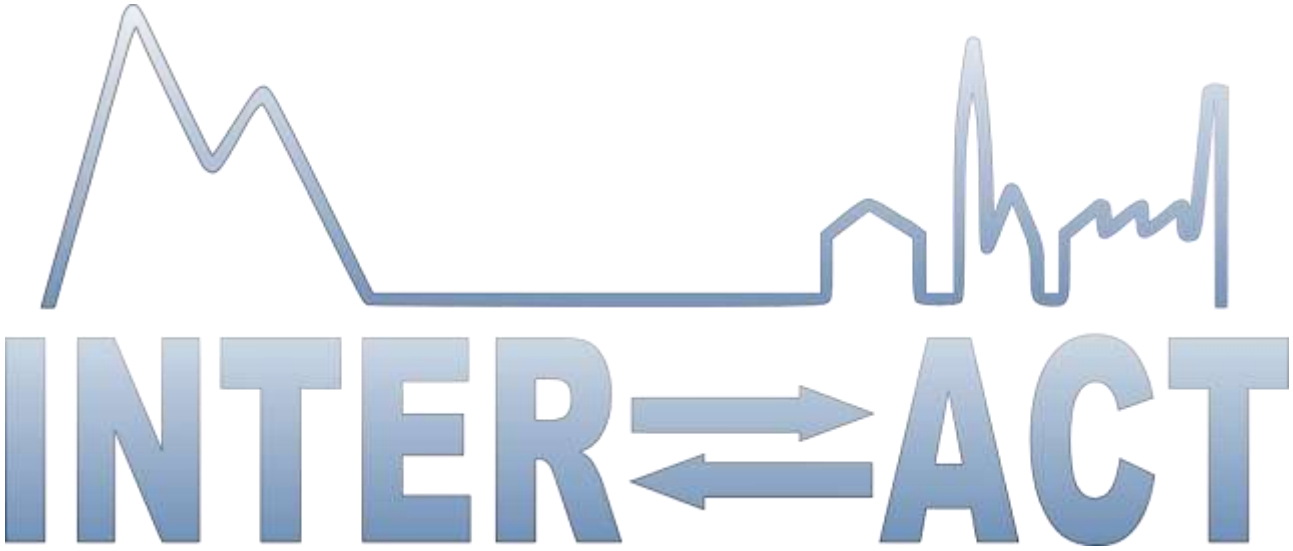
Integrating Activities for Advanced Communities



D4.4 – Report on the use of INTERACT station data to understand systematic forecast errors

Project No.871120– INTERACT

H2020-INFRAIA-2019-1

Start date of project: 2020/01/01 Duration: 48 months
Due date of deliverable: 2022/05/31 Actual Submission date: 2022/05/30

Lead partner for deliverable: ECMWF
Author: Jonathan Day

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the Consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

Table of Contents

Publishable Executive Summary

Although the quality of forecasts in the Arctic is improving it still lags behind the quality of forecasts in the mid-latitudes. The aim of Task 4.4 was to demonstrate the utility of data collected from INTERACT stations for evaluating and improving weather forecasts by using it to evaluate forecasts at a number of pilot sites.

To facilitate such comparisons between observatory data and forecast models the concept of Merged Observatory Data Files (MODFs) and Merged Model Data Files (MMDFs) was recently developed within the framework of the World Meteorological Organisation's Polar Prediction Project. In this report we briefly describe this data concept and provide some examples of the types of analysis one can do with this dataset to inform the model development process. This analysis is performed on a set of MMDFs for seven forecast models against a set of prototype MODFs. These were produced as part of the Year of Polar Prediction site Model Intercomparison Project (YOPPsiteMIP).

In the initial proposal it was proposed that the pilot sites to be used in this report would be Sodankylä (Finland), Utqiagvic (formerly known as Barrow, USA), Summit (Greenland) and Tiksi (Russia). Unfortunately, the MODF files for Summit are not yet available, so in this report we present the same diagnostics for another INTERACT station from Ny-Ålesund, Svalbard.

1. Introduction

Although the quality of weather forecasts and climate projections in the polar regions is improving, it still lags behind the quality of forecasts in lower latitudes (e.g. Jung et al., 2016; Bauer et al., 2016). In the case of weather forecasting this is partly the result of a relatively sparse conventional observing network (Lawrence et al., 2019), but polar regions also pose specific challenges for weather and climate models related to the relatively high frequency with which processes that are historically difficult to model occur. For example, situations which are prone to high forecast errors such as stable boundary layers (Atlaskin and Vihma, 2012; Sandu et al., 2013), the occurrence of mixed-phase clouds (Pithan et al., 2014, 2016), and atmosphere-snow-ice coupling (Day et al., 2020) occur more frequently and are therefore of higher importance in polar regions than in mid-latitudes.

Although process-focussed multi-model evaluations have been performed for global climate models (Pithan et al., 2014; Svensson and Karlsson, 2011) historically there has been relatively little effort to evaluate processes in weather models using in-situ datasets from the terrestrial Arctic and Antarctic, compared to the situation in mid-latitudes. Studies conducted as part of the WMO's Polar Prediction Project have started addressed this gap, focussing on both the large scales (Bauer et al., 2016) and evaluation of standard surface parameters (Køltzow et al., 2019) focussing on conventional verification. The Year of Polar Prediction site Model Intercomparison Project (YOPPsiteMIP) was designed to tackle this problem from a different angle: using observatory data to diagnose the causes of forecast error from a process-oriented perspective, to find common issues, and provide upstream guidance to model developers into how to improve the representation of polar processes in models. Because they are initialised from an actual point in time weather forecasts can be more easily compared with observed timeseries than climate models. Further, since many climate models are based on global weather forecasting systems, understanding the causes of forecast error after 1-2 days may help to inform the causes of error in climate models (Rodwell and Palmer, 2007).

To contribute to YOPPsiteMIP, forecasting centres and polar supersites were asked to provide data in a common file format, with consistent variable naming, units, sign conventions and meta data in order to facilitate fair comparisons between models and observations and interoperability between different sites to facilitate process-oriented evaluation. To achieve this, the concepts of Merged Observatory Data Files (MODFs) and Merged Model Data Files (MMDFs) were developed within the framework of the Polar Prediction Project (see Section 2 and dedicated paper on the MODF and MMDF concept by Uttal et al., in prep). The YOPPsiteMIP and MODF/MMDF concepts and the flow of information are incorporated into the schematic shown in Figure 1.

