at Tiksi (except CAPS), Sodankylä, and Utqiaġvik, and further show that the cold tail of the distribution is generally shorter in the models meaning there is a warm bias during cold periods. The warm bias in cold conditions is well known at Sodankylä and is typical of NWP systems (see Atlaskin and Vihma, 2012 and Day et al., 2020), but this feature has not been shown before at the other two sites to our knowledge.

Deleted:

The models typically also show differences in the distribution of the downwelling radiation at the surface, $LW\downarrow + SW\downarrow$ compared to observations (Fig 6d-f). The IQR is underestimated at Tiksi (except for CAPS) and Utqiaġvik. However, at Sodankylä all the models overestimate the IQR (except for CAPS) but also do not capture the highest values of incident radiation observed at the top of the distribution. Since errors in the incident radiation likely relate to interactions with clouds, which are not included in this iteration of the MODFs, we will not investigate the causes of these discrepancies between the observed and forecast radiation distributions further, leaving this for a more focussed future study, and will instead move on to focus on the response of the near-surface air temperature and the surface energy budget.

Deleted: In this study we

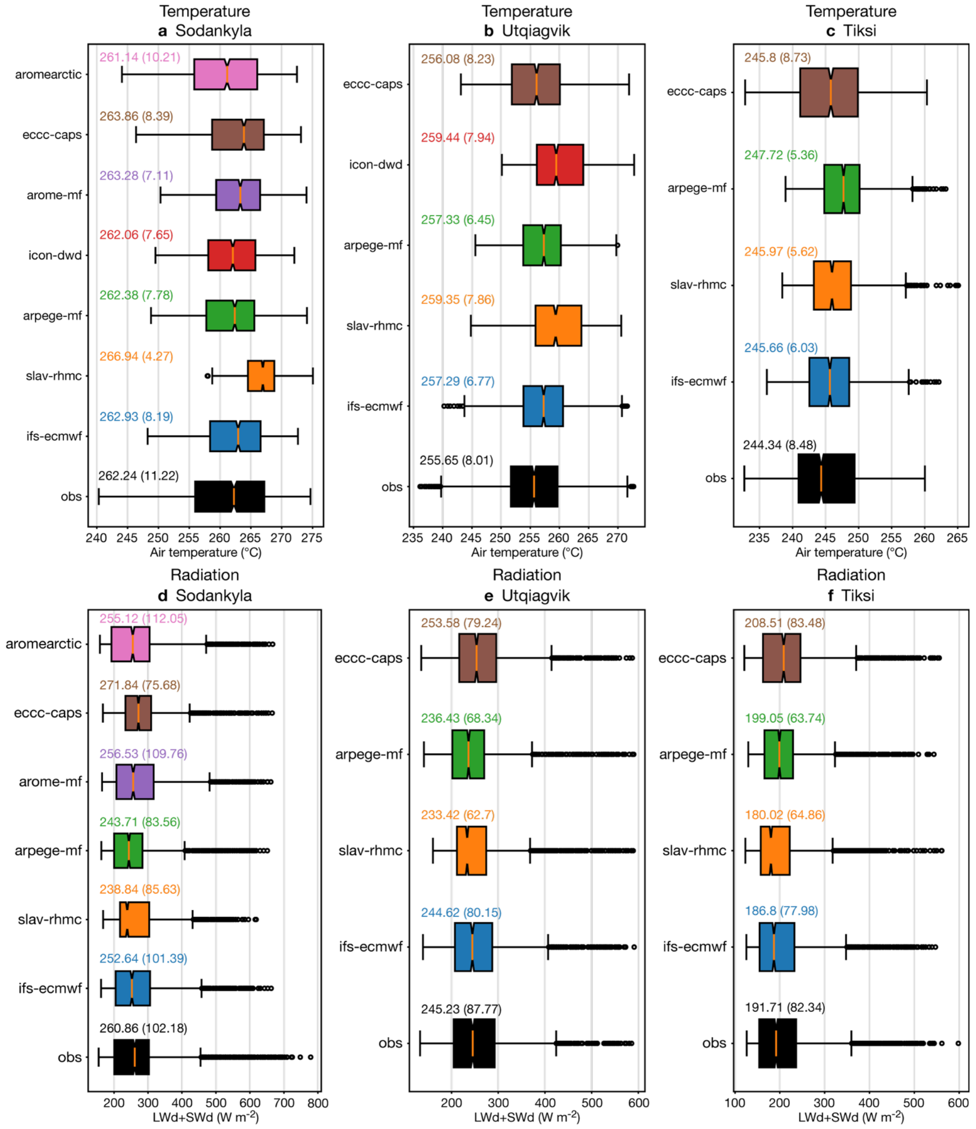


Fig 6. Boxplots of T2m (a-c) and LW↓+SW↓ (d-f) for Sodankylä, Utqiaġvik and Tiksi in observations and during the second day of the forecast. The text above the boxplots states the median (and inter-quartile-range) of each distribution, which are also shown by the orange line and box edges respectively. The 5-95% range is plotted by the whiskers and points outside this are shown in dots.

As $LW\downarrow + SW_{net}$ is the effective radiative forcing for the surface skin temperature (and indirectly for the 2m temperature), errors in 2 m air temperature are either due to errors in this driving term itself, the relationship between $LW\downarrow + SW_{net}$ and 2 m

Deleted: to

Deleted: a major driver of 2 m air temperature

temperature, or a more likely combination of both (assuming that errors in advection are negligible). Because the model median surface albedo (except for SLAV at Tiksi) is close to the observed estimate (Fig 7), then we can focus on how 2m temperature varies as a function of $LW\downarrow + SW_{net}$, to more deeply investigate the causes of error.

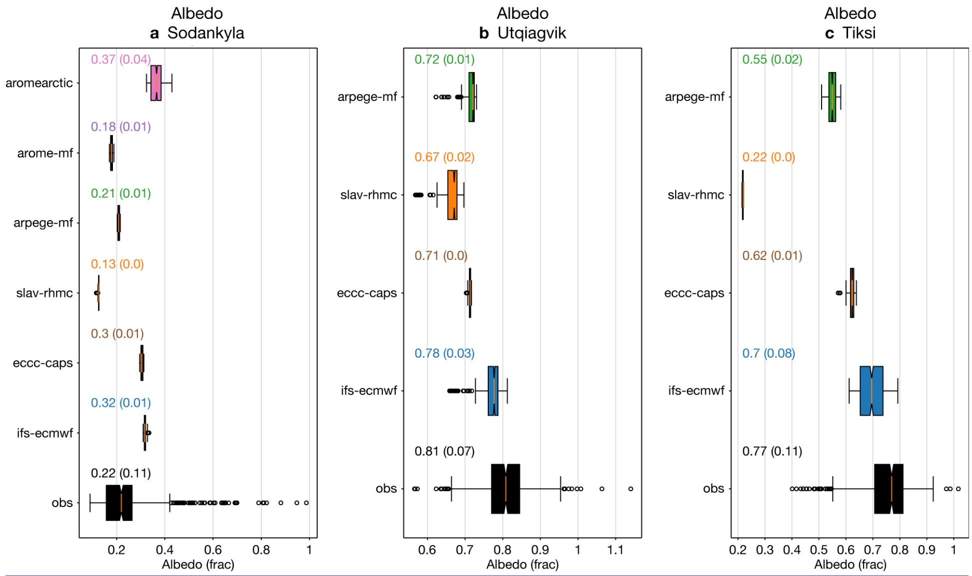


Figure 7. Boxplots of surface albedo for Sodankylä, Utqiaġvik and Tiksi in observations and during the second day of the forecast. The text above the boxplots states the median (and inter-quartile-range) of each distribution, which are also shown by the orange line and box edges respectively. The 5-95% range is plotted by the whiskers and points outside this are shown in dots.

At Sodankylä, Tiksi and Utqiagvik all the models have a warm 2m temperature bias at low levels of incoming radiation ($LW\downarrow + SW_{net}$) (see Fig 8). At Tiksi, Utqiaġvik and Sodankylä the overall sensitivity of T2m to radiative forcing, as measured by the slope of the regression coefficient between 2m-temperature and $LW\downarrow + SW_{net}$ is underestimated in all the models with one exception. The AROME-Arctic model seems to be too sensitive at Sodankylä according to this diagnostic, but captures the observed temperature range at low levels of $LW\downarrow + SW_{net}$.

Note that the LW components used for Sodankylä in this study, are not those provided in the SOP1 MODF, which are collected at the top of the 45m tower, rather they are from a dedicated radiation tower located near the sounding station where the downwelling component is at a height of 16m and the outgoing is at 2m. These were swapped due to a concern over the accuracy of the LW radiation data collected at the met tower (Roberta Pirazzini, personal communication).

Deleted: we find that, with the exception of the SLAV model at Tiksi, …

Deleted: investigate

Deleted: So looking at how the

Deleted: can

Deleted: provide additional information on

Commented [1]: why the observed albedo sometimes exceeds 1? why the spread of the obs albedo is so much larger than the one for the models?

Commented [2R1]: I think this is because when the sun is close to the horizon, as is at this time of the year, the "upward" sensor at 2m will sometimes be getting direct sunlight. I wonder if it is worth mentioning this. In the models it's just a lookup table for the surface type. I don't have a good explanation for the spread. Other than observations being noisy. I suspect it may not reflect true variability in the actual surface albedo. Although I suspect the true answer is somewhere in the middle.

