

Evaluation of a sub-kilometre NWP system in an Arctic fjord-valley system in winter

Teresa Valkonen , Patrick Stoll , Yurii Batrak , Morten Køltzow , Thea Maria Schneider , Emmy E. Stigter , Ola B. Aashamar , Eivind Støylen & Marius O. Jonassen

To cite this article: Teresa Valkonen , Patrick Stoll , Yurii Batrak , Morten Køltzow , Thea Maria Schneider , Emmy E. Stigter , Ola B. Aashamar , Eivind Støylen & Marius O. Jonassen (2020) Evaluation of a sub-kilometre NWP system in an Arctic fjord-valley system in winter, Tellus A: Dynamic Meteorology and Oceanography, 72:1, 1-21, DOI: [10.1080/16000870.2020.1838181](https://doi.org/10.1080/16000870.2020.1838181)

To link to this article: <https://doi.org/10.1080/16000870.2020.1838181>



Tellus A: 2020. © 2020 The Author(s).
Published by Informa UK Limited, trading as
Taylor & Francis Group.



Published online: 28 Oct 2020.



Submit your article to this journal [↗](#)



Article views: 396



View related articles [↗](#)



View Crossmark data [↗](#)

Evaluation of a sub-kilometre NWP system in an Arctic fjord-valley system in winter

By TERESA VALKONEN^{1*}, PATRICK STOLL², YURII BATRAK¹, MORTEN KØLTZOW¹, THEA MARIA SCHNEIDER³, EMMY E. STIGTER⁴, OLA B. AASHAMAR⁵, EIVIND STØYLEN¹, and MARIUS O. JONASSEN^{6,7}, ¹*Development Centre for Weather Forecasting, Norwegian Meteorological Institute, Oslo, Norway;* ²*Department of Physics and Technology, University of Tromsø, Tromsø, Norway;* ³*Department of Physics, University of Potsdam, Potsdam, Germany;* ⁴*Department of Physical Geography, Utrecht University, Utrecht, The Netherlands;* ⁵*Forecasting Department, Norwegian Meteorological Institute, Tromsø, Norway;* ⁶*Department of Arctic Geophysics, The University Centre in Svalbard, Longyearbyen, Norway;* ⁷*Geophysical Institute, The University of Bergen, Bergen, Norway*

(Manuscript Received 19 June 2020; in final form 12 October 2020)

ABSTRACT

Terrain challenges the prediction of near-surface atmospheric conditions, even in kilometre-scale numerical weather prediction (NWP) models. In this study, the ALADIN-HIRLAM NWP system with 0.5 km horizontal grid spacing and an increased number of vertical levels is compared to the 2.5-km model system similar to the currently operational NWP system at the Norwegian Meteorological Institute. The impact of the increased resolution on the forecasts' ability to represent boundary-layer processes is investigated for the period from 12 to 16 February 2018 in an Arctic fjord-valley system in the Svalbard archipelago. Model simulations are compared to a wide range of observations conducted during a field campaign. The model configuration with sub-kilometre grid spacing improves both the spatial structure and overall verification scores for the near-surface temperature and wind forecasts compared to the 2.5-km experiment. The sub-kilometre experiment successfully captures the wind channelling through the valley and the temperature field associated with it. In a situation of a cold-air pool development, the sub-kilometre experiment has a particularly high near-surface temperature bias at low elevations. The use of measurement campaign data, however, reveals some encouraging results, e.g. the sub-kilometre system has a more realistic vertical profile of temperature and wind speed, and the surface temperature sensitivity to the net surface energy is closer to the observations. This work demonstrates the potential of sub-kilometre NWP systems for forecasting weather in complex Arctic terrain, and also suggests that the increase in resolution needs to be accompanied with further development of other parts of the model system.

Keywords: numerical weather prediction; atmospheric boundary layer; cold-air pool; valley channelling; Svalbard

1. Introduction

Dramatic environmental changes are leading to a growth and diversification of human activity in the Arctic region (Dawson et al., 2017). Numerous sectors, such as transportation, tourism, fishing and scientific operations, are increasingly dependent on accurate weather information. Numerical weather prediction (NWP) models provide direct and indirect guidance for safe operations in the Arctic.

Arctic NWP capabilities have advanced in recent years due to, among other improvements, increased model resolution (Bauer et al., 2016). Several operational weather centres have started to run convection-permitting kilometre-scale models for limited areas at high latitudes. For example, the ALADIN-HIRLAM (Aire Limitée Adaptation dynamique Développement InterNational-High Resolution Limited Area Model) NWP system is run at 2.5 km horizontal grid spacing both for Greenland (Yang et al., 2018) and the European Arctic (Müller et al., 2017), and the Canadian Arctic Prediction System

*Corresponding author. e-mail: teresav@met.no

with 3-km grid spacing (Casati et al., in preparation) for a pan-Arctic domain. The added value of the kilometre-scale regional models over global models has been found to be most pronounced near complex coastlines and in mountainous terrain (e.g. Køltzow et al., 2019).

Complex coastlines and mountainous terrain at high latitudes strongly influence the atmospheric flow and generate phenomena with large spatial variability such as wind channelling (Nawri and Stewart, 2008), down-slope wind storms (Oltmanns et al., 2015) and cold-air pools in valleys and basins (Clements et al., 2003). Improved description of orography enables the models to capture the terrain-induced atmospheric features (e.g. Moore et al., 2015), but these fine-scale features still pose challenges for NWP even at kilometre-scale grid spacing.

A number of NWP studies have employed kilometre-scale grid spacing on Svalbard. Køltzow et al. (2019) show that large near-surface wind speed errors and a general cold bias, which change to a warm bias in the presence of a stable boundary layer (SBL), is common for NWP systems. The latter warm bias is a well-known problem in NWP (e.g. Atlaskin and Vihma, 2012; Sandu et al., 2013; Esau et al., 2018) and has also been investigated on Svalbard employing 1-km grid spaced models (Kilpeläinen et al., 2011; Kilpeläinen et al., 2012; Mayer et al., 2012). The reduced grid spacing gives only a minor improvement on temperature forecasts, but a substantial improvement for wind. The same studies additionally indicate that a better description of the orography improves wind forecast quality, while temperature forecasts are more sensitive to surface type.

Extending the horizontal resolution to sub-kilometre grid spacing has been proposed as a possible solution for further improving the forecast skill. A sub-kilometre set-up of the Canadian NWP system in wintertime mountainous terrain improved wind speed and temperature forecasts at high altitudes (Vionnet et al., 2015). Yang (2019) found that using 750 m grid spacing in the ALADIN-HIRLAM NWP system gave a substantial improvement in forecasting the spatial variability in wind extremes in Greenland. However, in relatively flat terrain around a mid-latitude airport Hagelin et al. (2014) found only minor improvements. These studies had a limited set of observations available to investigate the weather development and the model performance in detail. Particularly, the vertical structure of the atmosphere remained mainly unstudied. Furthermore, Køltzow et al. (2019) argued that a substantial part of the difference between forecast and observation comes from subgrid variability not resolved by the kilometre-scale models. Contrary to this, Atlaskin and Vihma (2012) argued that the subgrid variability plays a less important role for errors in the SBL in relatively flat terrain.

The purpose of this study is to evaluate the benefits of a sub-kilometre NWP system in predicting near-surface atmospheric conditions over an Arctic fjord-valley system compared to the 2.5-km system currently operational at the Norwegian Meteorological Institute. The ALADIN-HIRLAM NWP system with 0.5 km horizontal grid spacing is run for 5 days in February 2018 over a domain covering the Svalbard archipelago in the Arctic Ocean. The results are compared to both the same model system with 2.5 km grid spacing and a wide set of observations conducted during a measurement campaign in the Adventdalen valley. During the measurement campaign, several temporary weather stations were set up, components of the surface energy budget were measured, spatial variability of temperature was obtained from snowmobile transects and a tethersonde was utilised for measuring the vertical structure of the atmosphere. We use the campaign data, in addition to the data from permanent weather stations, to study the impact of the increased resolution on the forecast ability to represent the boundary layer in an Arctic valley. Compared to previous studies in Svalbard, we utilise finer horizontal and vertical resolution close to the surface. In addition, compared to previous sub-kilometre studies, we compare NWP results with a wider set of observations. We also analyse the surface energy budget (SEB) to investigate the coupling between the surface and the atmosphere in the performed experiments. The studied period is a part of the Year of Polar Prediction (YOPP) Special Observing Period Northern Hemisphere 1 (SOP-NH1). Therefore, this study is also a contribution to YOPP.