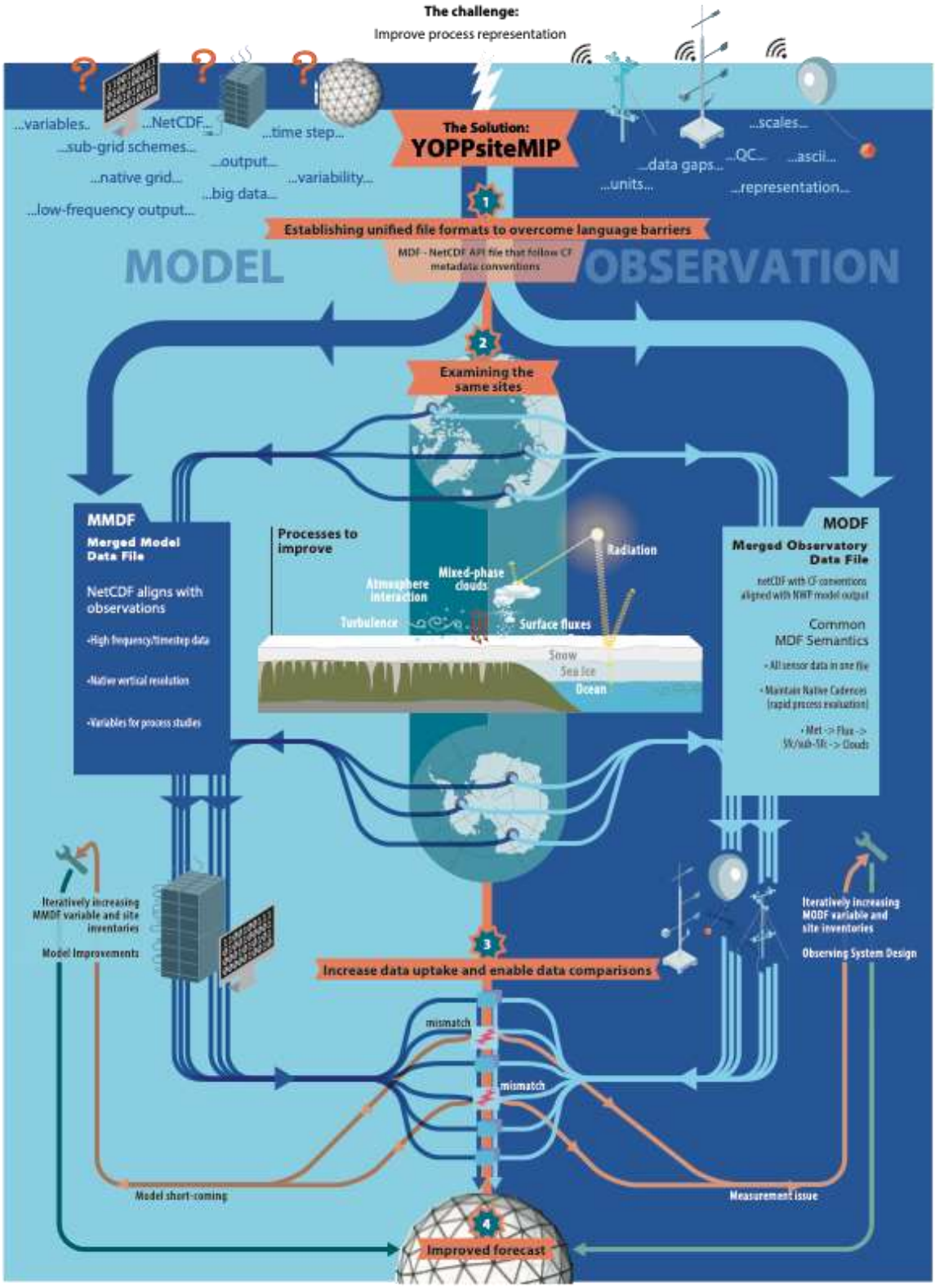


Figure 1. Schematic of the YOPPsiteMIP activity.

The supersites are locations where a wide range of measurements are collected through the full atmosphere-surface-snow-soil column. Measured quantities can include (but are not limited to) solar radiation, aerosol physical and chemical properties, air chemistry, trace gases, cloud properties, water vapor, ozone, temperatures, winds, surface radiation and turbulent fluxes (Uttal et al., 2015). The observations at these sites extend far beyond the traditional synoptic surface and upper-air observations traditionally used in the verification of Numerical Weather Prediction (NWP) systems, and offer opportunities to deepen our understanding of the physical processes governing the weather and climate in the polar environment and their representation in forecast models. The processes targeted by the YOPPsiteMIP include the surface energy budget, turbulence under stably stratified conditions, mixed-phase clouds and coupling between the atmosphere and underlying snow-covered land and/or ice.

The observatories where MODFs and MMDFs were requested are shown in Figure 2. Files were initially requested for the YOPP Special Observing Periods (SOP1: Feb–Mar 2018, and SOP2: Jul–Sep 2018, in the Northern Hemisphere). During these SOPs the observations at many stations were enhanced. For example, the frequency of radiosondes was doubled at many of the Arctic sites (Lawrence et al., 2019; Bromwich et al., 2020).

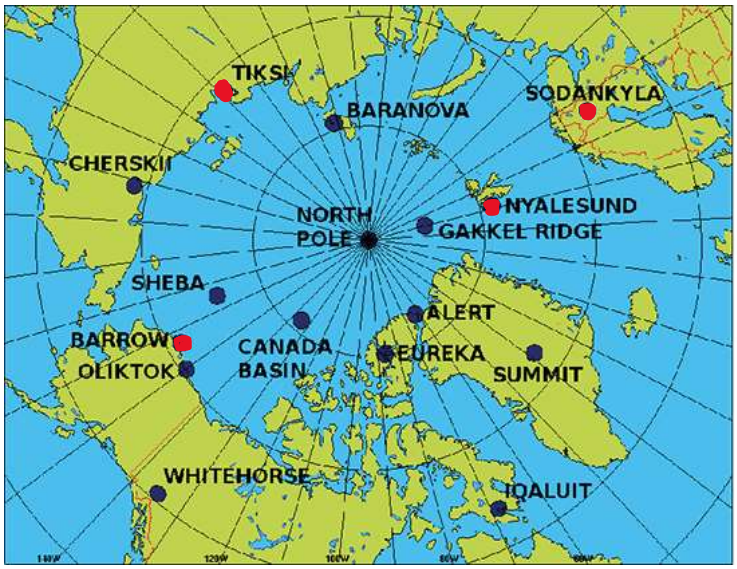


Figure 2. Map of Arctic observatories used in YOPPsiteMIP. The red dots, show the sites that will be discussed in this specific report.

In this report the YOPPsiteMIP dataset will be described, in its current form, and provide a basic description of the forecast files that have been archived (to date) at the YOPP portal, hosted by Met Norway, and provide a summary of the forecasting systems that produced them. Secondly, the paper performs some multi-model and multi-site evaluation of the forecasts (during the winter SOP (1)) at four observatories that have already produced prototype MODF files to give an example of the types of analysis that can be done with the dataset.

These sites are:

1. Sodankylä (Finland): produced by the Finnish Meteorological Institute (FMI)
2. Utqiagvic (formerly known as Barrow, USA): produced by the National Oceanographic and Atmospheric Administration (NOAA)
3. Ny-Ålesund (Svalbard, Norway): produced by the University of Stockholm (with input from Met Norway)
4. Tiksi (Russia): produced by NOAA

The examples will focus on the representation of mixed phase clouds, stable boundary layers and the representation of turbulent surface fluxes. Details of the MODF/MMDF concept and specifics of the data processing chain related to the production of the MODFs will be published separately.

2. Model and dataset descriptions

So far six NWP centres have submitted forecasts with seven forecasting systems for SOP1 & 2. Four of the systems are global: ECMWF-IFS, Meteo-France-ARPEGE, Roshydromet-SLAV and DWD-ICON. Three are regional: ECCC-CAPS with a pan-Arctic domain and the two versions of the AROME system with the same European-Arctic domain, one from Meteo France and the other from Met Norway. The position of the lateral boundaries of the regional systems can be seen in Figure 3.

The forecasts were initialised at both 0 and 12UTC for each day of the SOP. The SOPs covered by each model are indicated in table 2. The forecast length varies somewhat between each forecasting system but all forecasts are at least two days long (see Fig 4 & 5). Similarly, the cadence of the files varies between the frequency of the model timestep (~5-15mins) to hourly aggregated values.

A number of systems (CAPS, SL-AV, ARPEGE, AROME-MF) have also provided multiple grid-points centred around the station location to enable representativeness of the gridpoint to be studied. The others provide a single grid-point near to the location of the observatory. The grid resolution ranges from 2.5km to ~30km

For each forecast cycle, all variables (atmosphere, surface and land-surface) are combined in one netCDF file following the MMDF format which use the same nomenclature, metadata, and structure as the MODFs produced for the observatories. The actual variables archived vary from model to model, but a full list of what was requested can be found in Svensson (2020) The guidelines for producing the MMDFs (and MODFs) are described in the so-called H-K table (Hartten and Khalsa, 2022). They are available in both human-readable (PDF) and machine-readable (JSON) form. The H-K Table relies on standards and conventions commonly used in the earth sciences, including netCDF encoding with CF Conventions. The prescribed metadata make data provenance clear and encourage proper attribution of model (and observational) data.