Deleted: ¶

Deleted: conditional

Deleted: 7

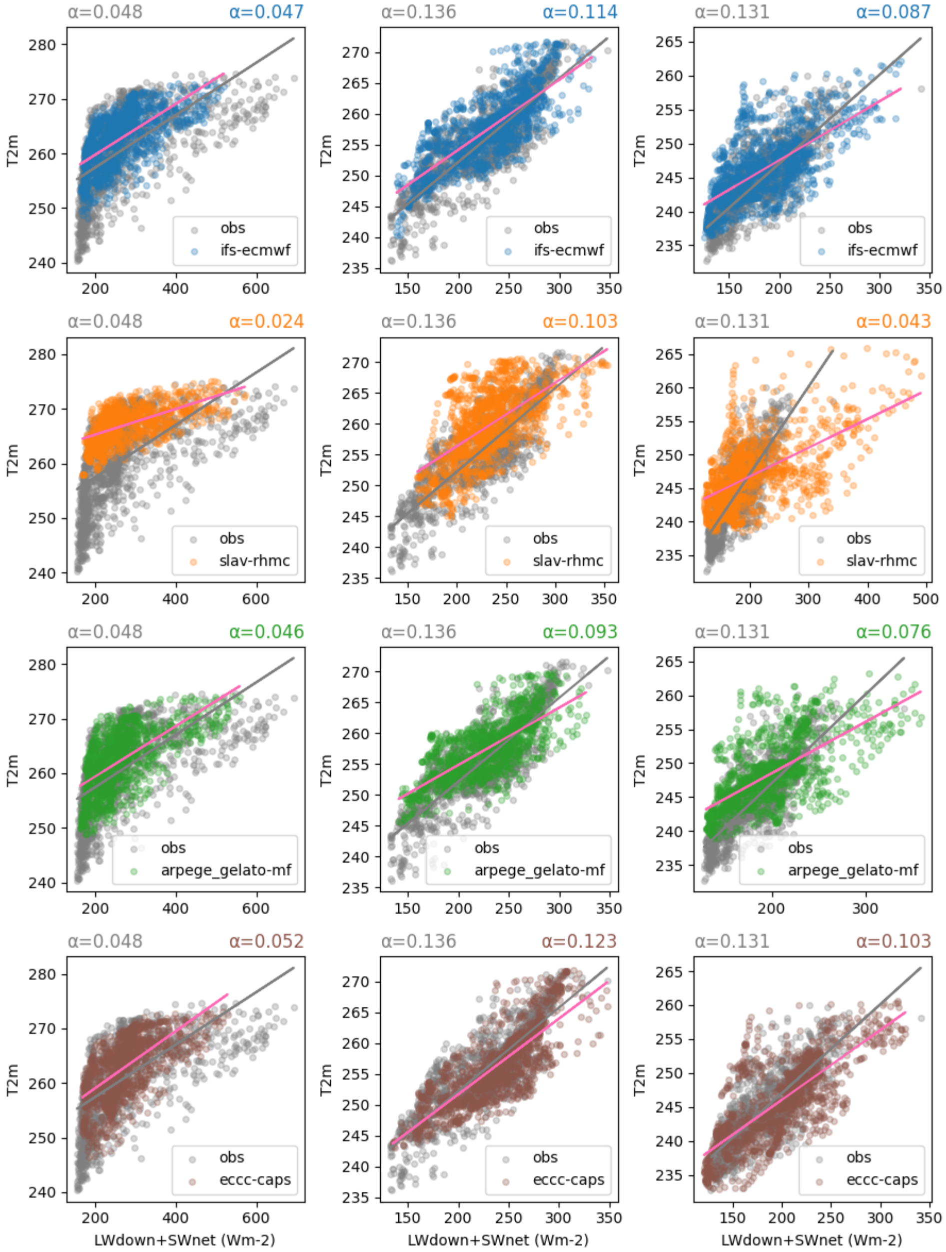
Deleted: is actually

Deleted: ,

Commented [3]: The relationship between T2m and LWdown+SWnet is not linear in Sodankylä, not for observations and not for models. There is snow on the ground which limits the temperatures to rise much above 0 degC no matter how much radiative forcing there is. When fitting a linear line into the data, it indeed looks like AROME-Arctic is too sensitive. But this is mainly because lack of high values of LWdown+SWnet, and not because the model is too sensitive in the cold end.

Commented [4R3]: It would actually be interesting to have a look at these sensitivities conditionally by stability (Ri) or simply binned by intervals of LWdown+SWnet. I am not asking you to revise the paper at this stage, especially because reviewers didn't comment this part, but concluding that AROME-Arctic is too sensitive is a bit contrary what we see.

Commented [5R3]: Agree this would be interesting to look at.



Fig<u>ure</u> <u>8</u>: Scatter plots of 2m temperature as a function of $LW\downarrow + SW_{net}$ for Sodankylä, Utqiaġvik and Tiksi (from left to right)<u>, for the second day of the forecast</u>. The regression slope between the 2m temperature and the $LW\downarrow + SW_{net}$ is stated in the title, for the observations (in grey) and each model (various colours).

Deleted: 7

Deleted: during the second day of the forecast

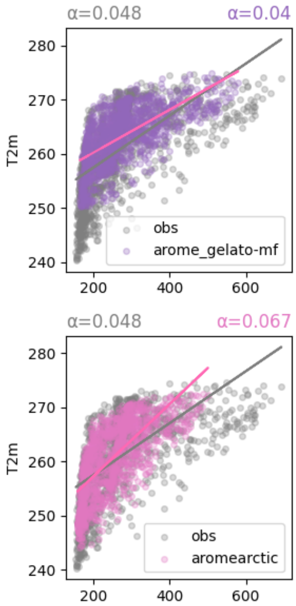


Figure 8 cont.

To investigate the role of surface-atmosphere decoupling in the 2m-temperature cold-tail warm bias and lack of 2m-temperature variability at low levels of incident radiation we plot the thermal stratification as a function of near-surface wind speed at the three sites (Fig 9) for situations where the model or observed $LW\downarrow + SW_{net}$ is below the 20th percentile. In the observations one can see the typical pattern seen at other sites (e.g. Ven de Weil et al., 2016) that inversions are weak for strong winds, whereas large inversions are found under weak-wind conditions with a transition found between those regimes at some critical wind speed. The models generally capture this qualitative regime behaviour (Fig 9), although the magnitude of the thermal stratification, the wind speed and the critical wind speed for the regime transition varies between the models.

Deleted: 7

Formatted: Superscript

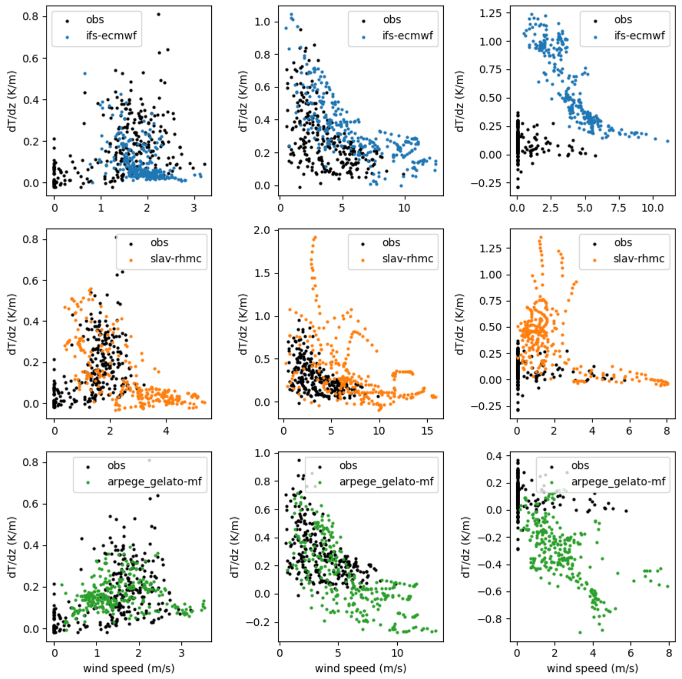
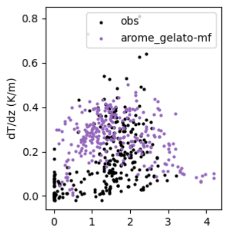


Figure 9. Scatter plots of thermal stratification ((T_{2m}-T_{lml})/height) as a function of wind speed on the lowest model at Sodankylä, Utqiaġvik and Tiksi (from left to right) for the observations (in black) and each model (various colours) during the second day of the forecast for situations where the model or observed $LW\downarrow + SW_{net}$ is below the 20th percentile.



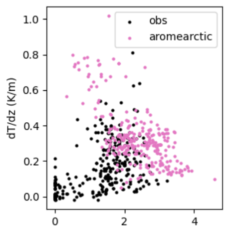


Figure 9. continued.

Formatted: Subscript

Deleted: grey

Formatted: Font: Italic

Formatted: Font: Italic, Subscript

3.4 Surface energy budget sensitivity to radiative forcing

Further insight into the role of the land-surface and surface exchange processes in the T2m errors outlined in the previous section, particularly the lack of T2m sensitivity to radiative forcing, can be gained by constructing surface energy budget sensitivity diagrams, following Miller et al. (2018) and Day et al. (2020). The idea here is that the surface energy budget can be separated into a "driving term" ($LW\downarrow + SW_{net}$) and "response terms" (*SHF, LHF, GHF*, and $LW\uparrow$). The relationship between the driving term and each response term can be summarised with regression coefficients, e.g. for the *SHF*:

$$SHF = \alpha_{SHF}(LW\downarrow + SW_{net}) + \beta_{SHF} \quad (1)$$

where each of the α's can be interpreted as a coupling strength parameter between the driving term and each response term. These α's provide direct information on the proportional response of each flux term, expressed as a fraction of the total change in radiative forcing. From this one can see that if, for example, the coupling to the ground heat flux and turbulent fluxes is too strong in the model (i.e. $|\alpha_{GHF_{mod}} + \alpha_{SHF_{mod}} + \alpha_{LHF_{mod}}| > |\alpha_{GHF_{obs}} + \alpha_{SHF_{obs}} + \alpha_{LHF_{obs}}|$) then $|\alpha_{LW\uparrow}|$ will be too small, i.e. surface temperature response will be too weak and vice versa. Similarly, compensating errors in the strength of the coupling to the turbulent fluxes ($\alpha_{SHF_{mod}} + \alpha_{LHF_{mod}}$) and ground heat flux($\alpha_{GHF_{mod}}$) could result in the right surface-temperature sensitivity, $\alpha_{LW\uparrow}$, but for the wrong reasons. As a result, by comparing the observed and modelled regression coefficients one can derive physical understanding of the causes of model error.

Note that in convective cases - the main driver of turbulent heat fluxes is indeed the convective instability at the surface driven by radiative forcing. However, in stratified conditions the main driver of turbulence in the boundary layer (and of the sensible and latent heat fluxes) is the mechanical forcing i.e. the large-scale wind speed (Van Hooijdonk et al. 2015, Van de Wiel et al. 2017, Vignon et al. 2017). As a result, one expects the turbulent fluxes to have little sensitivity to the radiative forcing in stable conditions, with the ground heat flux taking a larger role in balancing changes in radiative forcing and the converse in convective cases (see Day et al., 2020). As a result, at Utqiaġvik and Tiksi where stable conditions dominate, the ground heat flux varies with changes in radiative forcing, more than the turbulent fluxes as indicated by higher regression coefficients. At Sodankylä there is more of an even partitioning between the turbulent fluxes and the ground heat flux into the snow.

It is clear from Figures 10, 11 and 12 that all the models generally underestimate the surface temperature sensitivity to radiative forcing at Sodankylä, Utqiaġvik and Tiksi, because the rate of change in $LW\uparrow$ with changes in radiative forcing, $LW\downarrow + SW_{net}$, i.e. $\alpha_{LW\uparrow}$ is typically too low (i.e. $\alpha_{LW\uparrow_{mod}} < \alpha_{LW\uparrow_{obs}}$). Since the 2m temperature diagnostic in the models is calculated as a function of the surface skin temperature, the underestimation of the 2m-temperature and $LW\uparrow$ sensitivity to radiative forcing and the positive bias in those variables in cold conditions are likely to be closely related (i.e. comparing Fig 8 to Figs 10, 11 and 12). For example, at Sodankylä the CAPS model T2m and upwelling longwave ($LW\uparrow$) sensitivities are very close to what is observed, AROME-Arctic slightly overestimates these sensitivities and SLAV underestimates them. A similar proportionality can be seen between these properties of the models at the other two sites. Note that because the $LW\uparrow$ at Sodankylä was observed at 2m and so has rather a small footprint compared to the sensor on the 16m mast, the sensitivity is more representative of the bare snow than the forest canopy. As a result, one might expect the area mean $LW\uparrow$ sensitivity to be higher than the value presented here.

This mismatch in terms of $LW\uparrow$ sensitivity goes hand in hand with differences in the other α coefficients and by comparing the sensitivities of the other response terms in the surface energy budget we can develop some hypotheses about what is leading to this mismatch in surface temperature sensitivities. For example, at Utqiaġvik, all the models tend to overestimate the

Deleted: 8
Deleted: 9
Deleted: 0
Deleted: sensitivity parameter (the regression coefficient) for
Deleted: inability of the models to capture the lowest values of these …
Deleted: 7
Deleted: 8
Deleted: 9
Deleted: 0
Deleted: it

sensitivity of the *GHF*, α_{GHF}, which was calculated as the residual of the observed radiative and turbulent fluxes. This can be an indication of an indication of non-sufficient thermal representation of the land surface, for example lack of a multi-layer snow model (e.g. Day et al., 2020; Arduini et al., 2019). Unfortunately, we are not able to perform a similar calculation as performed for Sodankylä, to estimate the *GHF*, as the longwave observations thought to be most reliable, are not co-located with the other flux observations, or Tiksi, since we don't have the turbulent fluxes in the MODF. As a result, we cannot calculate the *GHF* as a residual of the other terms.