2. Data and methods

2.1. Study site and period

This study focuses on the Adventdalen fjord-valley system located in the central part of Spitsbergen, the largest island of the Svalbard archipelago (Fig. 1). The valley is approximately 4 km wide and 30 km long and has several smaller side valleys, which are approximately 1 km wide (Fig. 2). The valley ends at the 8 km long Adventfjorden, which is connected to the larger Isfjorden. The surrounding mountains reach elevations up to about 1000 m. The investigated area is a typical fjord-valley system on Spitsbergen.

The study period was from 12 to 16 February, 2018, at the end of the polar night. The period was characterised by a high pressure situation with prevailing weak easterly flow and very little precipitation. The land areas were fully covered with snow. The snow thickness was approximately 20 cm on flat areas in Adventdalen, while it varied locally elsewhere due to redistribution of snow by wind.

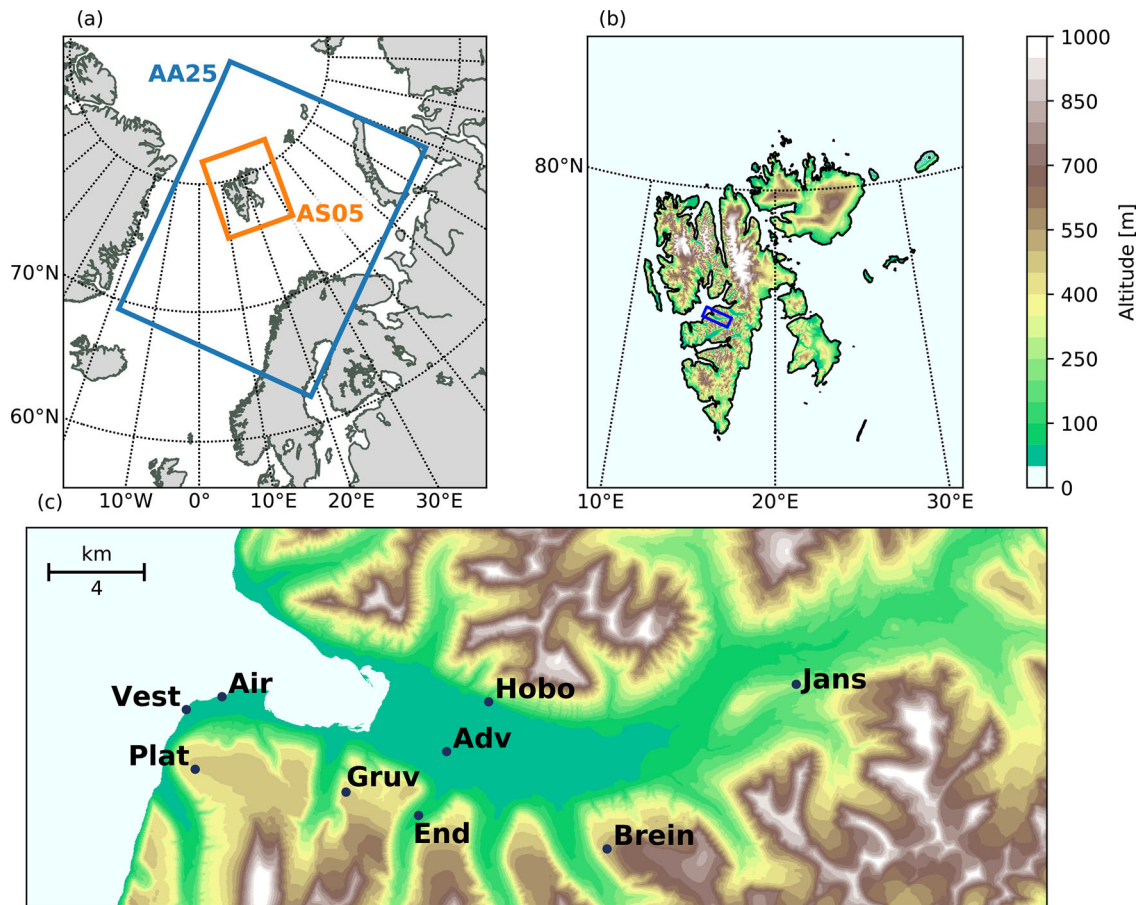


Fig. 1. Location of the model domains and weather stations in the Adventdalen valley. (a) The model domains of AA25 at 2.5 km and AS05 at 0.5 km horizontal grid spacing. (b) AS05 model domain with the representation of the model topography. The blue box shows the location of the Adventdalen valley. (c) Close-up view of the Adventdalen valley showing the local topography according to the Norwegian Polar Institute (2014) and the locations of the weather stations: Vestpynten (Vest), Svalbard Airport (Air), Platåberget (Plat), Gruvefjellet (Gruv), Adventdalen (Adv), Hobo, Endalen (End), Breinosa (Brein) and Janssonhaugen (Jans).

Sea ice was partly present in Adventfjorden. In the beginning of the study period, new ice started to form in the otherwise ice-free Adventfjorden. By the end of the period, a large part of the fjord was ice-covered. Isfjorden was free of sea ice during the whole period. The average 2-m temperature was -8.2°C at the Svalbard Airport from 12 to 16 February, and the study period can be considered representative of typical winter conditions in the region in the current climate (Isaksen et al., 2016).

2.2. Model simulations

The model system utilised in this study is the HARMONIE-AROME (HIRLAM-ALADIN Research on Mesoscale Operational NWP in Euromed-Application of Research to Operations at Mesoscale) configuration (Bengtsson et al., 2017) of the ALADIN-HIRLAM NWP model system version 40h1.1. Two model

experiments were set up with different model domains and resolution. The first one, AA25, was run with 2.5 km horizontal grid spacing and 65 vertical levels over a domain covering the European Arctic, including the Svalbard archipelago (Fig. 1a). The same model system for the same domain and model resolution is run operationally by the Norwegian Meteorological Institute under the name AROME-Arctic (Müller et al., 2017). The second experiment, AS05, was run with 0.5 km horizontal grid spacing and 90 vertical levels over a domain covering the Svalbard archipelago (Fig. 1b). While the lowest two model levels in AA25 are at 12 and 35 m above the ground level (AGL), the same levels in AS05 are 5 and 15 m AGL. Even with the much smaller domain, AS05 is about 14 times more computationally expensive than AA25 due to the higher number of grid points and the shorter integration timestep. Details of model experiments are listed in Table 1.

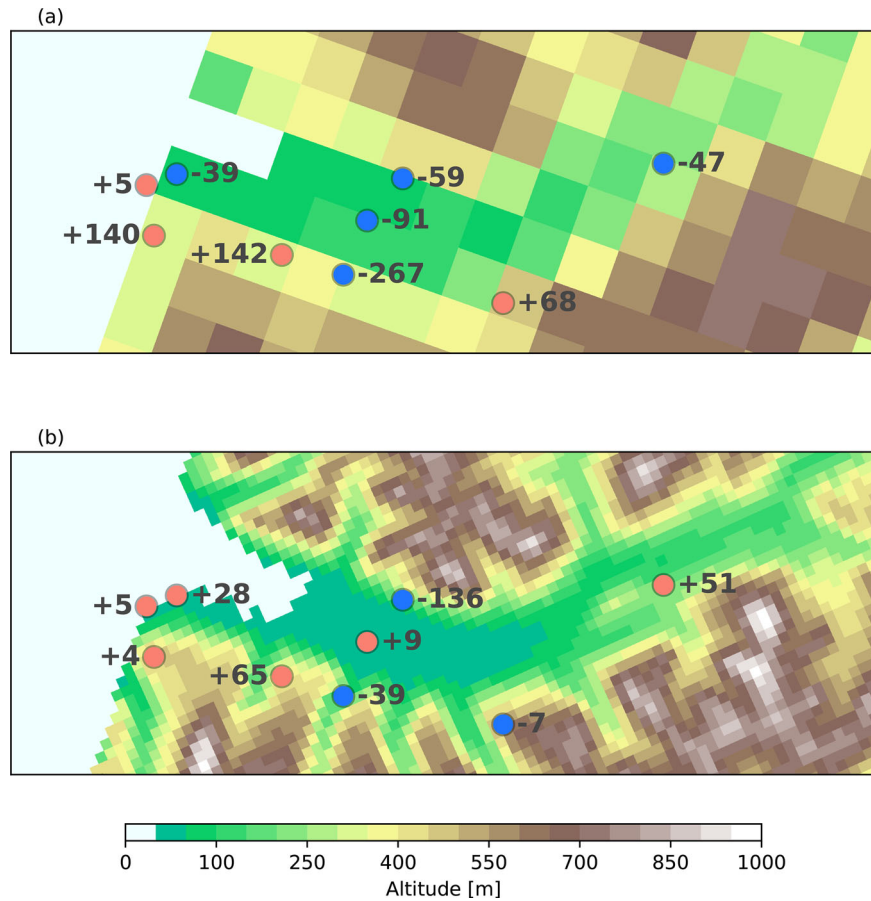


Fig. 2. The topography of the Adventdalen area and difference between the altitude of the weather stations and model topography: (a) AA25, (b) AS05. Positive altitude difference means that the altitude of the weather station is above the model terrain elevation of the nearest grid cell.