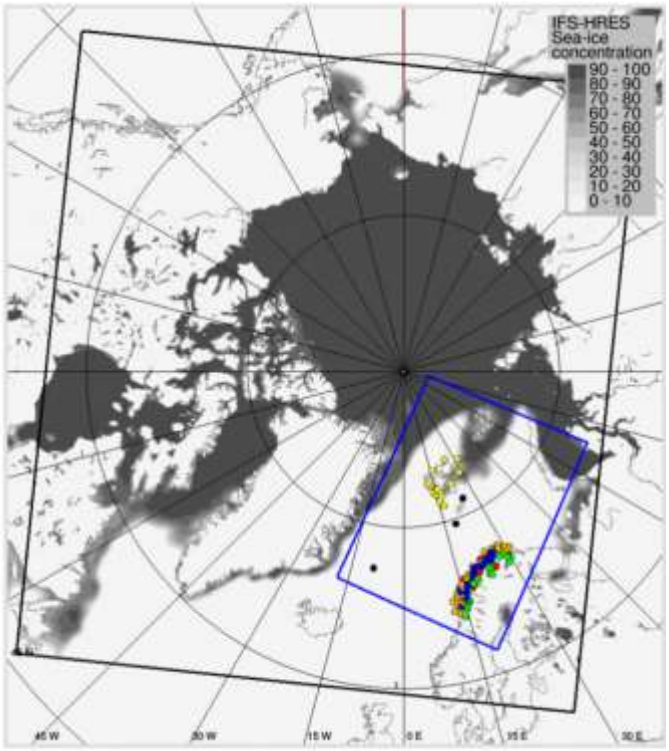


Figure 2. Map showing the outline of CAPS and AROME-Arctic regional domains (copied from Køltzow et al., 2019).

Table 1. Summary of forecasting systems

Centre	Model-name	Global/Regional Atmosphere resolution Version	References/Publications	SOPs in YOPP portal
ECMWF	IFS	Global: TCo1279 (~9km), 137L Version 43r3 (SOP1) and 45r1 (SOP2)	Buizza et al., (2017) (Buizza et al., 2017) Also documentation	SOP1, SOP2 & SOP-SH
MeteoFrance	ARPEGE	Global: T1198L105 (~10Km in the Arctic)	(Roehrig et al., 2020)	SOP1, SOP2 & SOP-SH

MeteoFrance	AROME-Arctic	Regional: 2.5km/ LBCs and ICs from ECMWF	(Seity et al., 2011)	SOP1 & SOP2
ECCC	CAPS	Regional: 3km Initial and LB conditions from the by the Canadian Global Deterministic Prediction System (GDPS, Buehner et al 2015)	(Milbrandt et al., 2016)	SOP1 & SOP2
DWD	ICON	Global r3b07 (~13km); 90 levels	(Zängl et al., 2015)	SOP1 & SOP2
HMCR	SL-AV (global, atmosphere only)	Global: SLAV20, 0.225x(0.16-0.24) lon-lat, 51 levels	(Tolstykh et al., 2018)	SOP1 & SOP2
MetNo	AROME-Arctic	HARMONIE-AROME cy40h 2.5-km, Regional, LBCs and ICs from ECMWF	(Müller et al. 2017) (Bengtsson et al., 2017)	SOP1 & SOP2

3. Results

3.1. Standard evaluation/scores

As mentioned in the introduction, supersites allow detailed process-oriented diagnostics to be performed for the models. However, it is first important to assess what the errors are in a given location, for standard variables such as 10m wind speed and 2m temperature, and how they change with the lead-time of the forecast. This first step is important since it can inform how typical errors at the supersite are of the wider region (when compared to surrounding SYNOP stations) and inform as to how stationary they are with leadtime. If they are quite stationary 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 mean bias and standard deviation of the forecast error are plotted in Figs 4 & 5. The 2m temperature errors have quite different properties at each site and for each model. The models are typically too warm at Barrow and Tiksi and too cold at Ny-Ålesund. At Sodankylä there is a distinct diurnal cycle to the bias, meaning that the night-time temperatures are typically too warm and the bias is smaller and, for some models too cold, during the day.

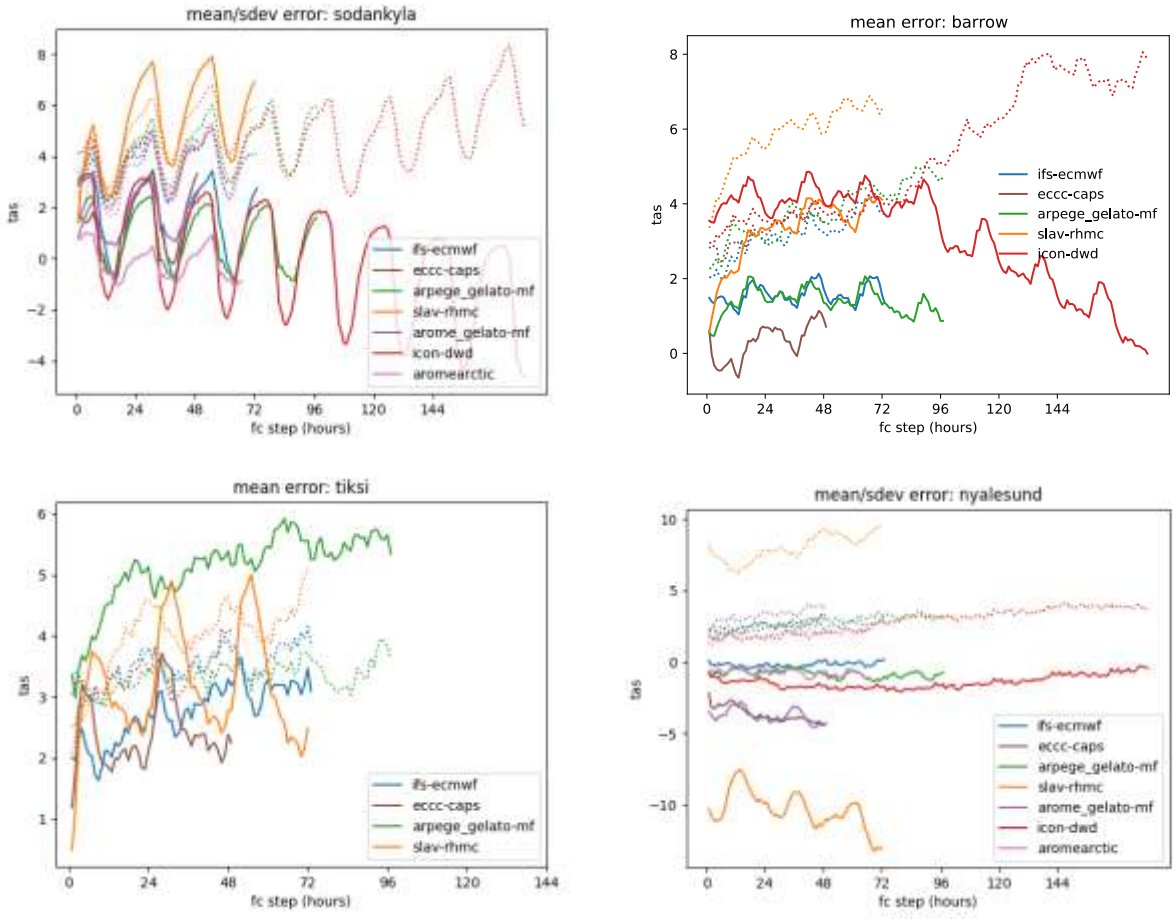


Figure 4. Mean bias and standard deviation of the 2m temperature error at various sites for 0z forecasts initialised during SOP1.

In terms of wind speed the forecasts winds are too fast at Tiksi and Barrow and too slow or too fast depending on the model at Sodankylä and Ny-Ålesund.