Where we have turbulent flux observations, we can also evaluate the α_{SHF} and α_{LHF} terms. At Utqiaġvik, an underestimation of the sensitivity of the turbulent fluxes, too low α_{SHF} and α_{LHF}in the ARPEGE and SLAV models goes hand in hand with an overestimation of α_{GHF} mentioned above. In the IFS and ECCC models are closer to observations with smaller values of α_{GHF} and larger values of α_{SHF} and α_{LHF}. At Sodankylä, the α_{SHF} varies quite a bit from model to model, but all the models where the LHF was available overestimate the α_{LHF}.

At all three sites the relative size of the coefficients varies between the sites, with $\alpha_{LW\uparrow}$, α_{SHF}, α_{GHF} typically being an order of magnitude larger than α_{LHF}. This is likely to be typical of cold dry snow-covered environments where the magnitude of the latent heat flux is low. However, the difference in the relative size of the other three terms varies quite a bit between sites with, for example, the turbulent flux playing a larger role at Sodankylä than at Tiksi and Utqiaġvik at this time of year. This reflects the larger surface roughness at Sodankylä associated with the trees at this site.

Before moving on it is worth noting that as well as being used to develop hypotheses about the causes of errors related to the surface energy budget, these process diagrams and sensitivity metrics could also be applied to test new configurations of NWP systems with modifications to the land-surface, boundary layer or related schemes and evaluate whether such modifications are improving the dynamic behaviour with respect to the surface energy budget in line with observed behaviour or not.

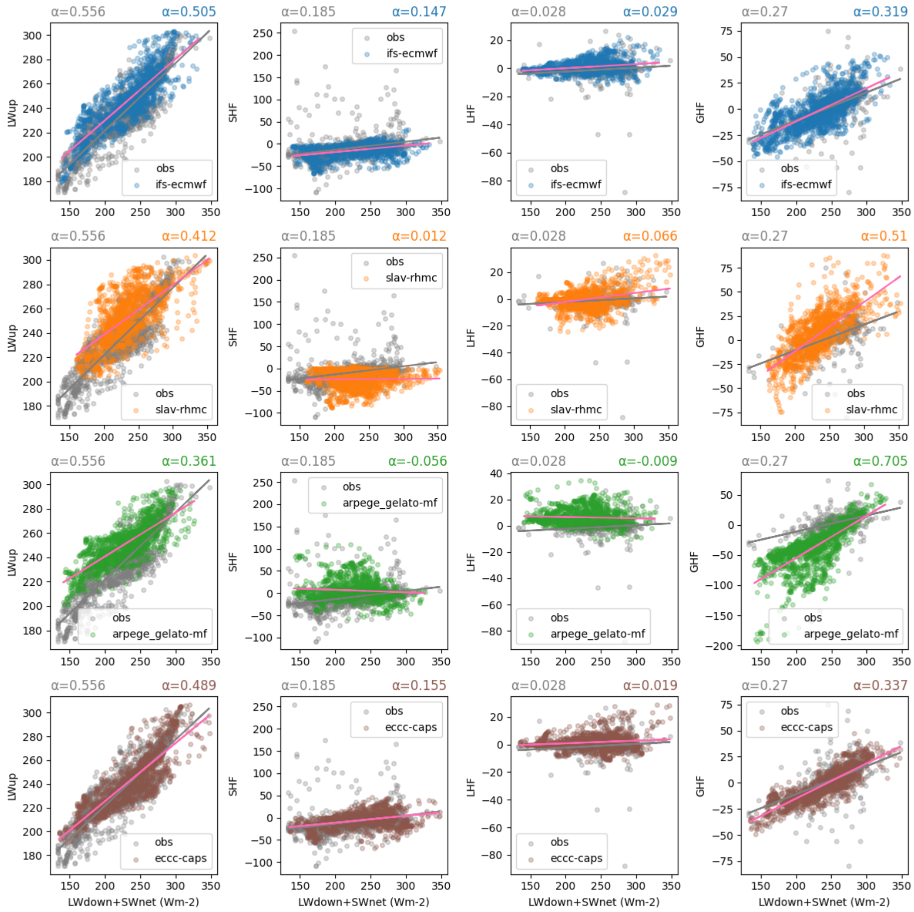


Figure 10: Process relationship diagrams and sensitivity parameters for upwelling longwave radiation (LWup; left), sensible heat flux (SHF; middle left), latent heat flux (LHF; middle right) and ground heat flux (GHF; right) at Utqiaġvik. Observed values are shown in grey, model values during the second day of the forecast are shown in colour. The line of best linear fit is shown for observations (gray line) and each model (pink line). The sensitivity parameters, a, describing the coupling strength between the driving ($LW\downarrow + SWnet$) and each response term are printed above each diagram, with observational (modelled) relationship on the left (right).

Deleted: 8

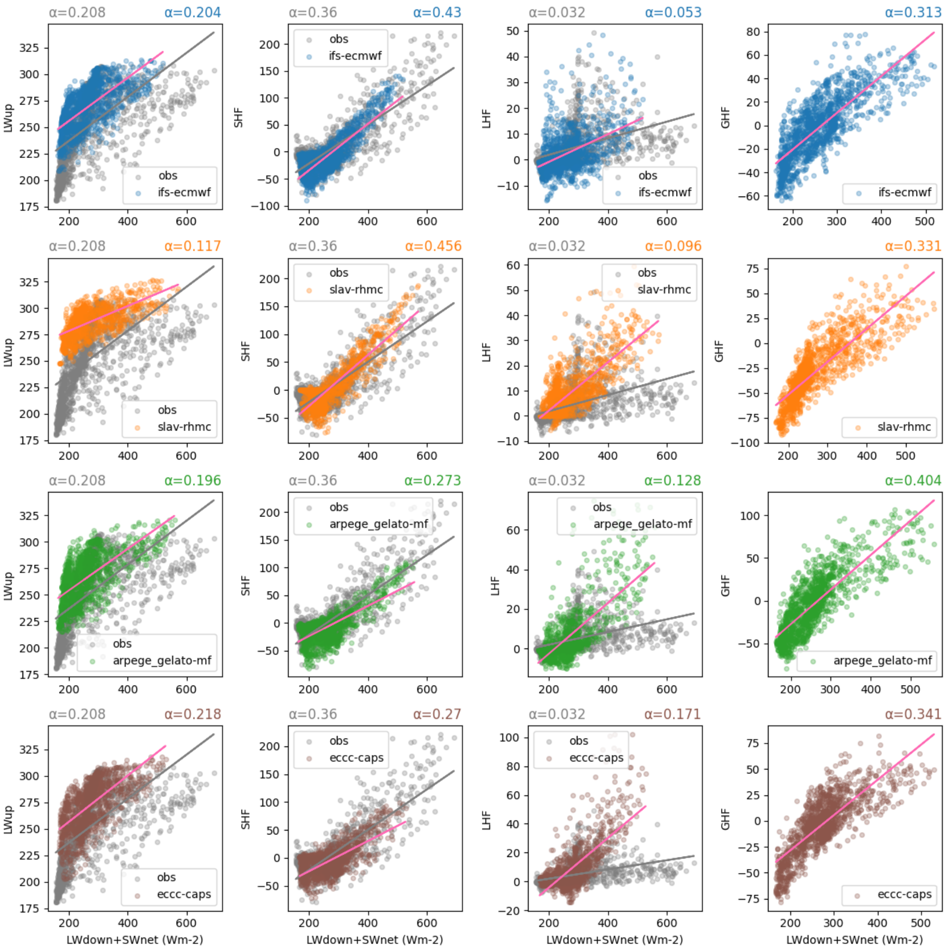


Figure 11: Same as Figure 8 but for Sodankylä.

Deleted: 9

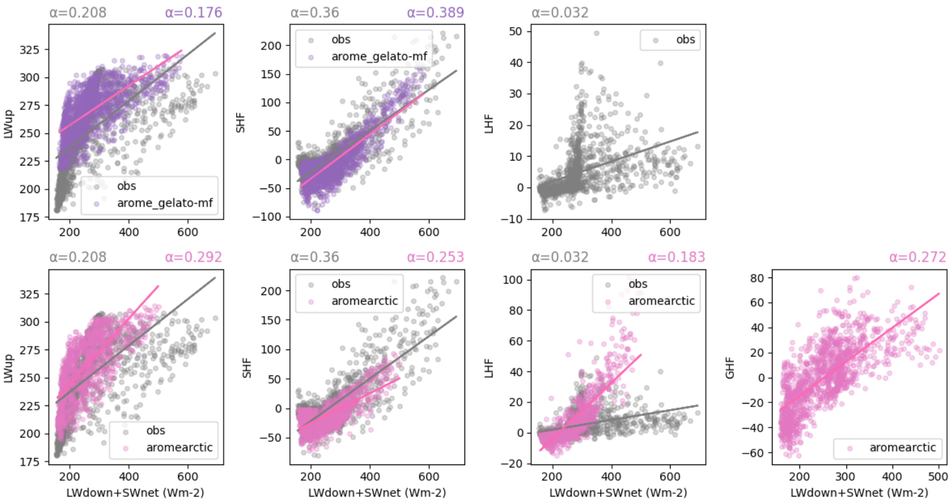


Figure 11: cont.

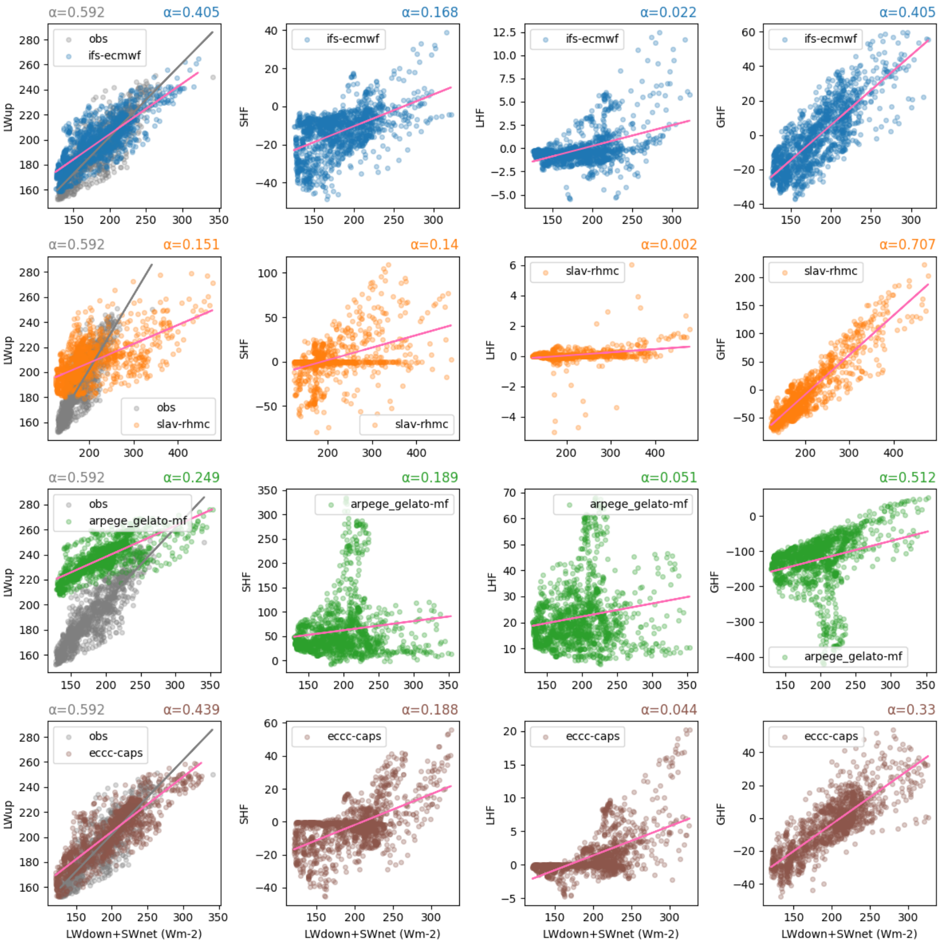


Figure 12: Same as Figure 8 but for Tiksi.