Table 1. Model experiments and their configurations.

	AA25	AS05
Horizontal grid spacing (km)	2.5	0.5
Number of horizontal grid points	750 × 960	1200 × 1440
Domain size (km × km)	1875 × 2400	600 × 720
Number of vertical levels	65	90
Height of two lowest model level (m)	12, 35	5, 15
Number of levels below 1000 m	21	23
Integration time step (s)	60	15
Lateral boundaries	IFS-HRES 1-hourly	AA25 1-hourly
Data assimilation atmospheric model	3DVAR 3-hourly, first guess from the operational AROME-Arctic	none
Data assimilation surface model	Optimal interpolation for T2m, RH2m and snow depth	Optimal interpolation for T2m, RH2m and snow depth
Forecast length	24 h starting from 00:00 UTC	24 h starting from 00:00 UTC

At the beginning of the simulation period at 11 February 21:00 UTC, atmospheric and surface forecasts from the operational AROME-Arctic were used as model

first guess in AA25 in order to avoid spin-up problems and to ensure a realistic description of surface variables, such as sea ice and snow properties. AS05 was started

Table 2. List of weather stations in the Adventdalen area and observations used in this study.

Station	Abbreviation	Altitude (m)	AA25 altitude (m)	AS05 altitude (m)	Temperature (m)	Wind (m)	Additional observations
Vestpynten	Vest	5	0	0	2	2.3, 4.7	—
Adventdalen*	Adv	15	106	6	0, 2	10	Radiation, sonic, tethersonde, snowmobile tracks
Svalbard Airport*	Air	28	67	0	2	10	Cloud cover
Hobo	Hobo	40	99	176	2	1.9, 3.4	Snowmobile tracks
Endalen	End	78	345	117	2	1.2, 3.2	—
Janssonhaugen*	Jans	251	298	200	1, 3	3	—
Platåberget*	Plat	450	310	446	2	10	—
Gruvefjellet*	Gruv	464	322	399	1, 3	3	—
Breinoså*	Brein	520	452	527	2	4.6	—

*Permanent weather stations.

from interpolated initial conditions from AA25 at the beginning of the simulation period. After this, data assimilation procedures were applied every 3 h for the atmosphere in AA25 and for the surface both in AA25 and AS05. In AA25, the use of the operational AROME-Arctic forecasts as model first guess continued 3-hourly throughout the whole simulation period. This made it possible to have AA25 practically identical with the operational model. In AS05, the atmosphere was run without observation assimilation or initialisation from AA25 allowing high-resolution features to develop. The surface parameterisation package of HARMONIE-AROME was configured to use the optimal interpolation method to assimilate snow depth, and screen-level temperature and relative humidity in both AA25 and AS05. Starting from 00:00 UTC analysis times, forecasts up to 24 h were performed.

Lateral boundary conditions for AA25 were provided hourly by the European Centre for Medium-Range Weather Forecasting (ECMWF) high resolution global model (IFS-HRES). The sub-kilometre AS05 received its lateral boundary conditions from AA25 allowing us to investigate directly the added value of the sub-kilometre forecasts over the kilometre-scale system and reduce the spin up distance from the lateral boundaries.

HARMONIE-AROME uses the surface parameterisation package SURFEX (SURFace EXternalisée) (Masson et al., 2013) for representing surface processes. SURFEX in AA25 and AS05 was configured similar to the operational AROME-Arctic system: land surfaces are parameterised by the three-layer version of the ISBA (Interactions between the Soil–Biosphere–Atmosphere)

force-restore scheme (Boone et al., 1999) with the single-layer snow model (Douville et al., 1995), open ocean is defined by the prescribed sea surface temperature field and turbulent fluxes calculated according to the ECUME (Exchange Coefficients from Unified Multi-campaigns Estimates) scheme (Belamari, 2005), sea ice cover is parameterised by a one-dimensional thermodynamic scheme (Batak et al., 2018). In AA25 and AS05 the 2-m and surface temperature and 2-m humidity are diagnostics computed by SURFEX. The 10-m wind speed in AA25 is computed by SURFEX. In AS05, where the lowest model level is below 10 m AGL, the 10-m wind speed is linearly interpolated between the two lowest model levels which are at 5 m and 15 m height.

The topography in the model experiments was described by Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010, Danielson and Gesch, 2011). We utilised the elevation data with a resolution of 7.5 arcseconds of latitude and longitude (approximately 250 m). In AS05 the altitude of the corresponding model-grid cell is more accurately represented at six, less accurately at two and similarly at one of the stations than in AA25 (Fig. 2 and Table 2). Averaged over all 9 stations, the mean absolute error of station altitude is considerably lower in AS05 (38 m) than in AA25 (95 m).

To pair the NWP model data with point observations we use the nearest neighbour interpolation for all NWP simulations. This choice preserves best the difference in resolution which we want to investigate in the study (e.g. Accadia et al., 2003). Forecast lead times included are from +1 h to +24 h starting from the forecasts initialised at 00:00 UTC each day.

2.3. Observations

Observational data for model comparison were obtained from (1) automatic weather stations, (2) a tethersonde operated at the Adventdalen weather station, and (3) snowmobile transects performed in the Adventdalen valley.

Both permanent and temporary weather stations provide data for the model comparison. Permanent weather stations are located in Adventdalen, Janssonhaugen, Gruvefjellet, Breinosa, Svalbard Airport and Platåberget (Fig. 1). The largest distance between the observation sites (Janssonhaugen and Platåberget) is approximately 25 km. During the campaign conducted by the University Centre in Svalbard (UNIS), temporary weather stations were set up in Endalen, at Vestpynten and on the Northern side of Adventdalen (Hobo). The latter provides observations only from 12 February 12:00 UTC to 16 February 03:00 UTC.

The station specifications are summarised in Table 2. At the Janssonhaugen and Gruvefjellet stations, the 2-m temperature is estimated by averaging the 1-m and 3-m temperatures. For stations with wind measurements from lower elevations, the 10-m wind speed was estimated by assuming a log wind profile and a roughness length of 0.003 m which is the same value as used in the model simulations. For temperature and humidity, the 1 min-average at the beginning of each hour is used for model comparison, while the 10 min-average is used for wind measurements. The measurement uncertainty of all instruments included in this study is considerably smaller than the difference between simulated and observed values and, therefore, have negligible effect on the results.

A radiometer is permanently installed at the Adventdalen weather station. Longwave up- and downwelling radiation were measured by a radiometer. The shortwave radiation was negligible during the observation campaign since the campaign was performed during the polar night.

A sonic anemometer was used to measure three-dimensional wind speed components as well as the sonic temperature at 20 Hz frequency at the Adventdalen weather station. The data processing to derive the surface sensible heat flux consisted of several steps, following Stigter et al. (2018). Only the correction method for the sensor tilt was chosen differently. Double rotation was used as this correction is more suitable for shorter time series consisting of only a few days. The derived sensible heat fluxes were quality checked according to Mauder and Foken (2004). Fluxes with a low-quality were exempted from the calculation. Finally, the 10-minute sensible heat fluxes were aggregated to hourly values to reduce flux sampling errors (Vickers and Mahrt, 1997). The latent heat fluxes were computed using the bulk-aerodynamic method and

station data is described in detail by Litt et al. (2015). The roughness lengths for momentum, humidity and sensible heat in the bulk-aerodynamic method were tuned with the observed sensible heat fluxes.

The low-level and total cloud cover is manually observed at the Svalbard Airport at three hour time intervals in the unit okta, i.e. one okta equals an eighth of the sky dome covered by clouds.

A tethersonde was operated near the Adventdalen weather station to observe the vertical atmospheric structure of temperature, humidity and the wind speed up to an altitude of 1 km. To measure spatial temperature variability in the Adventdalen area, temperature sensors were mounted on snowmobiles (at about 1 m height) and horizontal transects were made using this setup at semi-regular intervals.