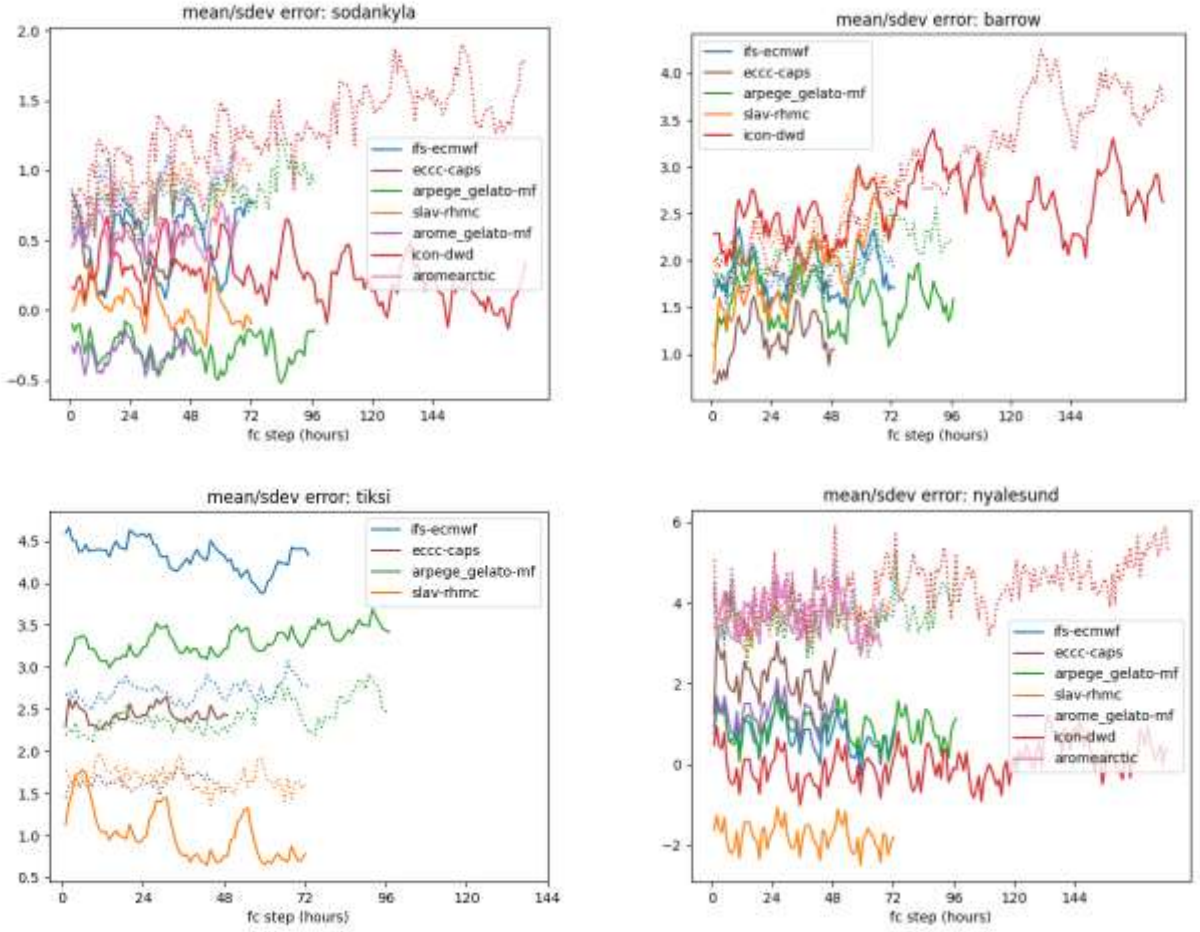


Figure 5. Mean bias and standard deviation of the 10m wind speed error at various sites for 0z forecasts initialised during SOP1.

Although there is some sub-daily variability with a diurnal frequency in the mean bias, the size of the biases does not grow dramatically with time. Two exceptions to this are the SL-AV system and ARPEGE system (for the Tiksi site only). This means that we can consider a 24hr time range between the T+25 and T+48 forecast steps (i.e. the second day of the forecast), simplifying the analysis.

3.2. Motivating case studies and diagnostics for the period

3.2.1. Vertical representation of mixed phase clouds (Ny-Ålesund)

In the Arctic during winter the radiation budget, and as a result the atmospheric boundary layer, is dominated by LW heating and cooling. This is in contrast to the lower latitudes, where solar heating plays a much larger role. LW radiation is emitted by the surface as a function of temperature and emissivity. Part of this returns to space, the remainder is absorbed by the atmosphere and emitted upward and downward. Variations in the downwelling LW are affected by atmospheric temperature and water vapour, but also on the presence and water phase of clouds (Serreze and Barry, 2009). As a result it is important to correctly represent air mass and cloud properties and their interaction

with LW radiation, such as cloud phase partitioning, cloud base height and frequency of cloud occurrence (e.g. Miller et al., 2018; Shupe and Intrieri, 2004).

To illustrate this, observed and modelled profiles of cloud liquid water and downwelling longwave radiation at Ny-Ålesund are shown. The beginning of the period is warm and cloudy, since Svalbard sits in a parcel of warm air that has been advected from the extra-tropics (see Fig 5 of Day et al., 2019). As this airmass passes it is replaced by a dry, cold and clear polar continental air-mass from the north. Between the cloudy and the clear period, the downwelling radiation drops from around 300W/m^2 to 200W/m^2.

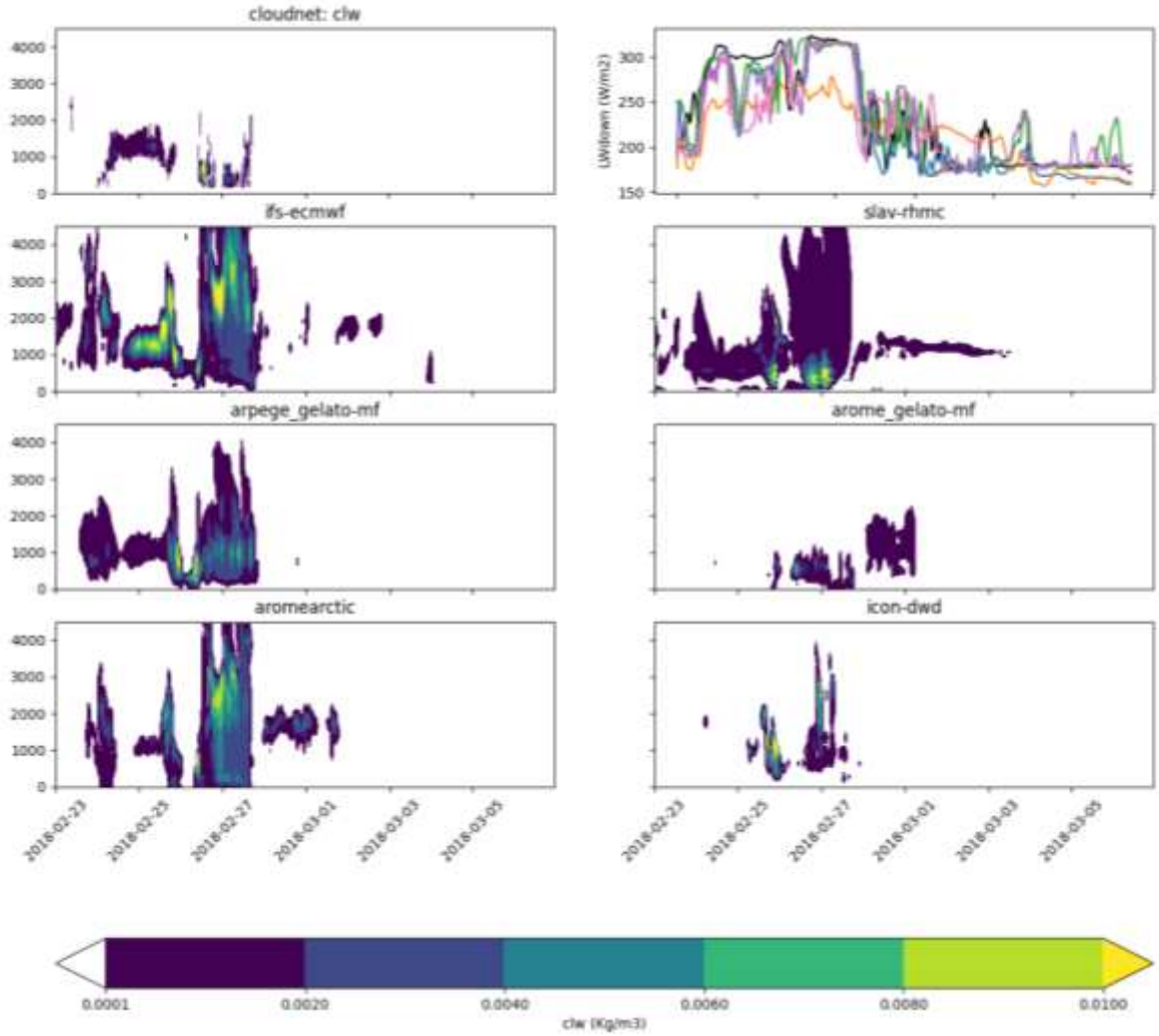


Figure 6. Observed (top left) and forecast (bottom three rows) profiles of cloud-liquid-water content at Ny-Ålesund, Svalbard, for the period 24th Feb to 6th March 2018. The downwelling longwave radiation timeseries for observations (black) and models (colours as in Fig 5) are shown in the top left panel. The model timeseries shown are produced by concatenating hourly output from day-2 of the forecast (T+24-T+47).