3.5 Evaluation of wind stress and sensible heat flux

The previous examples highlight discrepancies between forecast and observations and provide hints as to which processes are responsible for the documented errors. The observed conditions also provide multi-variate targets for updated forecasting systems. However, the observations can also help us evaluate a specific process and thereby target a specific parameter or parameterization to change.

The Sodankylä and Utqiaġvik MODFs include turbulent fluxes and profiles of wind speed and temperature allowing us to investigate the parameterisation of turbulent exchanges of heat and momentum at the surface. Turbulent surface fluxes in NWP models are often parameterised according to Monin-Obukhov (M-O) similarity theory where they are related to the gradient in the lowest atmosphere (e.g. Beljaars and Holtslag, 1991):

$$\tau = \rho C_M U_{ref}^2 \#(2)$$

$$SHF = \rho C_H U_{ref}(\theta_{ref} - \theta_{sfc}) \#(3)$$

Deleted: 0

Deleted: p

where τ is the wind stress, U is the wind speed, θ is potential temperature, ρ is the air density and the transfer coefficients, C_M and C_H, used to in each computation, are a function of the roughness length of momentum and heat, z_{oM} and z_{oH}, and a stability parameter. In these equations the U_{ref} and θ_{ref} are the wind speed and potential temperature at a reference height, which in the case of the models is the lowest atmospheric model level, the height of which varies from around 10 to 30 m above the surface depending on the model (see Table 3).

Successfully parameterizing τ and *SHF* relies on defining a reasonable function for C_M and C_H and selecting the appropriate parameters and a proper aggregation of the fluxes in the cases of a tiled surface. Because we have observed and forecast values for both the fluxes and the bulk parameters in equations 2 and 3 we can diagnose how appropriate the choices in each model are for the conditions at a particular site. This is done by examining the relationship between the bulk parameters, U and θ, and the fluxes τ and *SHF* (see Figures 13 to 16), as done previously by Tjernström et al. (2005) and more recently by Day et al. (2020).

Deleted: 1

Deleted: 4

In the case of wind stress, in neutral conditions, the points in Figures 13 and 14 would sit on the straight line following:

$$\tau = \rho \frac{k^2 U^2}{\left[ln\left(\frac{z_{ref}}{z_{0M}}\right)\right]^2}, \qquad \#(4)$$

Deleted: 1

Deleted: 2

where z_{ref} is the height of the lowest model level, k is the von Karman constant and z_{0m} is the aerodynamic roughness length. The slope of this line is determined by z_{0m}. However, this formula provides an overly simplified view as the atmospheric stability varies from neutral conditions and as a result there is scatter in the values of τ for any given wind speed.

The relationship between τ and U for Sodankylä (Figure 13) differs between the models and between the models and the observations. An estimate of the observed roughness length was calculated, following the equation above, after selecting for neutral conditions, and the value is presented in Table 4 along with the value used in each of the models. In the AROME-Arctic and ICON models, τ increases too slowly with increasing U. This is consistent with the fact that the roughness length for momentum is too low in these models, which have roughness lengths an order of magnitude lower than that derived from observations (see Table 4). Increasing z_{0m} in the AROME-Arctic and ICON models would likely reduce the positive bias in the wind median wind speed profile seen in Figure 4, however the other models which have roughness lengths closer to what was observed also have a positive wind speed bias suggesting another cause.

Deleted: 1

Deleted: also

Interestingly, all models fail to adequately capture the spread of τ for a given value of U, likely because the models underestimate the atmospheric stability as is suggested by the weaker than observed thermal stratification indicated by in Figs 4d and 5d. A more detailed study including numerical experimentation would be needed to demonstrate this further.

At Utqiaġvik, the aerodynamic roughness length is three orders of magnitude lower than at Sodankylä, reflecting the difference in surface type: snow covered tundra compared to the forested taiga of northern Finland (Table 4). Here the IFS and SLAV models have roughness lengths close to those derived from observations, whereas the ARPEGE and ICON have values that are higher. As a result, for a given wind speed the surface stress is too high in these two models (Figure 14).

Deleted: 2

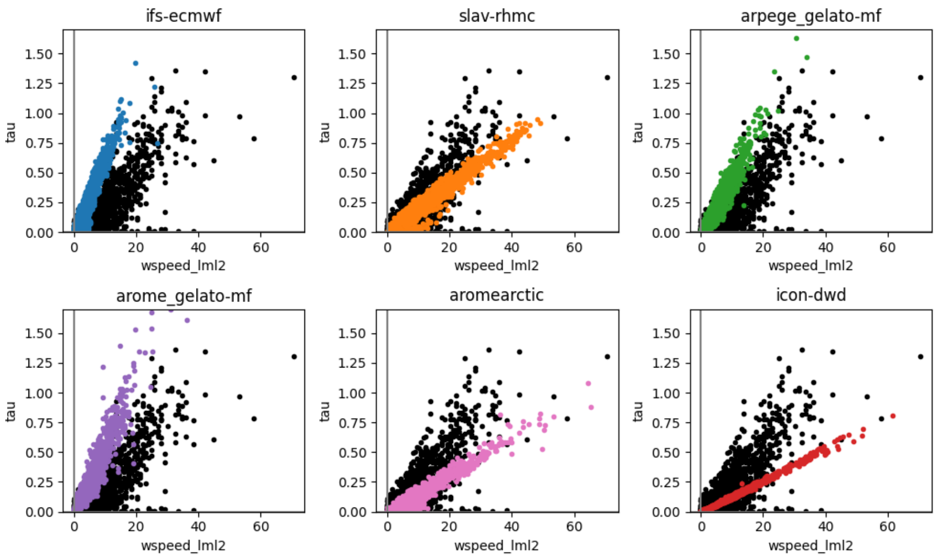


Figure 13: scatter plots of wind stress vs. the square of the near-surface (lowest model level) wind speed at Sodankylä. The observed points are shown in black and hourly values during the second day of the forecast forecast is shown in colours.

Deleted: 1

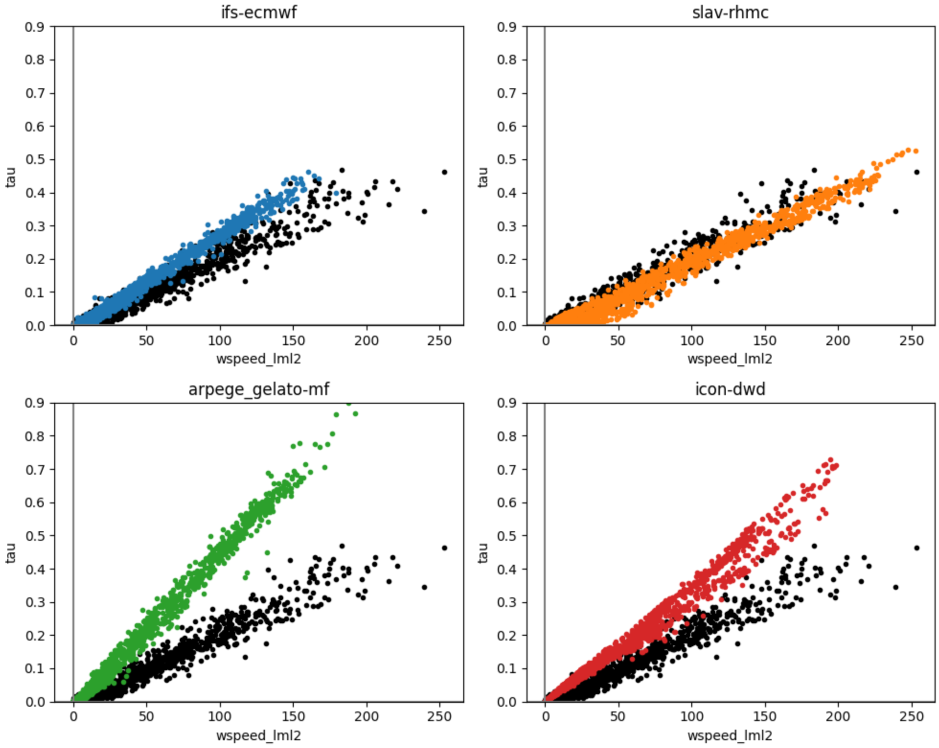


Figure 14: as Figure 13 but for Utqiaġvik.

	Sodankylä	Utqiaġvik
Obs	1.62	0.0012
IFS	1.83 (1.83-1.83)	0.00130 (0.00130-0.00130)
ARPEGE	1.50 (1.49-1.51)	0.00884 (0.00880-0.00891)
SLAV	1.60 (1.59-1.61)	0.00135 (0.00129-0.00144)
ICON-DWD	0.20 (0.20- 0.41)	0.00700 (0.00151-0.00981)
AROME-Arctic	0.45 (0.45-0.45)	Outside model domain

Table 4. Roughness lengths for momentum (m) at Sodankylä and Utqiaġvik from observations and models. For the models the mean is stated and the range of values is stated in parenthesis.

The scatterplots for the sensible heat flux (Figures 15,16) also provide some insights into the differences in the process representation between the models. All the models capture the link between the SHF and the temperature gradient dictated by M-O theory (see Eqn 3) however, the shape of the relationship varies between the models. For example, for the ARPEGE and AROME-MF models the sign of the sensible heat flux does not change in a binary way with ΔT, there is spread in the location along the x-axis where this occurs. This could be due to differences in the numerical formulation of the models, i.e. the timestep at which the flux and temperature terms are stored or due to the fact that we are looking at the gridbox mean values where the fluxes are aggregated from values computed on different surface tiles. At Sodankylä, the IFS, SLAV and AROME-ARCTIC

Deleted: 2

Deleted: 1

Deleted: 2

Deleted: 53

Deleted: 3

Deleted: 4

Deleted: The basic shape of the observed points is the same at both sites, but with fewer cases of instability at Utqiaġvik compared to Sodankylä. …

Deleted: -GELATO

model have a clear tapering in the scaled sensible heat flux towards zero for high values of ΔT. However, the AROME-MF, ARPEGE and ICON do not have such a tapering and the scaled heat flux continues to grow with larger ΔT, which is qualitatively inconsistent with the observations and will lead to higher fluxes in very stable conditions inhibiting cooling of the surface. There is also a clear difference in the range of ΔT between the different models however, in the models this is an aggregate of different surface types representing forest canopy top, bare snow and frozen water and because we do not have a trustable observation of the temperature of the top of the canopy frozen water during freezing conditions it is not clear what the realistic range should be. Note also that the SHF at Sodankylä is measured at 24.5 m and for process consistency ⊗T is calculated using the air temperatures observed at 18m and 32m which is not directly comparable with the models.

Except for ICON, differences between the models at Utqiaġvik are less pronounced. IFS, SLAV and ARPEGE have quite a similar shape, and all underestimate the magnitude of the scaled heat flux for low values of ΔT, potentially due to the slow bias in wind speeds near to the surface. Note that the large values of ΔT for the SLAV model are because the lowest model level is at ~30m, compared to ~10m for the other models. Note that the ICON model has a large fraction of open ocean in the grid cell considered and therefore the model tends to be biased towards convective conditions (i.e. most points are in the top left quadrant of Figure 16 where the sensible heat flux is heating the atmosphere), this is likely the main reason for the warm bias in surface skin-temperature and 2m-air temperature. For the other models shown in Figure 16, the grid-point considered is 100% land.