3. Evaluation of the operational AROME-Arctic

Before the comparison of AS05 and AA25 is presented in the next sections, we discuss some of the strengths and weaknesses of the operational AROME-Arctic for a winter period (December 2017 to February 2018). As explained, the AA25 is based on the same setup as the operational AROME-Arctic model, hence the results presented for an entire winter from AROME-Arctic is representative for the forecast capabilities of AA25. Here, AROME-Arctic is verified against the six permanent observation sites around Longyearbyen (Table 2).

The observed day-to-day variations in 2-m temperature are well captured by the model (Fig. 3a), with a temporal correlation of 0.95. AROME-Arctic has a cold bias (-1.3°C) for the period originating from a small, but relatively consistent underestimation enhanced by episodic substantial drops in the forecasted temperature. On the other hand, AROME-Arctic fails to identify days with increased spatial variability in 2-m temperature (Fig. 3b). The spatial correlation between forecasted and observed temperatures is high most of the time (on average 0.67), but with substantial drops for shorter periods (Fig. 3c). A closer investigation of these drops (not shown) reveals that they occur most frequently under conditions with strong static stability, as estimated using the difference in observed temperature between Adventdalen (15 m) and Gruvefjellet (464 m). For example, the average spatial correlation is 0.82 when Adventdalen is more than 1°C warmer than Gruvefjellet, while it is -0.01 when Gruvefjellet is more than 1°C warmer than Adventdalen. Hence, even if AROME-Arctic captures some of the local spatial variability of temperature (i.e. which sites are warmer/colder in the neighbourhood), it is unable to accurately predict the local spatial variability in the presence of strong static stability. Therefore, the amplitude of these local

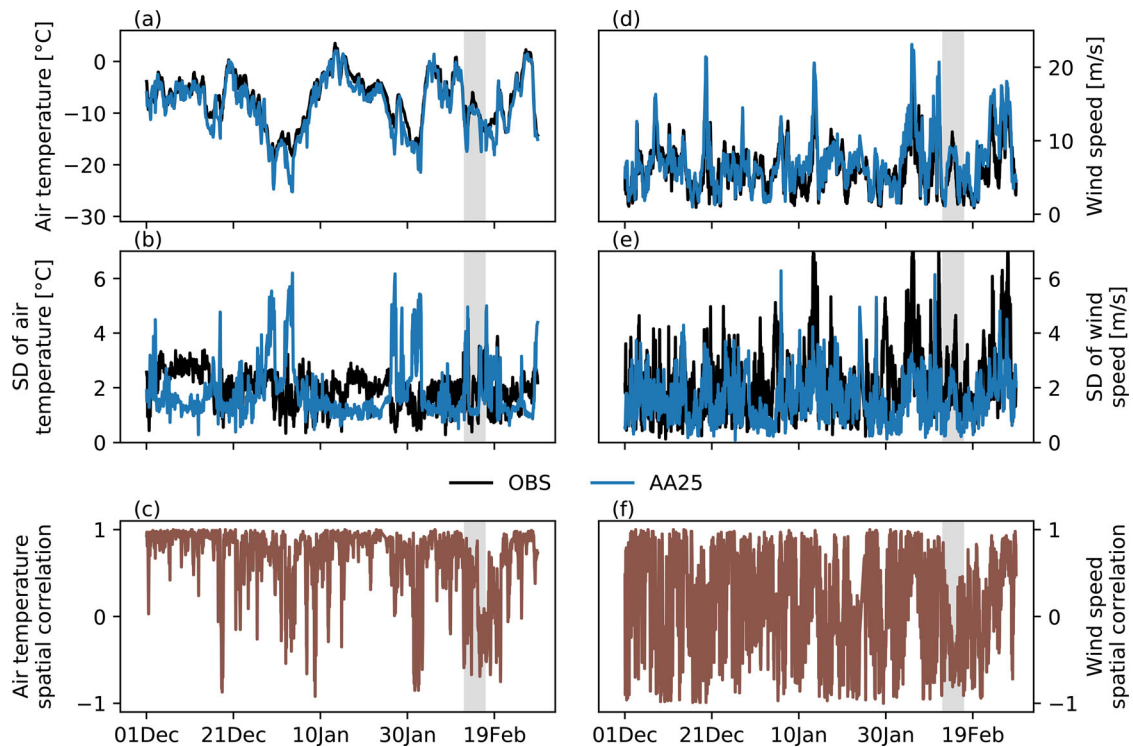


Fig. 3. Time series for DJF 2017/18 for temperature (left panels) and wind speed (right panels) for the 6 permanent stations. (a) Mean observed (black) and forecasted (blue) values of the two metre air temperature; (b) standard deviation of two metre temperature representing the spatial variability; (c) correlation between forecasts and observations; (d–f) same as (a–c), but for 10-m wind speed. The period for the detailed investigation in this study is highlighted (grey shading).

variations and which days are prone to enhanced local variability is not well forecasted (Fig. 3c).

Also for the 10-m wind speed the day-to-day variations are well forecasted by AROME-Arctic (Fig. 3d), with a temporal correlation of 0.81. However, AROME-Arctic has a positive mean error (~ 1.0 m/s), which is especially pronounced during a few high wind speed events. While AROME-Arctic is able to better identify days with high spatial variability in wind speed (Fig. 3e) than in 2-m temperature, the actual spatial correlation with observations is lower (on average 0.14, Fig. 3f), compared to 2-m temperature.

In summary, the operational AROME-Arctic is able to forecast the synoptic-scale forced temporal changes in the near-surface wind speed and temperature in the Adventdalen area. Earlier studies also showed that AROME-Arctic is competitive with other state-of-the-art NWP systems in the Arctic (e.g. Koltzow et al., 2019). However, weaknesses regarding the local spatial variability of temperature and wind speed are present. Having in mind the width of Adventdalen (~ 4 km) which is not resolved by 2.5 km grid spacing, this may be improved by higher-resolution forecasts. The time period for the rest of this study (marked with grey shading in Fig. 3) is

representative of the documented weaknesses in AROME-Arctic; the model performs relatively poorly with respect to reproducing the observed spatial variability in both temperature and wind.

4. Results from sub-kilometre simulations

In this section the sub-kilometre experiment, AS05, is validated against observations obtained during the field campaign between 12 and 16 February 2018, and compared to the model experiment with 2.5 km grid spacing, AA25. First, we show verification scores for all weather stations and the time series of near-surface meteorological variables at the Adventdalen (15 m) and Gruvefjellet (464 m) weather stations. Thereafter, two distinct weather situations are investigated in more detail. These are dominated by wind channelling through the Adventdalen valley on 14 February and a cold-air pool formation on 15–16 February.

4.1. Evaluation of surface parameters and surface energy budget

Both model experiments reproduced well the synoptic-scale conditions and the relative humidity follows the

Table 3. Mean error (ME) and standard deviation of error (SDE) between the observations and the modelled values averaged at weather stations for the forecast lead times from 1 to 24 hours.

Station	Exp	Temperature (°C)		Wind speed (m/s)	
		ME	SDE	ME	SDE
Vestpynten	AA25	2.23	3.00	0.99	2.12
	AS05	2.05	2.56	-0.16	2.36
Adventdalen	AA25	0.25	4.45	-1.09	3.38
	AS05	2.62	4.29	-0.26	2.67
Svalbard Airport	AA25	0.38	3.29	1.72	2.14
	AS05	1.50	2.54	0.65	2.04
Hobo	AA25	-0.58	3.53	5.25	2.99
	AS05	0.97	2.90	3.78	3.71
Endalen	AA25	-1.67	4.11	-0.38	2.74
	AS05	1.87	3.68	-1.28	1.99
Low elevation stations	AA25	0.16	3.94	1.07	3.42
	AS05	1.85	3.32	0.36	3.03
Janssonhaugen	AA25	-2.70	2.82	-0.72	2.76
	AS05	-0.41	1.43	-1.04	3.28
Platåberget	AA25	-1.03	2.57	2.86	2.71
	AS05	-2.17	2.03	-0.33	1.97
Gruvefjellet	AA25	-1.88	2.21	0.55	2.47
	AS05	-0.16	2.20	2.19	2.51
Breinosa	AA25	-2.82	2.97	3.91	2.37
	AS05	-1.70	2.99	1.36	2.26
High elevation stations	AA25	-2.11	2.75	1.65	3.16
	AS05	-1.11	2.38	0.55	2.85
All stations	AA25	-0.88	3.62	1.34	3.31
	AS05	0.49	3.28	0.44	2.95

Values marked with bold indicate significantly smaller absolute ME and SDE. The 95% confidence interval was calculated by bootstrapping.

temperature evolution (not shown). We therefore focus on the differences in near-surface temperature and wind speed, and understanding their errors, in the two forecast experiments.