Most of the models do well in capturing this dramatic shift in the downwelling radiation. They do this despite having very different cloud water phase partitioning and very different cloud base heights when compared to cloud liquid and ice derived from ground-based remote sensing (downloaded from the Cloudnet project (Illingworth et al., 2007)). For example, clouds in the ARPEGE and AROME model contain much more ice than liquid water in contrast to e.g. the IFS. The models tend to underestimate the downwelling radiation during the period where there is shallow stratus cloud around the 25th Feb. This underestimation in the radiation in the early part of the forecast likely contributes to the cold 2m-temperature bias present in all the models (see Fig 8).

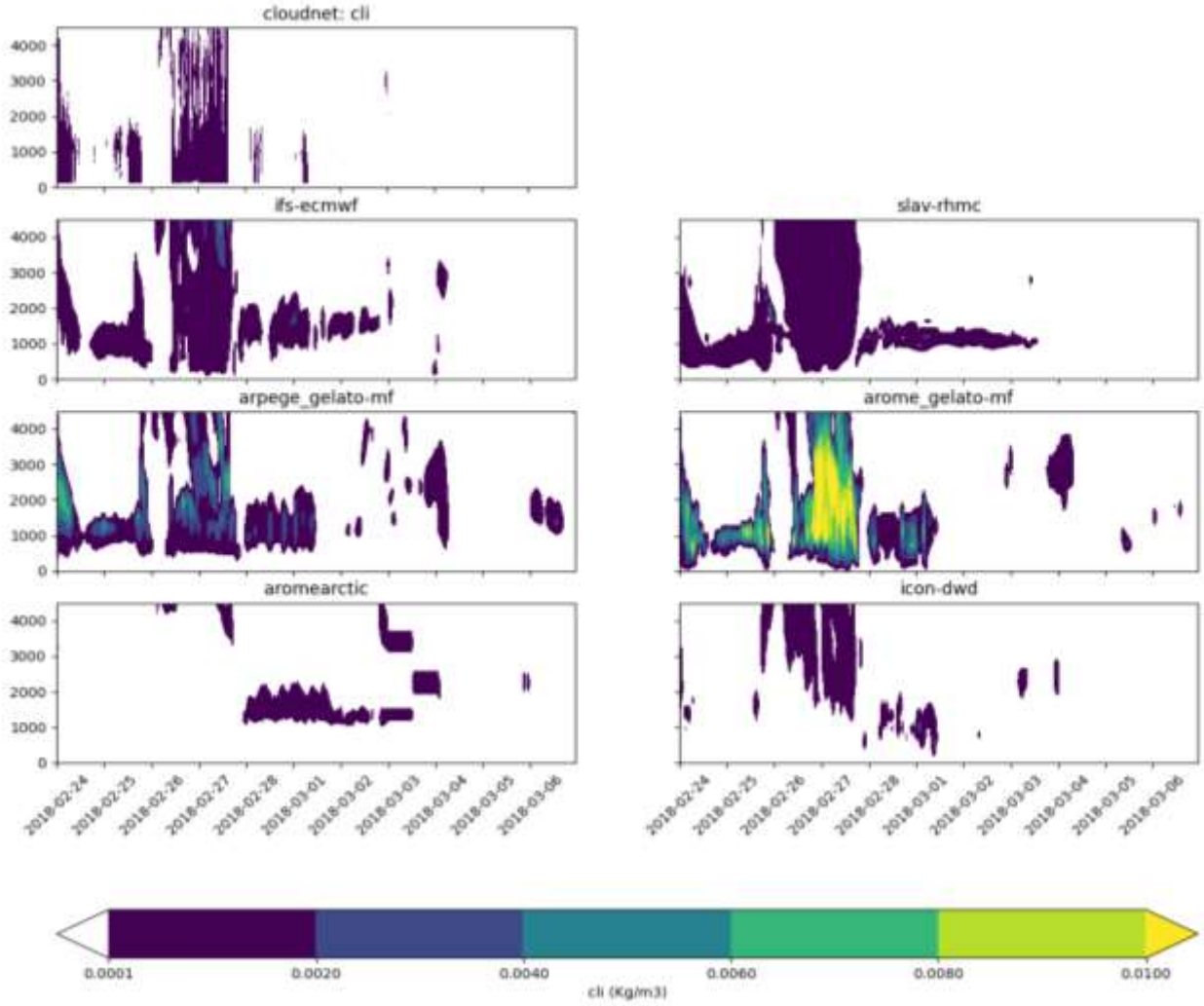


Figure 7. Observed and modelled Cloud-ice content at Ny-Ålesund, Svalbard, for the period 24th Feb to 6th March 2018. The model timeseries shown are produced by concatenating hourly output from day-2 of the forecast (T+24-T+47).

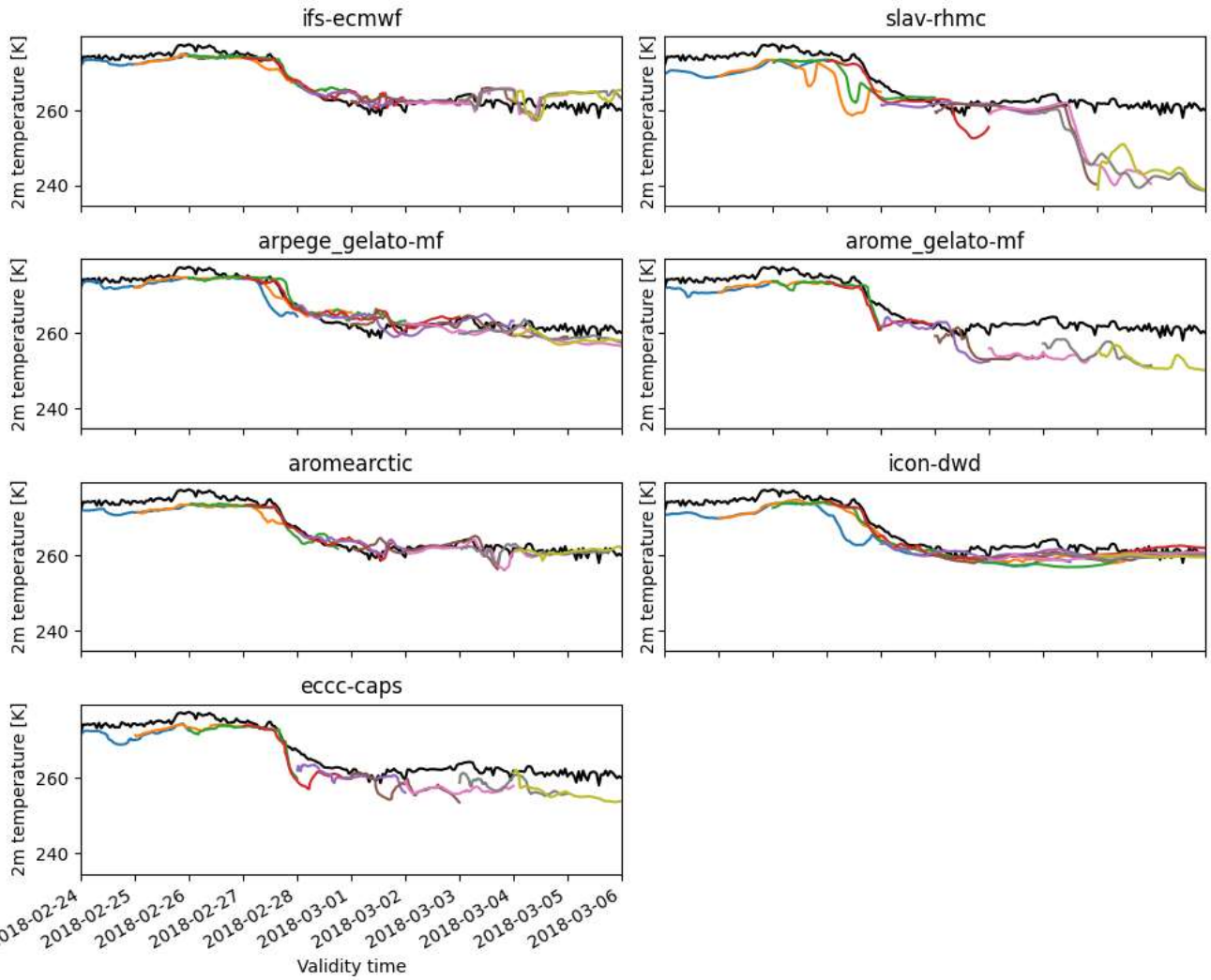


Figure 8. Observed and forecast 2m-temperature at Ny-Ålesund, Svalbard, for the period 24th Feb to 6th March 2018. Forecasts initialised on different days are represented in different colours.