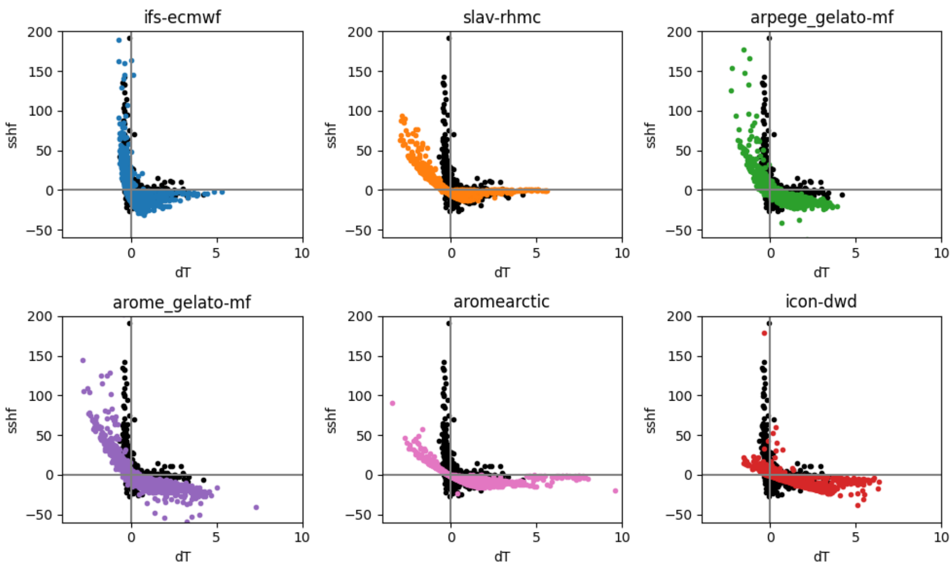


Figure 15: scatter plots of the scaled sensible heat flux (*SHF/U*) vs. thermal stratification, $\Delta T=T_{lml}-T_{skin}$, at Sodankylä. The observed points are shown in black and hourly values during the second day of the forecasts are shown in colours. Note that at Sodankylä the SHF is measured at 24.5 m and for process consistency ΔT is calculated using the temperatures observed at 18m and 32m so is not directly comparable with the models which use the skin temperature, T_{skin}, and the lowest model level, T_{lml}.

Deleted: , which has a large fraction of open ocean in the grid cell and therefore are biased towards convective conditions

Deleted: is

Deleted: 3

Formatted: Subscript

Formatted: Font: Italic

Formatted: Font: Italic, Subscript

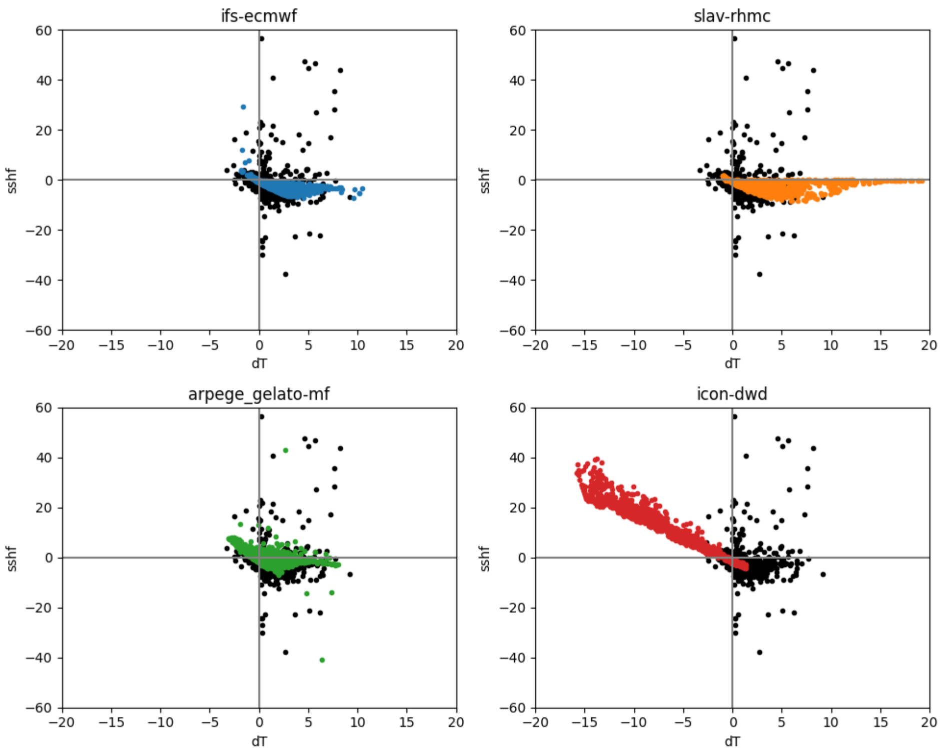


Figure 16: as Figure 15 but for Utqiaġvik. **Note that for the observations ΔT is calculated using the 10m air temperature and an estimate of the surface temperature from an infrared sensor.**

Deleted: 4

Deleted: 3

4. Conclusions and future plans

In this manuscript we have outlined the motivation for YOPPsiteMIP, documented the current status of the YOPPsiteMIP forecast MMDF data archived on the YOPP data portal (hosted by MET Norway), and presented some multi-model forecast evaluation examples to demonstrate the utility of the MMDFs and MODFs using data from the YOPP SOP1, which occurred during February and March 2018. The main conclusions from this analysis are that:

- Near-surface temperature and wind speed forecast errors vary considerably between the different sites, reflecting both a range of climate conditions and forecast performance across the selected sites.
- A common feature of several sites, namely Sodankylä, Barrow, Tiksi, Eureka, is a warm bias during periods of extreme cold which goes hand-in-hand with a lack of temperature variability in the lowest ~100m of the atmosphere.

Deleted: geographies represented by this selection of

Deleted: conditional

Deleted: in hand

- This lack of variability is investigated further at Utqiaġvik, Tiksi and Sodankylä where radiation components were observed and provided in the MODFs and MMDFs, which enabled us to investigate the sensitivity of T2m to radiative forcing:

- At all three sites the models tend to underestimate the sensitivity of T2m and the surface skin temperature (or $LW\uparrow$) to variations in radiative forcing and do not capture extreme minima in these variables, although the AROME-Arctic and CAPS models perform better in this regard.

- At Utqiaġvik and Sodankylä, since turbulent fluxes were provided in addition, we were able to investigate the link between these fluxes and the bulk parameters. This highlighted:
    - Differences in the parameterisation of turbulent fluxes, particularly the specification of the roughness length for momentum which varies by a little less than an order of magnitude between different models.
    - The high importance of the ground heat flux, particularly at the Utqiaġvik and Tiksi sites, where stable conditions dominate. Note that despite this importance, this flux is not observed at these two sites.

Process studies which compare point observations to gridded model output, need to be carried out in awareness of sub-tile representativeness issues. For fine resolution models it is alway recommended to provide output from multiple grid-points (as in this study), centred on the observatory, to be able to pair land-based observations to a model tile with dominant land-cover. For coarse resolution models, we recommend to provide variables for the different sub-tile components (bare soil, vegetation, water, ice, …). The more the site characteristics are matched to the correct model output, the more reliable diagnosis on the model capability to reproduce the observed physical process. In this study we found that the land-ocean contrast in the Arctic in winter does not significantly affect the surface energy budget sensitivity to radiative forcing in the CAPS model (in Section 3.4, the ocean-dominated Utqiagvik grid-points of CAPS do not stand out with respect to the other models), because the frozen ocean has similar characteristics to the snow-covered land surface. On the other hand, the ICON model, which has very low sea ice values (~10%) has much warmer temperatures than the other models at Utqiagvik, and as a result the sensible heat flux behaves differently compared to the other models). Accounting for the land-ocean contrast will be crucial in the sea-ice free summer SOP2 period that will be evaluated in the future.

The development of the MODFs and MMDFs is ongoing and will be completed in phases. The initial phase was to collect basic meteorology data and the main components of the radiation budget. Work on this initial phase is completed and the next phase will provide a wider range of parameters (e.g. turbulent fluxes and cloud parameters) included in the MODFs. This is a more complicated, but very necessary step since the models differ significantly in terms of surface heat and momentum fluxes as well as cloud properties (not shown). There are also plans to extend the MODF and MMDF concept to Antarctica, focussing on the Southern-hemisphere SOPs. These future phases of the YOPPsiteMIP will allow more detailed studies on e.g.:

- cloud cover, microphysics and radiative forcing,
- assessment of forecast models in Antarctica,
- testing of specific model developments,
- observatory representativeness studies.

This will allow a more process-focussed understanding of the forecasts in the YOPPsiteMIP archive, but also provide a testbed for model developers to use when testing new model formulations relevant for the Arctic. Further details on the MODF concept and the SOP1 and 2 MODFs can be found in Uttal et al., (2023) and Mariani et al., (2024) respectively. A Python based toolkit for producing the MODFs is under development, which it is hoped will speed up and simplify the production of MODFs and facilitate timely evaluation of forecast models to inform the model development process.

Appendix A: Table of acronyms

EDMF=Eddy Diffusivity Mass Flux.

FE=Finite Element,

Deleted: the thermal representation of the land surface as an issue with all forecasts, likely due to the single-layer representation of snow used in all the forecasts submitted to YOPPsiteMIP but potentially also due to the thermal representation of forest canopy at Sodankylä.

Deleted: ¶

Deleted: and ICON

Deleted: acts as

Deleted: the wind stress and sensible heat flux are affected by the ocean-vs-land tile characteristics (

Deleted: in Section 3.5, ICON

Deleted: with respect

Deleted: hugely

Deleted: stable boundary layers, diurnal cycles and surface exchange processes,

Deleted: ¶

Deleted: radiative forcing and albedo

Deleted: ¶

Deleted: vertical structure of the lower atmosphere,

Deleted: ¶

Deleted: assessment of cloud

Deleted: hydrometeors

Deleted: Morris

Deleted: 3

Deleted:

Deleted: available on gitlab and details can be found in Gallagher et al. (manuscript in preparation)

FD=Finite Difference,
FV=Finite Volume,
H=Hydrostatic,
HARATU = HARMONIE-AROME with RACMO Turbulence
HTESSEL=Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land,
ICE3 = Three-class ice parameterization
IQR = Inter-Quartile Range
ISBA= Interactions between Surface–Biosphere–Atmosphere,
NH=Non-hydrostatic,
SURFEX = Surface Externalisée,
TERRA = Land Surface module of the ICON weather forecast model.
TKE=Turbulent Kinetic Energy,

Data availability statement

All MMDF and MODFs are available on the YOPP Data Portal (https://yopp.met.no), hosted by the Norwegian Meteorological Institute, for perpetuity (ie. longer than 10 years). The YOPP Data Portal is relying on the Arctic Data Centre (https://adc.met.no) for data stewarding and the YOPPSiteMIP data can be programmatically accessed using the machine interface for the Arctic Data Centre or can be accessed directly from https://thredds.met.no/thredds/catalog/alertness/YOPP_supersite /obs/catalog.html, for the MODFs and https://thredds.met.no/thredds/catalog/YOPPSiteMIP-models/catalog.html, for the MMDFs.