Table 3 summarises the error statistics of AA25 and AS05 for all included weather stations. Even though there are exceptions on the level of each individual station, averaged over all stations, AS05 has significantly smaller mean error (ME) and standard deviation of error (SDE), both for 2-m temperature and 10-m wind speed than AA25. Distinguishing between average ME and SDE for low elevation stations (Vest, Adv, Air, Hobo, End) and high elevations stations (Jans, Plat, Gruv, Brein), gives the same conclusions. An exception is the average ME for 2-m temperature at the low elevation stations, which is better for AA25 than AS05, and this will be discussed later.

Figures 4 and 5 present time series of the measured and simulated near-surface temperature and wind speed

at the Adventdalen (lower elevation) and Gruvefjellet (higher elevation) weather stations. For the Adventdalen weather station, both AA25 and AS05 overestimate 2-m temperature, but this is more pronounced in AS05 (Table 3). As illustrated in Fig. 4a, there is a particularly strong warm bias (on average 7°C) in AS05 from the latter half of 15 February and onwards (cold-pool formation situation). Both AA25 and AS05 underestimate the 10-m wind speed (Table 3). The smallest underestimation is seen in AS05, which is largely due to a smaller (positive) bias in the period between about 13 and 15 February (wind channelling situation, Fig. 4c).

For the Gruvefjellet weather station, the underestimation of 2-m temperature is smaller for AS05 than for AA25, and the 2-m temperature ME at this location is the smallest of all stations (Table 3). Temperature errors at the Gruvefjellet station in AA25 are related to cold biases in the latter half of 12 and of 15 February (Fig. 5a). Both AS05 and AA25 do not accurately reproduce the difference in 2-m temperature between the Gruvefjellet and Adventdalen stations under conditions with stable static stratification (negative difference) as seen in Fig. 5c. However, AS05 shows a moderate improvement compared to AA25 in resolving the temperature difference between the two stations, e.g. around midnight on 13 February and in the latter half of 15 February. In contrast, AA25 has the smallest ME of the two model experiments for the 10-m wind speed at Gruvefjellet (Table 3). The time series of observed and simulated 10-m wind speed reveal that this is related to a consistent (stronger) positive wind speed bias throughout most of the study period in AS05. However, this is an exception, as AS05 has on average a less pronounced positive wind bias than AA25 for the higher elevated sites.

Surface-layer temperatures are mainly driven by the surface energy budget (SEB). We evaluate the SEB simulations with observations from the Adventdalen weather station. We first investigate the driving terms of the SEB following the ideas of Miller et al. (2018) and Day et al. (2020), i.e. investigate those terms that are not directly influenced by the surface properties, shortwave and longwave downward radiation. Shortwave radiation is negligible in mid-February in Svalbard and, therefore, longwave radiation unsurprisingly dominates the forcing (Fig. 6). Both AS05 and AA25 have relatively large negative biases in downwelling longwave radiation (Table 4). Downwelling radiation (Fig. 6a) is strongly affected by cloud cover, in particular low clouds. By considering the cloud fractions, observed at Svalbard airport, presented in Fig. 7, we see that the periods with the strongest downwelling radiation biases (around 12 and 14

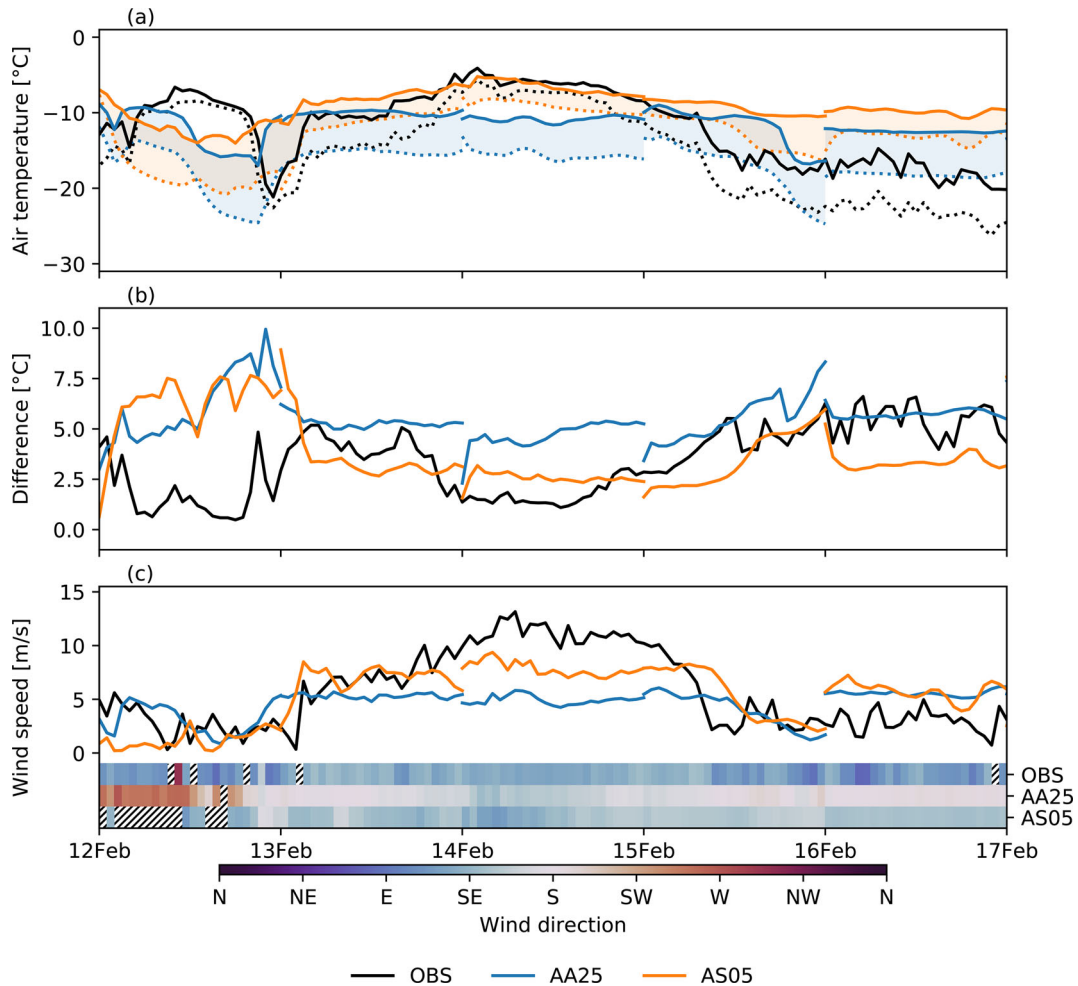


Fig. 4. Time series of simulated and observed (a) 2-m air temperature (solid lines) and surface temperature (dotted lines); (b) difference between surface and 2-m air temperature; and (c) 10-m wind speed from the Adventdalen weather station. Wind direction is indicated in (c) using coloured shading. Hatched areas indicate wind speed less than 1 m/s.

February), coincide with the occurrence of a strong negative cloud fraction bias in both simulations.

The response terms (i.e. the terms depending on the state of the surface) in the SEB are dominated by the upwelling longwave radiation and sensible heat flux (Fig. 6). Considering upwelling longwave radiation (Fig. 6b), the model errors vary in sign throughout the period and display a correlation with surface temperature errors (0.72 in AA25, 0.51 in AS05). Sensible heat flux was observed to be positive (downwards) during most of the study period (Fig. 4d). Despite the underestimation of wind speed at this station, both model experiments overestimate this flux on average (Table 4). Following Tjernström et al (2005), Fig. 8a shows sensible heat flux divided by 10-m wind speed against temperature difference between 2m and the surface. The sensible heat flux is proportional to both the wind speed and the near-surface temperature gradient in the bulk-flux formulation

which the model system makes use of, and thus, the slope represents the heat transfer coefficient in Fig. 8a. In observations, the wind-scaled heat flux has close to linear dependence on the temperature difference in weakly stable conditions when the temperature difference is small. Figure 8a suggests that both model experiments have the heat transfer coefficient close to the observed coefficient of weakly stable conditions regardless of the temperature difference in the model experiments. This indicates that the positive turbulent flux bias is due to overestimation of the transfer coefficient in very stable conditions. The overestimation can be partly explained by the surface roughness length in the model being approximately one order of magnitude larger than determined from the sonic measurements, but it is unknown how representative the observation point is for a model grid box. In addition, the static stability in the surface layer is often overestimated. The temperature difference between 2-m and the