3.2.2. Representation of stable boundary layer

One of the benefits of having the MODFs and MMDFs for many different sites is that we can look for common issues across the Arctic region. One of the meteorological situations which is poorly represented in models is stably stratified boundary layers (Atlaskin and Vihma, 2012; Sandu et al., 2013).

To compare the models with observations during stably stratified conditions, we need selection criteria for sub-setting. To define this, we plot boundary-layer regime diagrams for the four sites we have MODFs (Fig 9). These 2-D histograms show the strength of the temperature inversion at each location and the net LW radiation. One can see that temperature inversions tend to be accompanied by longwave cooling and we select points at each site with inversion strength (delT) larger than 0.5

and LW_{net} lower than -20W/m2 (i.e. the points to the bottom and left of the horizontal and vertical red lines). At all of the sites these criteria select the most stable cases.

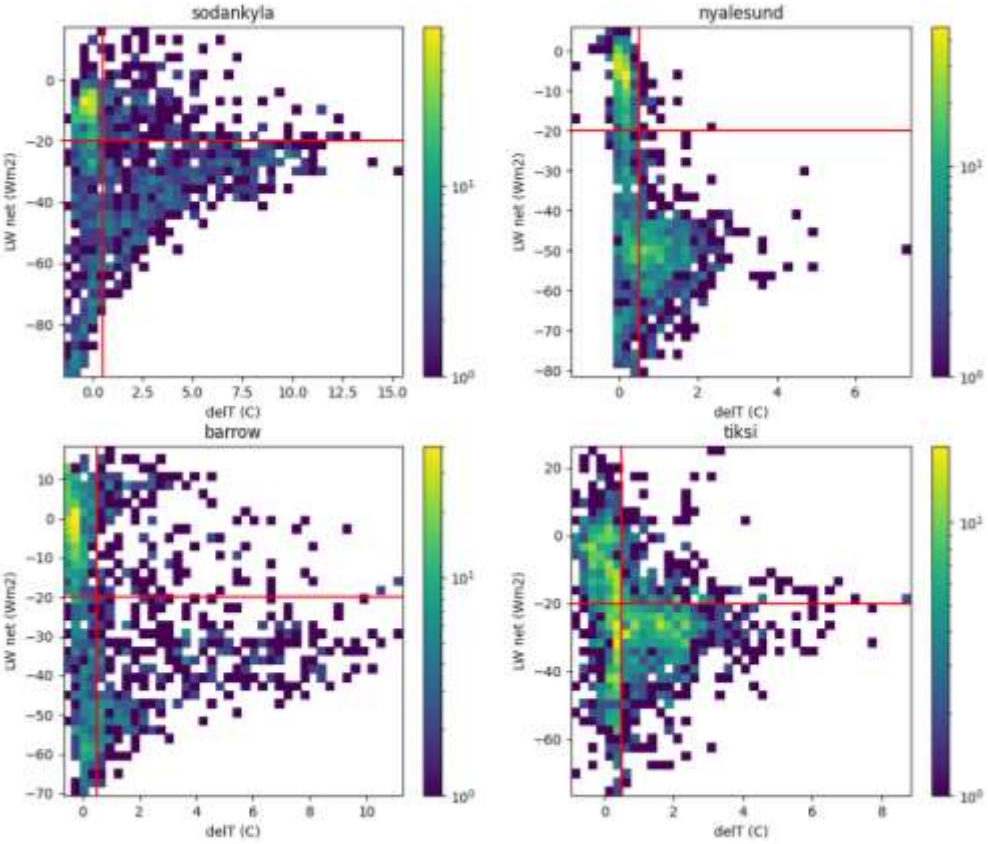


Figure 9. Boundary layer regime histograms (net longwave radiation vs the temperature difference between the top and bottom temperature at the met-tower at each site) for Sodankylä, Ny-Ålesund, Barrow and Tiksi.

The composite profiles, averaged when the observations meet the criteria set out above, are shown in Figures 10 and 11. Note that the same times are selected from the forecasts and the observations. Figure 10 shows the profiles over the lowest 100m of the atmosphere, hence the profiles of observations from the micro-meteorological tower are shown. Figure 10 shows that, with the exception of Ny-Ålesund, air temperature and humidity tend to be too high close to the surface. This error is largest in the 2m level.

At Tiksi, which has the largest inversion depth (up to 1000m), the models are too warm through the whole inversion depth and do not capture the strength of the temperature or humidity inversion. Temperature and humidity inversions are also too weak (or not present in the mean model profile) at Sodankylä and Barrow.