The SOP1 and SOP2 MODFs for each station shown in white in Fig 1 has been assigned a separate DOI, as described in Mariani et al. (2024). In the case of the MMDFs a DOI is assigned to the data for each forecast model:

Deleted: Morris
Deleted: submitted

- ECMWF-IFS: https://doi.org/10.21343/A6KA-7142,
- ARPEGE-MF: https://doi.org/10.21343/T31Z-J391,
- SLAV-RHMC: https://doi.org/10.21343/J4SJ-4N61
- DWD-ICON: https://doi.org/10.21343/09KM-BJ07,
- ECCC-CAPS: https://doi.org/10.21343/2BX6-6027,
- AROME-MF: https://doi.org/10.21343/JZH3-2470,
- AROME-Arctic: https://doi.org/10.21343/47AX-MY36.

Formatted: Font colour: Auto

Code availability statement

Apart from the ECMWF-IFS, for which an open access version of the code is available here: https://confluence.ecmwf.int/display/OIFS, the model codes are not open access.

Formatted: No bullets or numbering

Funding

- JD was supported by European Union's Horizon 2020 Research and Innovation program through Grant Agreement 871120 (INTERACTIII).

- MT was partially supported with Russian Science Foundation, Grant 21-17-00254
- RP was supported by European Union's Horizon 2020 Research and Innovation program through Grant Agreement 101003590 (PolarRES)

- TR was supported by the Norwegian Research Council project no. 280573 ‘Advanced models and weather prediction in the Arctic: enhanced capacity from observations and polar process representations (ALERTNESS)’

Author contributions

The initial YOPPsiteMIP, MODF and MMDF concepts were developed by GS, JD, BC, TU, SK, LMH, AS and EB. JD, BC, EB, NA, HF, TR, RF & MT produced or ran simulations to make MMDFs. TU, EA, MG, LXH, JH, ZM, SM, EO, IS, MG, JT and RP produced or were involved in the production of MODFs. LF, MD and ØG were responsible for the YOPPsiteMIP archive hosted at MET Norway. JD produced the figures and wrote the manuscript with comments and input from all co-authors.

Competing Interests

The authors declare that they have no conflict of interest.

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. We would specifically like to thank Thomas Jung, Jeff Wilson and wider PPP steering group for their tireless support of YOPPsiteMIP.

References

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

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

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

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

Atlaskin, E. and Vihma, T.: Evaluation of NWP results for wintertime nocturnal boundary-layer temperatures over Europe and Finland, Q. J. R. Meteorol. Soc., 138, 1440–1451, https://doi.org/10.1002/qj.1885, 2012.

Baines, P. G., & Palmer, T. N. (1990). Rationale for a new physically-based parametrisation of sub-grid scale orographic effects. Tech. Rep. 169, European Centre for Medium-Range Weather Forecasts, Reading, UK.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., & Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, Monthly Weather Review, 139(12), 3887-3905, 2011

Deleted: ¶

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, J. Hydrometeorol., 10, 623–643, https://doi.org/10.1175/2008JHM1068.1, 2009.

Batrak, Y. and Müller, M.: On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice, Nat. Commun., 10, 1–8, https://doi.org/10.1038/s41467-019-11975-3, 2019.

Bauer, P., Magnusson, L., Thépaut, J.-N., and Hamill, T. M.: Aspects of ECMWF model performance in polar areas, Q. J. R. Meteorol. Soc., 142, 583–596, https://doi.org/10.1002/qj.2449, 2016.

Bazile, E., & Azouz, N. (2023a). Merged model Data Files (MMDFs) for the Meteo France ARPEGE global forecast model for various Polar sites. Norwegian Meteorological Institute. https://doi.org/10.21343/T31Z-J391

Bazile, E., & Azouz, N. (2023b). MMDFs for the Meteo France AROME regional forecast model for various Arctic sites. Norwegian Meteorological Institute. https://doi.org/10.21343/JZH3-2470

E.Bazile, N. Azouz, A. Napoly, C. Loo: Impact of the 1D sea-ice model GELATO in the global model ARPEGE. France 6-03. http://bluebook.meteoinfo.ru/index.php?year=2020&ch_=2

Bazile, E., Marquet, P., Bouteloup, Y, and Bouyssel, F.: The Turbulent Kinetic Energy (TKE) scheme in the NWP models at Météo-France, ECMWF GABLS Workshop on Diurnal cycles and the stable boundary layer, 7-10 November 2011

Bélair, S., Brown, R., Mailhot, J., Bilodeau, B., & Crevier, L. (2003). Operational Implementation of the ISBA Land Surface Scheme in the Canadian Regional Weather Forecast Model. Part II: Cold Season Results, Journal of Hydrometeorology, 4(2), 371-386.

Bélair, S., J. Mailhot, C. Girard, and P. Vaillancourt, 2005: Boundary Layer and Shallow Cumulus Clouds in a Medium-Range Forecast of a Large-Scale Weather System. Mon. Wea. Rev., 133, 1938–1960, https://doi.org/10.1175/MWR2958.1.

Beljaars, A. C. M. and Holtslag, A. a. M.: Flux Parameterization over Land Surfaces for Atmospheric Models, J. Appl. Meteorol., 30, 327–341, https://doi.org/10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2, 1991.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., Rooy, W. de, Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K.-I., Lenderink, G., Niemelä, S., Nielsen, K. P., Onvlee, J., Rontu, L., Samuelsson, P., Muñoz, D. S., Subias, A., Tijm, S., Toll, V., Yang, X., and Køltzow, M. Ø.: The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System, Mon. Weather Rev., 145, 1919–1935, https://doi.org/10.1175/MWR-D-16-0417.1, 2017.

Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo, 2008: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. Quarterly Journal of the Royal Meteorological Society, 134 (634), 1337–1351, https://doi.org/10.1002/qj.289, URL https://onlinelibrary.wiley.com/doi/ 10.1002/qj.289.

Bechtold, P., E. Bazile, F. Guichard, P. Mascart and E. Richard, 2001 : A mass flux convection scheme for regional and global models. Quart. J. Roy. Meteor . Soc., 127, 869-886.

Bougeault, P.: Cloud ensemble relations for use in higher order models of the planetary boundary layer, J. Atmos. Sci, 39, 2691-2700, 1982.

Bougeault,P.(1985). A simple parameterisation of the large scale effects of cumulus convection. Mon. Wea. Rev., 4:469–485.

Bromwich, D. H., Werner, K., Casati, B., Powers, J. G., Gorodetskaya, I. V., Massonnet, F., Vitale, V., Heinrich, V. J., Liggett, D., Arndt, S., Barja, B., Bazile, E., Carpentier, S., Carrasco, J. F., Choi, T., Choi, Y., Colwell, S. R., Cordero, R. R., Gervasi, M., Haiden, T., Hirasawa, N., Inoue, J., Jung, T., Kalesse, H., Kim, S.-J., Lazzara, M. A., Manning, K. W., Norris, K., Park, S.-J., Reid, P., Rigor, I., Rowe, P. M., Schmithüsen, H., Seifert, P., Sun, Q., Uttal, T., Zannoni, M., and Zou, X.: The Year of Polar Prediction in the Southern Hemisphere (YOPP-SH), Bull. Am. Meteorol. Soc., 101, E1653–E1676, https://doi.org/10.1175/BAMS-D-19-0255.1, 2020.

Buizza, R., Bidlot, J.-R., Janousek, M., Keeley, S., Mogensen, K., and Richardson, D.: New IFS cycle brings sea-ice coupling and higher ocean resolution, ECMWF Newsl. - Number 150, 14–17, https://doi.org/10.21957/xbov3ybily, 2017.

Casati, B. (2023). MMDFs for the Environment and Climate Change Canada-CAPS regional forecast model for various Arctic sites. Norwegian Meteorological Institute. https://doi.org/10.21343/2BX6-6027

Casati, B., T. Robinson, F. Lemay, M. Køltzow, T. Haiden, E. Mekis, F. Lespinas, V. Fortin, G. Gascon, J. Milbrandt, G. Smith (2023) Performance of the Canadian Arctic Prediction System during the YOPP Special Observing Periods. Atmosphere-Ocean, https://doi.org/10.1080/07055900.2023.2191831

Casati, B.: Model process-based evaluation using high-frequency multi-variate observations at the Arctic and Antarctic supersites during the Year of Polar Prediction: the YOPPsiteMIP, n.d.

Deleted: ¶

Deleted: ¶

Deleted: Bélair, S., Brown, R., Mailhot, J., Bilodeau, B., & Crevier, L. (2003). Operational Implementation of the ISBA Land Surface Scheme in the Canadian Regional Weather Forecast Model. Part II: Cold Season Results, Journal of Hydrometeorology, 4(2), 371-386.

Deleted: ¶

Deleted: Belair et al., 2005: Boundary Layer and Shallow Cumulus Clouds in a Medium-Range Forecast of a Large-Scale Weather System. Monthly Weather Review, Vol. 133, 1938-1960

Deleted: ¶

Catry, B., Geleyn, J. F., Bouyssel, F., Cedilnik, J., Brožková, R., & Derková, M. (2008). A new sub-grid scale lift formulation in a mountain drag parameterisation scheme. Meteorologische Zeitschrift, 17(2), 193–208. https://doi.org/10.1127/0941-2948/2008/0272

Cheng, Y., V. M. Canuto, and A. M. Howard, 2002: An Improved Model for the Turbulent PBL. J. Atmos. Sci., 59, 1550–1565, https://doi.org/10.1175/1520-0469(2002)059<1550:AIMFTT>2.0.CO;2.

Coté, J.,S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998: The operational CMC–MRD Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. Mon. Wea. Rev., 126, 1373–1395, doi:10.1175/ 1520-0493(1998)126,1373:TOCMGE.2.0.CO;2.

Cuxart, J., Bougeault, P. and Redelsperger, J.-L. (2000), A turbulence scheme allowing for mesoscale and large-eddy simulations. Q.J.R. Meteorol. Soc., 126: 1-30. https://doi.org/10.1002/qj.49712656202

Deleted: Curry, J.: On the Formation of Continental Polar Air, J. Atmospheric Sci., 40, 2278–2292, https://doi.org/10.1175/1520-0469(1983)040<2278:OTFOCP>2.0.CO;2, 1983.

Deleted: ¶

Cuxart, J., Holtslag, A. A. M., Beare, R. J., Bazile, E., Beljaars, A., Cheng, A., Conangla, L., Ek, M., Freedman, F., Hamdi, R., Kerstein, A., Kitagawa, H., Lenderink, G., Lewellen, D., Mailhot, J., Mauritsen, T., Perov, V., Schayes, G., Steeneveld, G.-J., Svensson, G., Taylor, P., Weng, W., Wunsch, S., and Xu, K.-M.: Single-Column Model Intercomparison for a Stably Stratified Atmospheric Boundary Layer, Boundary-Layer Meteorol, 118, 273–303, https://doi.org/10.1007/s10546-005-3780-1, 2006.

Day, J. (2023). MMDFs for the ECMWF-IFS global forecast model for various Polar sites. Norwegian Meteorological Institute. https://doi.org/10.21343/A6KA-7142

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., 145, 3846–3862, https://doi.org/10.1002/qj.3673, 2019.

Day, J. J., Arduini, G., Sandu, I., Magnusson, L., Beljaars, A., Balsamo, G., Rodwell, M., and Richardson, D.: Measuring the Impact of a New Snow Model Using Surface Energy Budget Process Relationships, J. Adv. Model. Earth Syst., 12, https://doi.org/10.1029/2020MS002144, 2020.

Day, J. J., Keeley, S., Arduini, G., Magnusson, L., Mogensen, K., Rodwell, M., Sandu, I., and Tietsche, S.: Benefits and challenges of dynamic sea ice for weather forecasts, Weather Clim. Dyn., 3, 713–731, https://doi.org/10.5194/wcd-3-713-2022, 2022.