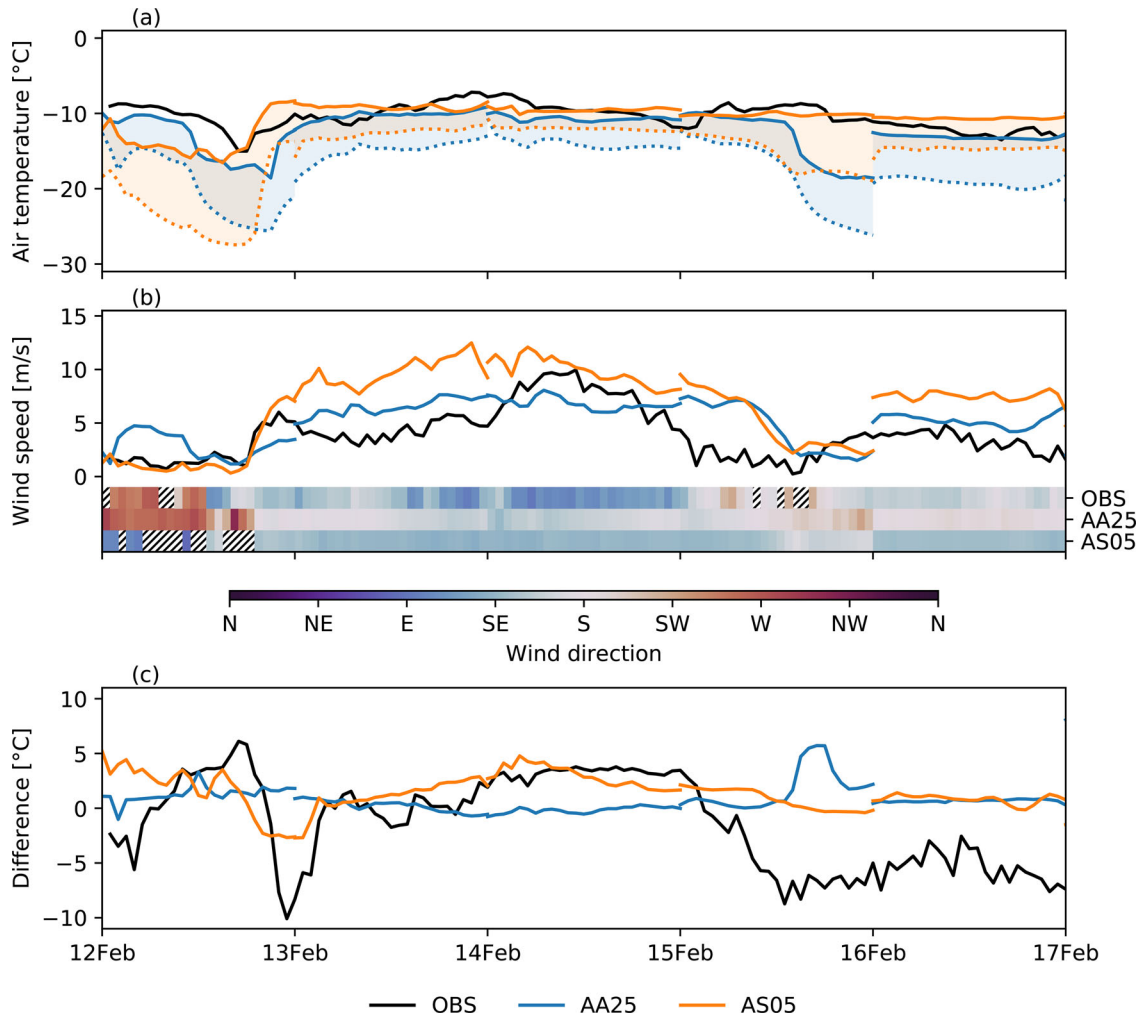


Fig. 5. Time series of (a) simulated and observed 2-m air temperature (solid lines) and simulated surface temperature (dotted lines); (b) simulated and observed 10-m wind speed from the Gruvefjellet weather station; and (c) simulated and observed difference between 2-m temperature at Adventdalen and Gruvefjellet. Observed surface temperature was not available for Gruvefjellet. Also on (b) the observed and modelled wind direction is shown. Hatched areas indicate wind speed less than 1 m/s.

surface is overestimated on average by 2.1°C in AA25 and by 0.4°C in AS05 (Fig. 4b). The observed latent heat flux (Fig. 6e) is small compared to the sensible heat flux and varied in sign during the campaign, both in the observations and in the simulations.

To study the sensitivity on how the surface responds to the net energy flux we show surface temperature and 2-m temperature against the net energy flux in Fig. 8b and c. These temperature sensitivity diagnostics provide information about the thermal inertia of the surface and indirectly also about the coupling between the surface and atmosphere. In AA25, the overall surface temperature sensitivity to the net energy flux is too low ($0.01^{\circ}\text{C}/\text{Wm}^{-2}$ compared to $0.24^{\circ}\text{C}/\text{Wm}^{-2}$ of observations, based on the regression line slopes in Fig. 8b and c), whereas in AS05, the surface temperature sensitivity

($0.19^{\circ}\text{C}/\text{Wm}^{-2}$) is more comparable to the observations. At 2-m, temperature sensitivity to the net energy flux is slightly lower because the 2-m temperature is diagnosed not only with the help of surface temperature but the lowest model level. AS05 has the 2-m temperature sensitivity ($0.11^{\circ}\text{C}/\text{Wm}^{-2}$) closer to the observed value ($0.18^{\circ}\text{C}/\text{Wm}^{-2}$) than AA25 ($0.08^{\circ}\text{C}/\text{Wm}^{-2}$). In AS05, the temperature sensitivities are higher at lower elevations than at higher elevations, while AA25 does not show this pattern (not shown). This behaviour can be interpreted as a more realistic physical response because the higher elevated sites experience a more rapid exchange of air masses and are less dominated by local processes working over time.

In general, AS05 does not show clear improvements in simulating the surface energy budget components