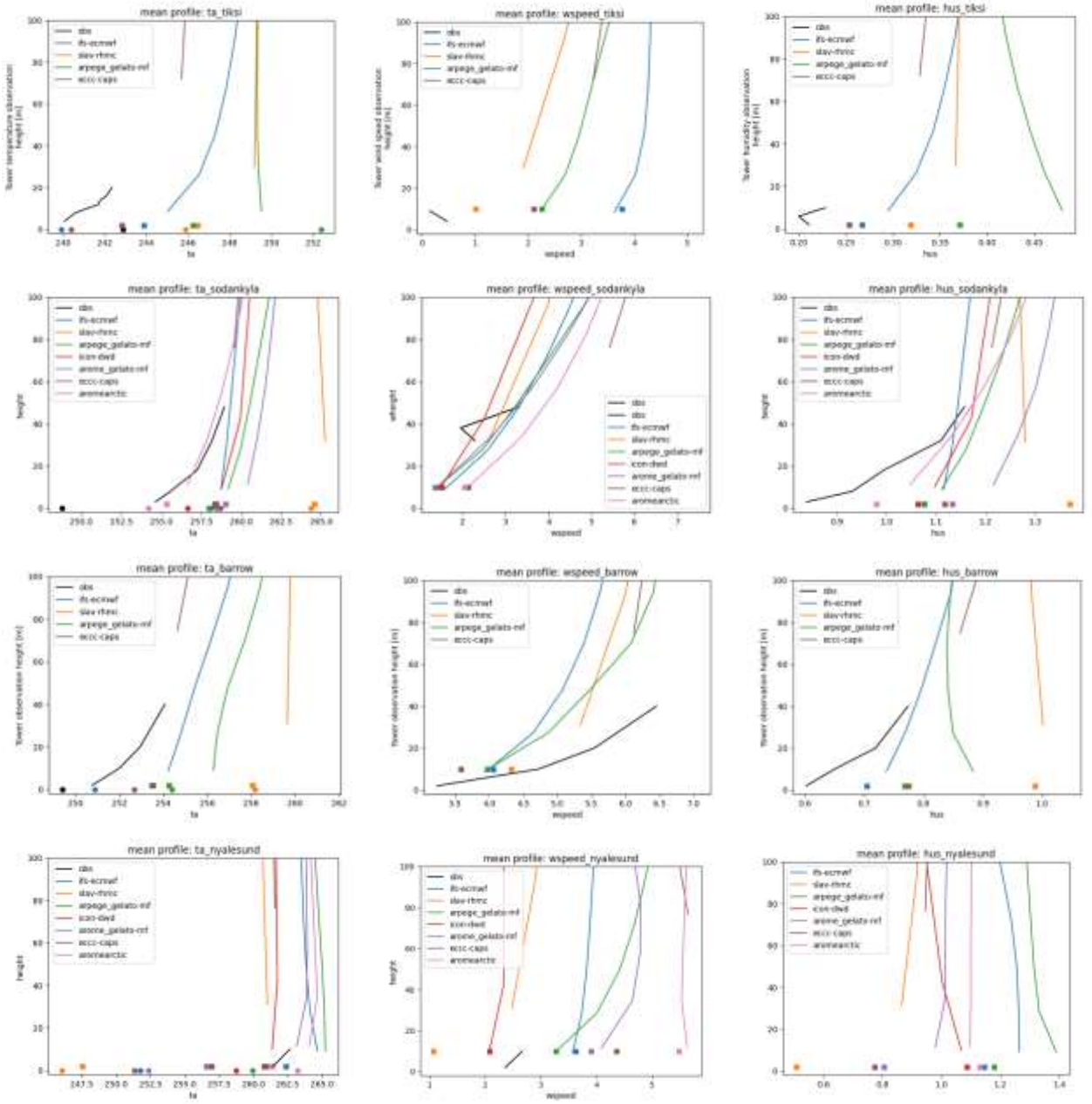


Figure 10. Mean observed (from micro-meteorological tower) and modelled atmospheric temperature (left), wind speed (middle) and specific humidity profiles (right) for SBL/LW cooling conditions. Dots show the mean surface temperature, and the squares show the 2m temperature (left) 10m wind speed (middle) and the 2m specific humidity (right).

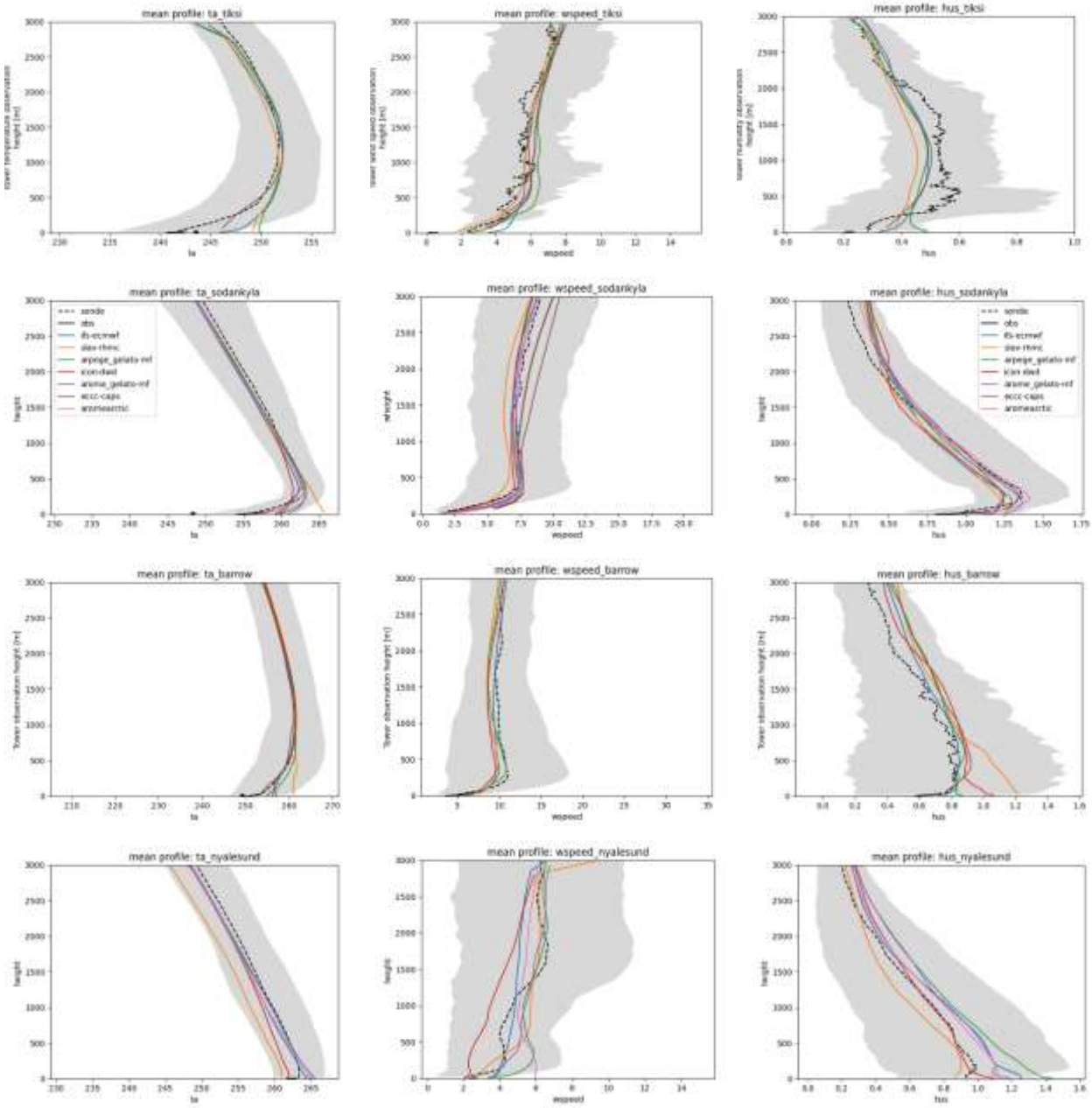


Figure 11. Mean observed (from micro-meteorological tower [solid line] and radiosonde [dashed line]) and modelled atmospheric temperature (left), wind speed (middle) and specific humidity profiles (right) for the lower troposphere during SBL/LW cooling conditions.

At both Barrow and Sodankylä the mean wind speed profiles have a local wind speed maximum above the first few hundred meters. This general feature is captured by the models, but the speed in this layer is generally too low in the models at Barrow.

The rate of cooling of the atmosphere during SBLs is intimately connected to the surface energy budget since LW deficit at the surface must be balanced by the turbulent fluxes from the atmosphere and conductive heat flux from the snow. This balance can be seen in the observed fluxes

at Sodankylä, where a radiative deficit (R_{net}) of 25W/m2 is approximately balanced by the flux from the surface (Q_{net}=18W/m2)) and the sensible heat flux (SHF=8W/m2). Looking across the different models one can see that the R_{net} is typically more than 30% larger, partly due to the warmer surface temperatures in the model and partly the result of differences in the albedo (not shown). This difference in R_{net} is balanced by larger than observed turbulent fluxes from the atmosphere and conductive fluxes from the snow.

Table 2. Mean fluxes at Sodankylä during SBL conditions (based on 474 cases). All fluxes are positive upward.

	Obs	**IFS**	**SLAV**	**ARPEGE**	**ICON**	**AROME-MF**	**CAPS**	**AROME-MN**
SHF	-8.2	-20.3	-9.0	-14.9	-17.8	-19.7	-9.5	-17.0
LHF	-0.2	-1.3	3.8	-1.0	-2.6	0.0	1.2	-1.8
Rnet	25.6	44.7	55.3	45.8	42.5	41.0	33.6	37.8
Qnet	18.0	23.1	50.2	29.9	22.1	21.3	25.3	19.1
LWdown	181.7	196.0	205.5	194.8	N/A	197.7	212.2	188.7

3.2.3. Parameterisation of wind stress

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

At Sodankylä turbulent fluxes exist as well as profiles of wind speed, temperature and humidity in the prototype MODF. Turbulent fluxes in the models are parameterised according to Monin-Obukhov similarity theory (Beljaars and Holtslag, 1991). For example, the surface stress (τ) is expressed as:

$$\tau = \rho C_M U^2$$

where U is the wind speed at the lowest model level, ρ is the air density and C_M is the transfer coefficient for momentum. C_M is a complicated function of the roughness length and depends on the atmospheric stability. So the challenge of parameterizing τ comes down to defining a reasonable function and selecting the appropriate parameters. Because we have observed and forecast τ and U, we can check how well this was done in each model by plotting these parameters against each other (see Figure 12) as done previously by Tjernstrom et al. (2005).

If conditions were always neutral, the points in Figure 12 would sit on the straight line:

$$\tau = \rho \frac{kU^2}{\ln(\frac{z1}{z0m})},$$

where z1 is the height of the lowest model level, k is the von Karman constant and z0m is the aerodynamic roughness length. The slope of this line is determined only by the roughness length. However, this is not the case as the atmosphere is often stably stratified and as a result there is

much scatter in the values of τ for any given wind speed. In practice this is taken into account by adding a stability correction to this function.