Delage, Y. (1997): Parametrizing sub-grid scale vertical transport in atmospheric models under statically stable conditions. Boundary-Layer Meteorology 82, 23–48 (1997). https://doi.org/10.1023/A:1000132524077

Delage, Y., and C. Girard (1992): Stability functions correct at the free convection limit and consistent for for both the surface and Ekman layers. Boundary-Layer Meteorol 58, 19–31. https://doi.org/10.1007/BF00120749

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, Remote Sens. Environ., 116, 140–158, https://doi.org/10.1016/j.rse.2010.10.017, 2012.

Bašták-Ďurán, I. B., J. Geleyn, F. Váňa, J. Schmidli, and R. Brožková, 2018: A Turbulence Scheme with Two Prognostic Turbulence Energies. *J. Atmos. Sci.*, **75**, 3381–3402, https://doi.org/10.1175/JAS-D-18-0026.1.

Dyer, A. J. (1974). A review of flux-profile relationships. Boundary-Layer Meteorol., 7, 363–372.

Emmerson, C., and G. Lahn, 2012: Arctic opening: Opportunity and risk in the high north. Lloyds Rep., 59 pp. [Available online at www.chathamhouse.org /sites/files/chathamhouse/public/Research/Energy,%20 Environment%20and%20Development/0412arctic.pdf.]

Forbes, R. M. and Ahlgrimm, M.: On the Representation of High-Latitude Boundary Layer Mixed-Phase Cloud in the ECMWF Global Model, Mon. Weather Rev., 142, 3425–3445, https://doi.org/10.1175/MWR-D-13-00325.1, 2014.

Frank, H. (2023). MMDFs for the DWD-ICON global forecast model for various Arctic sites. Norwegian Meteorological Institute. https://doi.org/10.21343/09KM-BJ07

Gerard, L. and Geleyn, J.-.-F. (2005), Evolution of a subgrid deep convection parametrization in a limited-area model with increasing resolution. Q.J.R. Meteorol. Soc., 131: 2293-2312. https://doi.org/10.1256/qj.04.72

Gerard, L., J. Piriou, R. Brožková, J. Geleyn, and D. Banciu, 2009: Cloud and Precipitation Parameterization in a Meso-Gamma-Scale Operational Weather Prediction Model. *Mon. Wea. Rev.*, **137**, 3960–3977, https://doi.org/10.1175/2009MWR2750.1.

Girard, C., and Coauthors, 2014: Staggered vertical discretization of the Canadian Environmental Multiscale (GEM) model using a coordinate of the log-hydrostatic-pressure type. Mon. Wea. Rev., 142, 1183–1196, doi:10.1175/MWR-D-13-00255.1.

Goessling, H. F., and Coauthors, 2016: Paving the way for the Year of Polar Prediction. Bull. Amer. Meteor. Soc., 97, ES85–ES88, doi:10.1175/BAMS-D-15-00270.1.

Haiden, T., Sandu, I., Balsamo, G., Arduini, G., and Beljaars, A.: Addressing biases in near-surface forecasts, ECMWF, https://doi.org/10.21957/eng71d53th, 2018.

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

Heise E., Ritter B. & Schrodin R. 2006. Operational implementa- tion of the multilayer soil model. COSMO Technical Reports No. 9. Offenbach am Main, Germany: Consortium for Small-Scale Modelling.

Hogan, R. J. and Bozzo, A.: A Flexible and Efficient Radiation Scheme for the ECMWF Model, J. Adv. Model. Earth Syst., 10, 1990–2008, https://doi.org/10.1029/2018MS001364, 2018.

Högström, U. (1988). Non-dimensional wind and temperature profiles in the atmospheric surface layer: A re-evaluation. Boundary-Layer Meteorol., 42, 55–78.

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

Holtslag, A. A. M., & De Bruin, H. A. R. (1988). Applied Modeling of the Nighttime Surface Energy Balance over Land, Journal of Applied Meteorology and Climatology, 27(6), 689-704.

Holtslag, A. A. M., Svensson, G., Baas, P., Basu, S., Beare, B., Beljaars, A. C. M., Bosveld, F. C., Cuxart, J., Lindvall, J., Steeneveld, G. J., Tjernström, M., & Van De Wiel, B. J. H. (2013). Stable Atmospheric Boundary Layers and Diurnal Cycles: Challenges for Weather and Climate Models, Bulletin of the American Meteorological Society, 94(11), 1691-1706.

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

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

Iacono, M., Delamere, J., Mlawer, E., Shephard, M., Clough, S. and Collins, W. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. J. Geophys. Res., 113D, 13103.

Ingleby, B., Arduini, G., Balsamo, G., Boussetta, S., Ochi, K., Pinnington, E., and De Rosnay, P.: Improved two-metre temperature forecasts in the 2024 upgrade, ECMWF Newsletter - Number 178, 2024.

Jung, T. and Matsueda, M. (2016), Verification of global numerical weather forecasting systems in polar regions using TIGGE data. Q.J.R. Meteorol. Soc., 142: 574-582. https://doi.org/10.1002/qj.2437

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.: Advancing Polar Prediction Capabilities on Daily to Seasonal Time Scales, Bull. Am. Meteorol. Soc., 97, 1631–1647, https://doi.org/10.1175/BAMS-D-14-00246.1, 2016.

Kain, J. S., and J. M. Fritsch, 1990: A One-Dimensional Entraining/Detraining Plume Model and Its Application in Convective Parameterization. J. Atmos. Sci., 47, 2784–2802, https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2.

Deleted: Iacono, M., Delamere, J., Mlawer, E., Shephard, M., Clough, S. and Collins, W. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. J. Geophys. Res., 113D, 13103.

Deleted: ¶

Deleted: ¶

Deleted: Illingworth, A. J., Hogan, R. J., O'Connor, E. j., Bouniol, D., Brooks, M. E., Delanoé, J., Donovan, D. P., Eastment, J. D., Gaussiat, N., Goddard, J. W. F., Haeffelin, M., Baltink, H. K., Krasnov, O. A., Pelon, J., Piriou, J.-M., Protat, A., Russchenberg, H. W. J., Seifert, A., Tompkins, A. M., van Zadelhoff, G.-J., Vinit, F., Willén, U., Wilson, D. R., and Wrench, C. L.: Cloudnet, Bull. Am. Meteorol. Soc., 88, 883–898, https://doi.org/10.1175/BAMS-88-6-883, 2007.

Formatted: Space After: 0 pt, Line spacing: 1.5 lines

Formatted: Space After: 0 pt, Line spacing: 1.5 lines

Kähnert, M., Sodemann, H., Remes, T.M. et al. Spatial Variability of Nocturnal Stability Regimes in an Operational Weather Prediction Model. Boundary-Layer Meteorol 186, 373–397 (2023). https://doi.org/10.1007/s10546-022-00762-1

Karlsson J. and G. Svensson, 2013: Consequences of poor representation of Arctic sea-ice albedo and cloud-radiation interactions in the CMIP5 model ensemble. Geophys. Res. Lett., 40, 4374–4379, doi:10.1002/grl.50768

Køltzow, 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, https://doi.org/10.1175/WAF-D-19-0003.1, 2019.

Köhler, M., Ahlgrimm, M. and Beljaars, A. (2011). Unified treatment of dry convective and stratocumulus topped boundary layers in the ECMWF model. Q. J. R. Meteorol. Soc., 137, 43–57.

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. R. Meteorol. Soc., 145, 3432–3454, https://doi.org/10.1002/qj.3628, 2019.

Lenderink, G. and Holtslag, A.A.M. (2004), An updated length-scale formulation for turbulent mixing in clear and cloudy boundary layers. Q.J.R. Meteorol. Soc., 130: 3405-3427. https://doi.org/10.1256/qj.03.117

Li, J. and H.W Barker, 2005: A radiation algorithm with correlated-k distribution. Part I: Local thermal equilibrium. J. Atmos. Sci., 62, 286-309.

Lott, F. and Miller, M. J. (1997). A new subgrid-scale orographic drag parametrization: Its formulation and testing. Q. J. R. Meteorol. Soc., 123, 101–127.

Louis, J. F.: A parametric model of vertical eddy fluxes in the atmosphere | SpringerLink, Bound.-Layer Meteorol., 1979.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev., 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, 2013.

Milbrandt, J. A. and Morrison, H.: Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part III: Introduction of Multiple Free Categories, J. Atmospheric Sci., 73, 975–995, https://doi.org/10.1175/JAS-D-15-0204.1, 2016.

Milbrandt, J. A., Bélair, S., Faucher, M., Vallée, M., Carrera, M. L., and Glazer, A.: The Pan-Canadian High Resolution (2.5 km) Deterministic Prediction System, Weather Forecast., 31, 1791–1816, https://doi.org/10.1175/WAF-D-16-0035.1, 2016.

Miller, N. B., Shupe, M. D., Lenaerts, J. T. M., Kay, J. E., de Boer, G., & Bennartz, R. (2018). Process-based model evaluation using surface energy budget observations in central Greenland. Journal of Geophysical Research: Atmospheres, 123, 4777–4796. https://doi.org/10.1029/2017JD027377

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J. and Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res., 102D, 16663–16682.

Mariani, Z., Morris, S., Uttal, T., Akish, E., Crawford, R., Huang, L., Day, J., Tjernström, J., Godøy, Ø., Ferrighi, L., Hartten, L., Holt, J., Cox, C., O'Connor, E., Pirazzini, R., Maturilli, M., Prakash, G., Mather, J., Strong, K., Fogal, P., Kustov, V., Svensson, G., Gallagher, M., and Vasel, B.: Special Observing Period (SOP) Data for the Year of Polar Prediction site Model Intercomparison Project (YOPPsiteMIP), Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2023-497, in review, 2024.

Morrison, H. and Milbrandt, J. A.: Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part I: Scheme Description and Idealized Tests, J. Atmospheric Sci., 72, 287–311, https://doi.org/10.1175/JAS-D-14-0065.1, 2015.

Deleted: Miller, N. B., Shupe, M. D., Cox, C. J., Noone, D., Persson, P. O. G., and Steffen, K.: Surface energy budget responses to radiative forcing at Summit, Greenland, The Cryosphere, 11, 497–516, https://doi.org/10.5194/tc-11-497-2017, 2017.

Deleted: ¶

Deleted: Morris et al. (submitted to ESSD): Special Observing Period (SOP) Data for the Year of Polar Prediction site Model Intercomparison Project (YOPPsiteMIP)

Morrison, H., Milbrandt, J. A., Bryan, G. H., Ikeda, K., Tessendorf, S. A., and Thompson, G.: Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part II: Case Study Comparisons with Observations and Other Schemes, J. Atmospheric Sci., 72, 312–339, https://doi.org/10.1175/JAS-D-14-0066.1, 2015.

Müller, M., and Coauthors, 2017: AROME-MetCoOp: A Nordic Convective-Scale Operational Weather Prediction Model. Wea. Forecasting, 32, 609–627, https://doi.org/10.1175/WAF-D-16-0099.1.

Noilhan, J., & Planton, S. (1989). A Simple Parameterization of Land Surface Processes for Meteorological Models, Monthly Weather Review, 117(3), 536-549.

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

Pailleux, J., Geleyn, J.-F., Hamrud, M., Courtier, P., Thépaut, J.-N., Rabier, F., Andersson, E., Burridge, D., Simmons, A., Salmond, D., Khatib, E., and Fischer, C.: Twenty-five years of IFS/ARPEGE, https://doi.org/10.21957/FTU6MFVY, 2014.