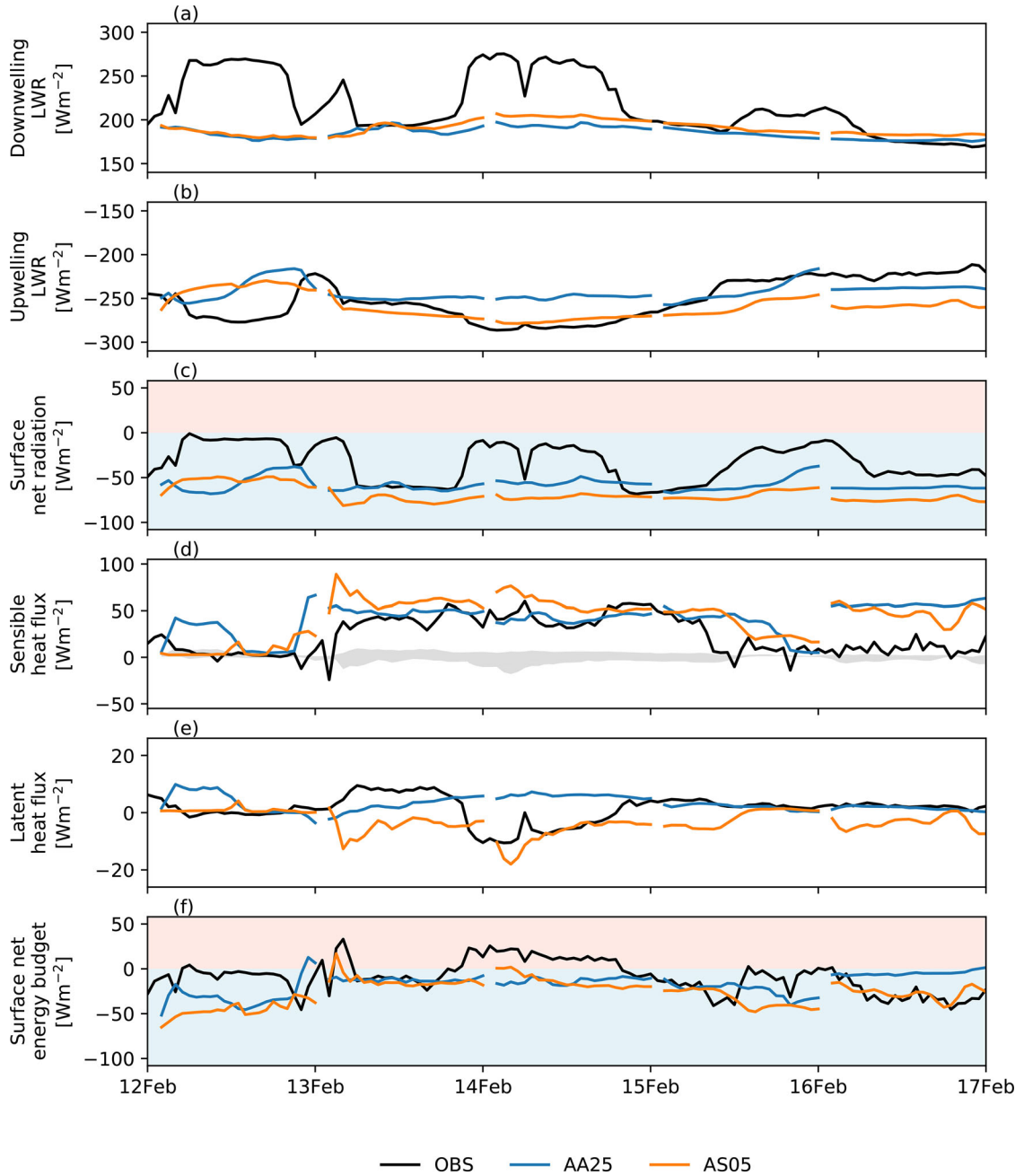


Fig. 6. (a) Downwelling longwave radiation flux at the surface; (b) upwelling longwave radiation flux emitted by the surface; (c) net radiation flux at the surface including the contributions from both shortwave and longwave radiation fluxes; (d) sensible heat flux from surface to atmosphere; (e) latent heat flux from surface to atmosphere (note the different scale on y-axis); (f) surface net energy budget at the Adventdalen weather station. Also on (d) the latent heat flux from the panel (e) is outlined to scale. All the fluxes are downwards-positive.

compared to AA25 (Table 4). The net surface energy budget (Fig. 6f), defined as a sum of surface net radiation, sensible heat flux and latent heat flux, has both negative and positive values in the observations, whereas both model experiments simulate negative values through virtually the whole study period. This is because both

model experiments fail to capture the cloudy periods. The temperature sensitivity diagnostics, however, suggest that the sub-kilometre system improves the sensitivity of the near-surface temperatures to the net energy flux even though there are no changes in the surface description of the model.

The stronger underestimation of the surface net energy in AS05 than in AA25 for Adventdalen (Table 4) contrasts a substantially warm bias seen in 2-m temperature in AS05 (Table 3). It is likely that part of this can be explained by the force-restore method employed where the surface temperature evolves not only due to the forcing by the surface energy budget, but also due to a restoring term towards deep ground temperature (Boone et al., 1999). During the studied

Table 4. Mean error (ME) and standard deviation of error (SDE) between the observations and the modelled values for components of the surface energy budget at the Adventdalen weather station.

Variable	Exp	ME	SDE
LWR down	AA25	-32.7	32.1
	AS05	-27.0	32.2
LWR up	AA25	8.4	21.8
	AS05	-7.6	23.6
LWR net	AA25	-24.2	20.1
	AS05	-34.7	17.7
Sensible heat flux	AA25	18.5	22.2
	AS05	19.1	18.2
Latent heat flux	AA25	1.7	6.1
	AS05	-4.6	5.1
Surface net energy	AA25	-4.0	22.7
	AS05	-20.1	18.8

All units in W/m^2 . Values marked with bold indicate significantly smaller absolute ME and SDE. The 95% confidence interval was calculated by bootstrapping.

period this restoring term warms, on average, the surface layer more in AS05 (2.6°C/h) than in AA25 (1.4°C/h). This is partly due to a slightly warmer initial deep layer temperature in the beginning of the studied period originating from warmer model forecasts during the model spin-up. In addition, since AS05 captures the relatively warm air temperatures in the middle of the period better, AS05 builds up an energy reservoir in the ground from where excessive energy is released at the time of the cooling in the last part of the period. This shows the importance of describing the ground processes adequately and highlights the need for developing the surface model at the same time as moving towards higher resolution. In addition, the surface assimilation scheme behaves differently in the two experiments in terms of producing different surface temperature increments (not shown), which also affects the evolution of the surface temperature during the period studied.

In order to investigate the difference between the two model experiments in more detail, two distinct and rather persistent time periods are analysed: First, a period of wind channelling through the Adventdalen valley on 14 February. Secondly, the formation of a cold-air pool in the Adventdalen valley on 15 February.

4.2. Wind channelling through the Adventdalen valley

The highest temperatures and strongest winds during the campaign were measured on 14th February, at the Hobo

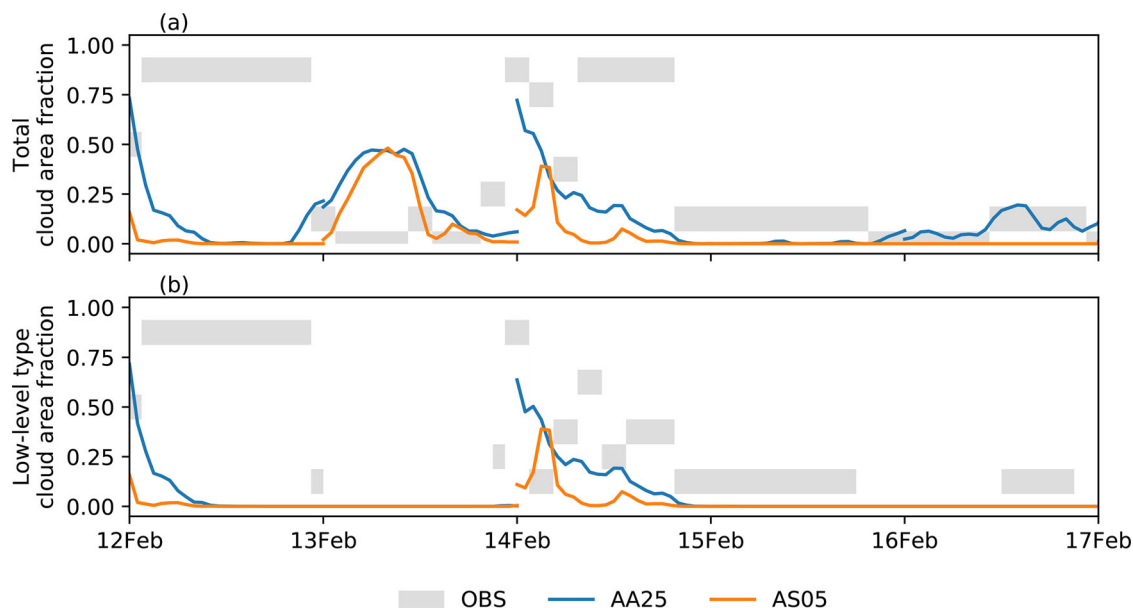


Fig. 7. Time series of (a) total cloud cover and (b) low-level cloud cover as observed at Svalbard Airport (grey shading) and in the AA25 and AS05 experiments (colors).

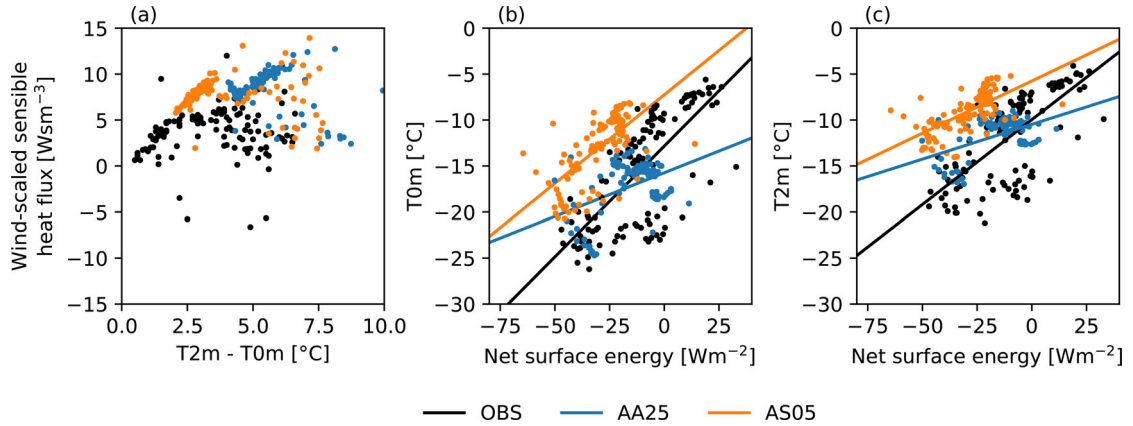


Fig. 8. (a) Sensible heat flux divided by the 10-m wind speed against temperature difference between 2m and surface, (b) surface temperature, and (c) 2-m temperature against the net energy ($LW_{net} + SH + LE$) at the Adventdalen station in observations and in the AA25 and AS05 experiments. Least-squares regression lines are included in (b) and (c).