Looking at the relationship between τ and U (Figure 12) one can see quite different behaviours in different models and compared to observations. The roughness length for the observations was also calculated from the previous equation after selecting data for neutral conditions and is presented along with the values used by each model in Table 3. 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 a roughness length an order of magnitude lower than that derived from observations. More generally even the models that correct for stability do not adequately capture the spread τ for a given value of U. This may be because the model underestimates the atmospheric stability, as is suggested by the too-weak inversions shown in Fig 10. Or it could be a more general issue with the formulation of the stability functions. A more detailed study would be needed to investigate this.

Interestingly, in the ICON model the points sit on a straight line since the C_M used is a neutral transfer coefficient and there is no correction for atmospheric stability.

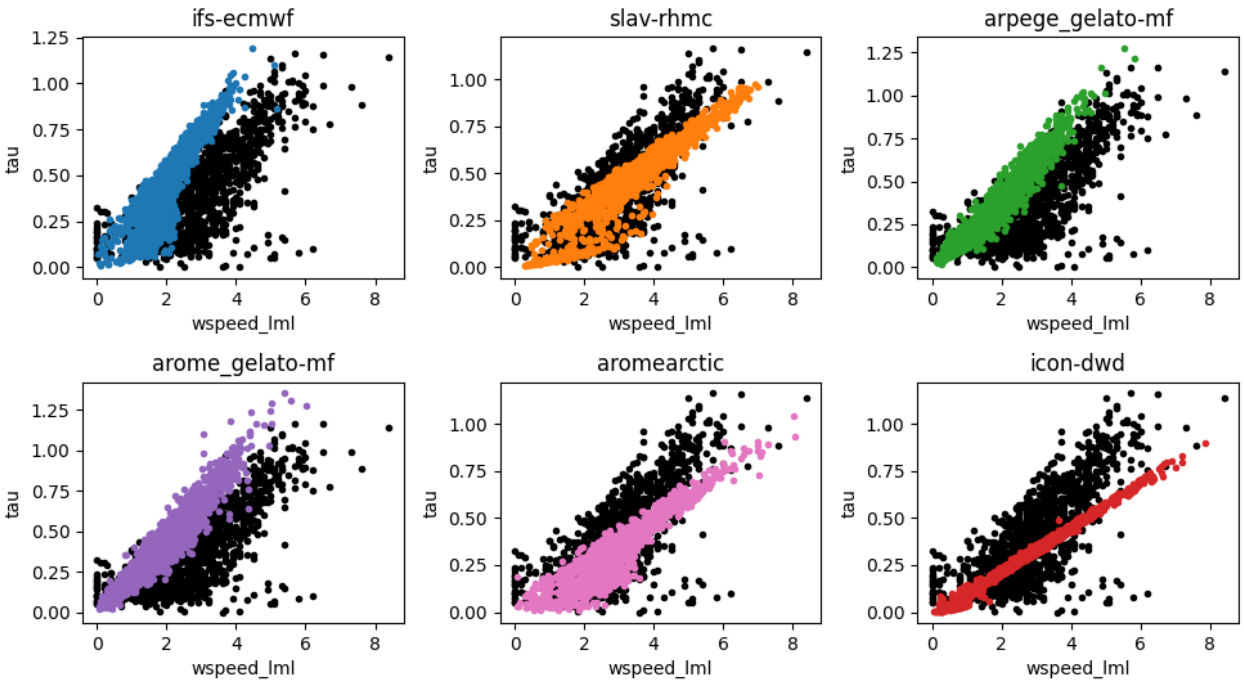


Figure 12. Scatter plots of wind stress vs. the square of the near surface wind at Sodankylä. The observed points are shown in black and the day-2 forecast is shown in colours.

Table 3. Roughness lengths for momentum at Sodankylä.

	Sodankylä
Obs	1.62
IFS	1.82
ARPEGE	1.50
SLAV	1.53
ICON-DWD	0.2
CAPS	Not in file
AROME-Arctic-No	0.45

4. Conclusions and next steps

Data from Arctic research stations has been under-utilised for evaluating forecast models. This is partly an issue of interoperability of observational data. Relevant data from different sites is typically stored in different archives, usually those of the organisation maintaining the site, and there has been little overlap in the data format, quality control or data processing between different archives. This has made it difficult to extend analysis, such as the forecast evaluation across different sites. To address this the MODF and MMDF formats for observational and forecast data have been developed and described above.

This report documents the current status and presents some multi-model forecast evaluation examples to demonstrate the utility of the MMDF and MODF approach and identifies issues with the performance of the models that are common across a number of Arctic locations. This helps to inform how common they are and provides some guidance of how models can be improved. For example, the parameterization of momentum fluxes was evaluated at Sodankylä indicating that the roughness lengths in the ICON and AROME models was too high resulting in unrealistic surface stresses for a given value of the wind speed.

The development of the MODFs is ongoing and will be completed in phases. The initial phase was to collect basic meteorology data and potentially turbulent fluxes. Work on this initial phase is still ongoing but is expected to be finished in advance of the YOPP final summit, to be held in Montreal in August, where it will be presented. The next phase will be to get more cloud information into the MODFs. This is more complicated, but very necessary since the models differ hugely in terms of cloud properties as shown in section 3.2.1.

The next step is then to use the knowledge derived from the diagnostics developed with the supersite data to motivate and evaluate new development of forecasting systems.

5. References

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

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

Beljaars, A. C. M. and Holtslag, A. a. M.: Flux Parameterization over Land Surfaces for Atmospheric Models, J. Appl. Meteor., 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, 145, 1919–1935, https://doi.org/10.1175/MWR-D-16-0417.1, 2017.

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), 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, 14–17, https://doi.org/10.21957/xbov3ybily, 2017.

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, 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, 12, https://doi.org/10.1029/2020MS002144, 2020.

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.

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. Amer. Meteor. Soc., 88, 883–898, https://doi.org/10.1175/BAMS-88-6-883, 2007.

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. Amer. Meteor. Soc., 97, 1631–1647, https://doi.org/10.1175/BAMS-D-14-00246.1, 2016.

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, Wea. Forecasting, 34, 959–983, https://doi.org/10.1175/WAF-D-19-0003.1, 2019.

Lawrence, H., Bormann, N., Sandu, I., Day, J., Farnan, J., and Bauer, P.: Use and impact of Arctic observations in the ECMWF Numerical Weather Prediction system, 145, 3432–3454, https://doi.org/10.1002/qj.3628, 2019.

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, 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., and Bennartz, R.: Process-Based Model Evaluation Using Surface Energy Budget Observations in Central Greenland, 123, 4777–4796, https://doi.org/10.1029/2017JD027377, 2018.

Pithan, F., Medeiros, B., and Mauritsen, T.: Mixed-phase clouds cause climate model biases in Arctic wintertime temperature inversions, 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.

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

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, 12, e2020MS002075, https://doi.org/10.1029/2020MS002075, 2020.

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?, Journal of Advances in Modeling Earth Systems, 5, 117–133, https://doi.org/10.1002/jame.20013, 2013.

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

Serreze, M. C. and Barry, R. G.: The Arctic Climate System, Cambridge University Press, 2009.

Shupe, M. D. and Intrieri, J. M.: Cloud Radiative Forcing of the Arctic Surface: The Influence of Cloud Properties, Surface Albedo, and Solar Zenith Angle, 17, 616–628, https://doi.org/10.1175/1520-0442(2004)017<0616:CRFOTA>2.0.CO;2, 2004.

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

Svensson, G., Casati, B., Day, J. J., Uttal, T., Godøy, Ø., and Hartten, L. M.: YOPPsiteMIP – The YOPP site Model Inter-comparison Project Experiment description, 2020.

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,' 117, 337–381, https://doi.org/10.1007/s10546-004-7954-z, 2005.

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 SL-AV: the Results of Medium-range Weather Forecasts, 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. Amer. Meteor. Soc., 97, 1033–1056, https://doi.org/10.1175/BAMS-D-14-00145.1, 2015.

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, 141, 563–579, https://doi.org/10.1002/qj.2378, 2015.