Pergaud, J., Masson, V., Malardel, S. et al. A Parameterization of Dry Thermals and Shallow Cumuli for Mesoscale Numerical Weather Prediction. Boundary-Layer Meteorol 132, 83–106 (2009). https://doi.org/10.1007/s10546-009-9388-0

Pinty, J.-P. and P. Jabouille (1998), A mixed-phase cloud parameterization for use in a mesoscale non-hydrostatic model: Simulations of a squall line and of orographic precipitation, Proc. Conf. on Cloud Physics, 217–220.

Pithan, F., Medeiros, B., and Mauritsen, T.: Mixed-phase clouds cause climate model biases in Arctic wintertime temperature inversions, Clim. Dyn., 43, 289–303, https://doi.org/10.1007/s00382-013-1964-9, 2014.

Pithan, F., Ackerman, A., Angevine, W. M., Hartung, K., Ickes, L., Kelley, M., Medeiros, B., Sandu, I., Steeneveld, G.-J., Sterk, H. a. M., Svensson, G., Vaillancourt, P. A., and Zadra, A.: Select strengths and biases of models in representing the Arctic winter boundary layer over sea ice: the Larcform 1 single column model intercomparison, J. Adv. Model. Earth Syst., 8, 1345–1357, https://doi.org/10.1002/2016MS000630, 2016.

Prill, F, Reinert, D., Rieger, D., Zängl, G.: ICON Tutorial - Working with the ICON model, https://doi.org/10.5676/dwd_pub/nwv/icon_tutorial2020, 2020

Raschendorfer, M., 2001: The new turbulence parameterization of LM. COSMO Newsl., 1, 89–97.

Remes, T. (2023). MMDFs for the MetNorway AROME regional forecast model for various Arctic sites. Norwegian Meteorological Institute. https://doi.org/10.21343/47AX-MY36

Rodwell, M. J. and Palmer, T. N.: Using numerical weather prediction to assess climate models, Q. J. R. Meteorol. Soc., 133, 129–146, https://doi.org/10.1002/qj.23, 2007.

Sandu, I., Beljars, A., Bechtold, P., Mauritsen, T., and Balsamo, G.: Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models?, J. Adv. Model. Earth Syst., 5, 117–133, https://doi.org/10.1002/jame.20013, 2013.

Sedlar, J., Tjernström, M., Rinke, A., Orr, A., Cassano, J., Fettweis, X., et al. (2020). Confronting Arctic troposphere, clouds, and surface energy budget representations in regional climate models with observations. Journal of Geophysical Research: Atmospheres, 125. https://doi.org/10.1029/2019JD031783

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France Convective-Scale Operational Model, Mon. Weather Rev., 139, 976–991, https://doi.org/10.1175/2010MWR3425.1, 2011.

Seifert, A.: A revised cloud microphysical parameterization for COSMO-LME. COSMO News Letter No. 7, http://www.cosmo-model.org, 2008

Seity, Y., Lac, C., Bouyssel, F., Riette, S., & Bouteloup, Y. (2012, November). Cloud and microphysical schemes in ARPEGE and AROME models. In Proceedings of the Workshop on Parametrization of Clouds and Precipitation (ECMWF), Reading, UK (pp. 5-8).

Deleted: Roehrig, R., Beau, I., Saint-Martin, D., Alias, A., Decharme, B., Guérémy, J.-F., Voldoire, A., Abdel-Lathif, A. Y., Bazile, E., Belamari, S., Blein, S., Bouniol, D., Bouteloup, Y., Cattiaux, J., Chauvin, F., Chevallier, M., Colin, J., Douville, H., Marquet, P., Michou, M., Nabat, P., Oudar, T., Peyrillé, P., Piriou, J.-M., Salas y Mélia, D., Séférian, R., and Sénési, S.: The CNRM Global Atmosphere Model ARPEGE-Climat 6.3: Description and Evaluation, J. Adv. Model. Earth Syst., 12, e2020MS002075, https://doi.org/10.1029/2020MS002075, 2020.

Deleted: ¶

Solomon, A., Shupe, M.D., Svensson, G., Barton, N.P., Batrak, Y., Bazile, E., Day, J.J., Doyle, J.D., Frank, H.P., Keeley, S., Remes, T., and Tolstykh, M., 2023. The Winter Central Arctic Surface Energy Budget; A Model Evaluation using Observations from the MOSAiC Campaign. Elem Sci Anth, in press.

Soares, P., P. Miranda, A. Siebesma, and J. Teixeira, 2004: An eddy-diffusivity/mass-flux parametrization for dry and shallow cumulus convection. Quart. J. Roy. Meteor. Soc., 130, 33653383, doi:10.1256/qj.03.223.

Siebesma, A. P., P. M. Soares, and J. Teixeira, 2007: A combined eddy diffusivity mass-flux approach for the convective boundary layer. J. Atmos. Sci., 64, 1230–1248, doi:10.1175/JAS3888.1.

Svensson, G. and Karlsson, J.: On the Arctic Wintertime Climate in Global Climate Models, J. Clim., 24, 5757–5771, https://doi.org/10.1175/2011JCLI4012.1, 2011.

Tarasova, T. A., and B. A. Fomin, 2007: The Use of New Parameterizations for Gaseous Absorption in the CLIRAD-SW Solar Radiation Code for Models. J. Atmos. Oceanic Technol., 24, 1157–1162, https://doi.org/10.1175/JTECH2023.1.

Tiedtke, M., 1993: Representation of Clouds in Large-Scale Models. Monthly Weather Review, 121 (11), 3040–3061,

Tjernström, M., Žagar, M., Svensson, G., Cassano, J. J., Pfeifer, S., Rinke, A., Wyser, K., Dethloff, K., Jones, C., Semmler, T., and Shaw, M.: 'Modelling the Arctic Boundary Layer: An Evaluation of Six Arcmip Regional-Scale Models using Data from the Sheba Project,' Bound.-Layer Meteorol., 117, 337–381, https://doi.org/10.1007/s10546-004-7954-z, 2005.

Tjernström, M., Svensson, G., Magnusson, L., Brooks, I. M., Prytherch, J., Vüllers, J., and Young, G.: Central Arctic weather forecasting: Confronting the ECMWF IFS with observations from the Arctic Ocean 2018 expedition, Q. J. R. Meteorol. Soc., 147, 1278–1299, https://doi.org/10.1002/qj.3971, 2021.

Tolstykh, M. (2023). MMDFs for the Roshydromet-SLAV global forecast model for various Arctic sites. Norwegian Meteorological Institute. https://doi.org/10.21343/J4SJ-4N61

Tolstykh, M., Shashkin, V., Fadeev, R., and Goyman, G.: Vorticity-divergence semi-Lagrangian global atmospheric model SL-AV20: dynamical core, Geosci. Model Dev., 10, 1961–1983, https://doi.org/10.5194/gmd-10-1961-2017, 2017.

Tolstykh, M. A., Fadeev, R. Yu., Shashkin, V. V., Goyman, G. S., Zaripov, R. B., Kiktev, D. B., Makhnorylova, S. V., Mizyak, V. G., and Rogutov, V. S.: Multiscale Global Atmosphere Model SLAV: the Results of Medium-range Weather Forecasts, Russ. Meteorol. Hydrol., 43, 773–779, https://doi.org/10.3103/S1068373918110080, 2018.

Uttal, T., Starkweather, S., Drummond, J. R., Vihma, T., Makshtas, A. P., Darby, L. S., Burkhart, J. F., Cox, C. J., Schmeisser, L. N., Haiden, T., Maturilli, M., Shupe, M. D., De Boer, G., Saha, A., Grachev, A. A., Crepinsek, S. M., Bruhwiler, L., Goodison, B., McArthur, B., Walden, V. P., Dlugokencky, E. J., Persson, P. O. G., Lesins, G., Laurila, T., Ogren, J. A., Stone, R., Long, C. N., Sharma, S., Massling, A., Turner, D. D., Stanitski, D. M., Asmi, E., Aurela, M., Skov, H., Eleftheriadis, K., Virkkula, A., Platt, A., Førland, E. J., Iijima, Y., Nielsen, I. E., Bergin, M. H., Candlish, L., Zimov, N. S., Zimov, S. A., O'Neill, N. T., Fogal, P. F., Kivi, R., Konopleva-Akish, E. A., Verlinde, J., Kustov, V. Y., Vasel, B., Ivakhov, V. M., Viisanen, Y., and Intrieri, J. M.: International Arctic Systems for Observing the Atmosphere: An International Polar Year Legacy Consortium, Bull. Am. Meteorol. Soc., 97, 1033–1056, https://doi.org/10.1175/BAMS-D-14-00145.1, 2015.

Uttal, T., Hartten, L. M., Khalsa, S. J., Casati, B., Svensson, G., Day, J., Holt, J., Akish, E., Morris, S., O'Connor, E., Pirazzini, R., Huang, L. X., Crawford, R., Mariani, Z., Godøy, Ø., Tjernström, J. A. K., Prakash, G., Hickmon, N., Maturilli, M., and Cox, C. J.: Merged Observatory Data Files (MODFs): An Integrated Observational Data Product Supporting Process-Oriented Investigations and Diagnostics, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-2413, 2023.

Van de Wiel, B. J. H., Vignon, E., Baas, P., van Hooijdonk, I. G. S., van der Linden, S. J. A., Antoon van Hooft, J., Bosveld, F. C., de Roode, S. R., Moene, A. F., & Genthon, C. (2017). Regime Transitions in Near-Surface Temperature Inversions: A Conceptual Model, Journal of the Atmospheric Sciences, 74(4), 1057-1073. doi: https://doi.org/10.1175/JAS-D-16-0180.1

van Meijgaard, E., L. van Ulft, G. Lenderink, S. De Roode, E. L. Wipfler, R. Boers, and R. van Timmermans, 2012: Refinement and application of a regional atmospheric model for climate scenario calculations of Western Europe. KVR Research Rep. 054/12, 44 pp. [Available online at http://library.wur.nl/ WebQuery/wurpubs/fulltext/312258.]

Wallace, J. M., Tibaldi, S., & Simmons, A. J. (1983). Reduction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography. Quarterly Journal of the Royal Meteorological Society, 109(462), 683–717. https://doi.org/10.1002/qj.49710946202

Deleted: Seifert, A.: A revised cloud microphysical parameterization for COSMO-LME. COSMO News Letter No. 7, http://www.cosmo-model.org, 2008

Deleted: ¶

Deleted: Shupe, M.D., Rex, M., Blomquist, B., Persson, P.O.G., Schmale, J., Uttal, T., Althausen, D., Angot, H., Archer, S., Bariteau, L. and Beck, I., 2022. Overview of the MOSAiC expedition: Atmosphere. Elem Sci Anth, 10(1), p.00060.

Deleted: ¶

Formatted: Space After: 0 pt, Line spacing: 1.5 lines

Deleted: Uttal, T. L.M. Hartten, S.J. Khalsa, B. Casati, G. Svensson, J. Day, M. Gallagher, J. Holt, E. Aksh, S. Morris, E. O'Connor, R. Pirazzini, L. Huang, R. Crawford, Z. Mariani, O. Godoy, J. A.K. Tjerrnstrom, G. Prakesh, N. Hickmon, M. Maturilli, and C. Cox, Merged Observatory Data Files (MODFs): An Integrated Research Data Product Supporting Process Oriented Investigations and Diagnostics. 2023.

Formatted: Line spacing: Multiple 1.15 li

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, Q. J. R. Meteorol. Soc., 141, 563–579, https://doi.org/10.1002/qj.2378, 2015.

Deleted: Wexler, H.: Cooling in the lower atmosphere and the structure of polar continental air, Mon. Weather Rev., 64, 122–136, https://doi.org/10.1175/1520-0493(1936)64<122:CITLAA>2.0.CO;2, 1936.

Deleted: ¶