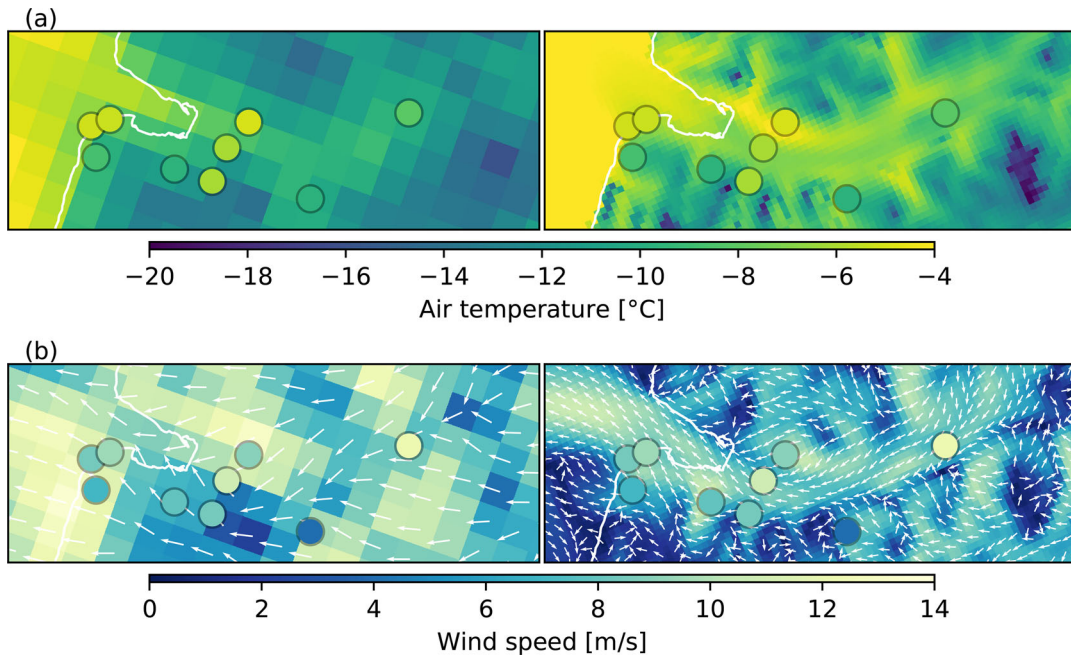


Fig. 9. (a) Two-metre temperature and (b) 10-m wind speed and wind direction in AA25 (left column) and AS05 (right column) on 14 February at 12 UTC. For AS05, only the wind direction in each second grid cell is presented. The white contour shows the coastline according to the local topography by the Norwegian Polar Institute (see Fig. 1). Superimposed (filled circles) are in-situ observations of 2-m temperature and 10-m wind speed.

(-1.0°C) and Janssonhaugen (14.7 m/s) weather stations. The lower elevated stations along the Adventdalen valley (Vest, Air, Adv, Hobo, Jans, see Fig. 1) measured on average 1 m/s stronger winds and 3°C higher temperatures than the nearby sites at higher elevation (Plat, Gruv, End, Brein) on 14 February, as illustrated in Fig. 9. This indicates intensified wind channelling through the Adventdalen valley.

The spatial differences in temperature and wind between the two experiments are striking (Fig. 9). AS05 matches the indicated near-surface observations at most measurement stations far better than AA25. The mean absolute error of wind speed for all stations is 2.2 m/s for AS05 and 3.0 m/s for AA25 on 14 February. As in the observations, AS05 shows clear signs of wind channelling through the valley (Fig. 9b). For example, AS05 captures

the mean wind speed difference between the valley bottom in Adventdalen and the highest located station Breinosa on 14 February (3.0 m/s compared to 3.2 m/s in observations). In contrast, the channelling is only recognised to a small extent in AA25, and the mean wind speed difference between Adventdalen and Breinosa is even wrong in sign in AA25 (−1.9 m/s). This reveals that AS05 effectively resolves the topography of the Adventdalen valley, which is not the case for AA25. Another feature simulated by AS05 is increased wind speed close to and on downstream mountain slopes. The occurrence of this phenomenon is supported by the intensified flow observed at the Gruvefjellet weather station, located close to a mountain slope. On the other hand, both AS05 and AA25 simulate easterly downslope wind intensification for the Hobo weather station, where in fact weak westerly wind was measured, which is an indication for a recirculation in the Adventdalen valley. In addition, at the entrance of Adventdalen (Vest and Air) the simulated wind direction of AS05 differs by approximately 30° from the observations (not shown in figures). These examples imply that despite the remarkable improvements in capturing the topographic impact, an NWP model with increased spatial resolution (0.5 km grid spacing) does not necessarily fully resolve all topographic wind effects in complex terrain.

The temperature field simulated by AS05 highly agrees with the observations from weather stations for this situation (Fig. 9a). The warmest temperatures occur at the valley bottom, especially at the northern side of the valley, and lower temperatures occur at higher elevation sites. AA25 simulates too low temperatures for the sites in the valley.

In order to measure the spatial variability of the temperature in the valley, a transect of the near-surface temperature was performed with a snowmobile on 14 February (Fig. 10a). Measurements show noticeable spatial variability with the temperature varying between −8 and −4 °C in the Adventdalen valley with increased temperatures around the Hobo station. AS05 is in this situation, compared to AA25, in much better agreement with the snowmobile measurements (Fig. 10c) and shows higher spatial correlation with observations (0.63 compared to 0.04) and smaller bias (−0.2 °C compared to −2.7 °C). The results suggest that topographic effects play a main role here also for the temperature field in a situation of wind channelling, possibly through enhanced vertical mixing of the air mass in the valley

4.3. Cold-air pool in the Adventdalen valley

On 15 February 2018 a cold-air pool developed in the Adventdalen valley under nearly clear sky conditions.

The wind speed dropped from 8 m/s to below 3 m/s and the 2-m temperatures from −10 to −15 °C between 06:00 and 12:00 UTC at the Adventdalen weather station (Fig. 4a) while the high elevation stations measured higher temperatures (Fig. 11). The maximum temperature difference between Adventdalen station (at 15 m elevation) and Gruvefjellet (464 m), which are less than 5 km apart in horizontal distance, was 9 °C on this day (Fig. 5c). After the aforementioned rapid temperature drop in the Adventdalen valley, temperature decreased gradually both at the valley bottom and higher elevations until the end of the campaign, without any major warming.

The observed net radiation, mainly driven by the clear sky conditions, varied from −70 to −10 W/m² at the Adventdalen weather station on 15 February (Fig. 6c). As the observed sensible heat flux decreased from 50 W/m² to close to zero at the time of the wind speed decrease on 15 February (Fig. 6d), the net surface energy budget (Fig. 6f) remained negative the whole day, having its minimum at 06–12:00 UTC when the temperature drop was steepest. This led to cooling of the surface and the near-surface air creating a stable surface layer with a strong temperature inversion.

Both model experiments capture the wind weakening on 15 February, though hours later than observed (Fig. 4c). AA25 simulates the steep temperature drop around 18:00 UTC, approximately 12 hours later than observed (Fig. 4a), and wrongly predicts a temperature drop at the high-elevated stations as well. Therefore, the temperature difference between Adventdalen and Gruvefjellet in AA25 does not resemble the observations (Fig. 5c). AS05, instead, creates a small temperature decrease in the valley around 12:00 UTC at the time of the wind speed decrease and correctly keeps the temperature nearly constant for the higher elevation stations (e.g. Gruvefjellet, Fig. 5a). The 2-m temperature decrease at the Adventdalen station (Fig. 4a) is, however, substantially underestimated in AS05, which results in a strong positive bias of the forecast.

Also on 15 February a transect of the near-surface temperature was performed with a snowmobile to measure the spatial variability (Fig. 10b and d). Unlike the day before, AS05 does not show improved spatial correlations compared to AA25. The spatial correlations between forecast and observations are 0.05 and −0.01 for AS05 and AA25, respectively. In addition, a considerable positive bias is present in both AS05 (+5.8 °C) and AA25 (+4.0 °C).

Both AS05 and AA25 overestimated the outgoing longwave radiation noticeably at the time of the cold-air pool generation (Fig. 6b). Simultaneously, the sensible heat flux was overestimated in both experiments and the corresponding error compensation led to a net surface

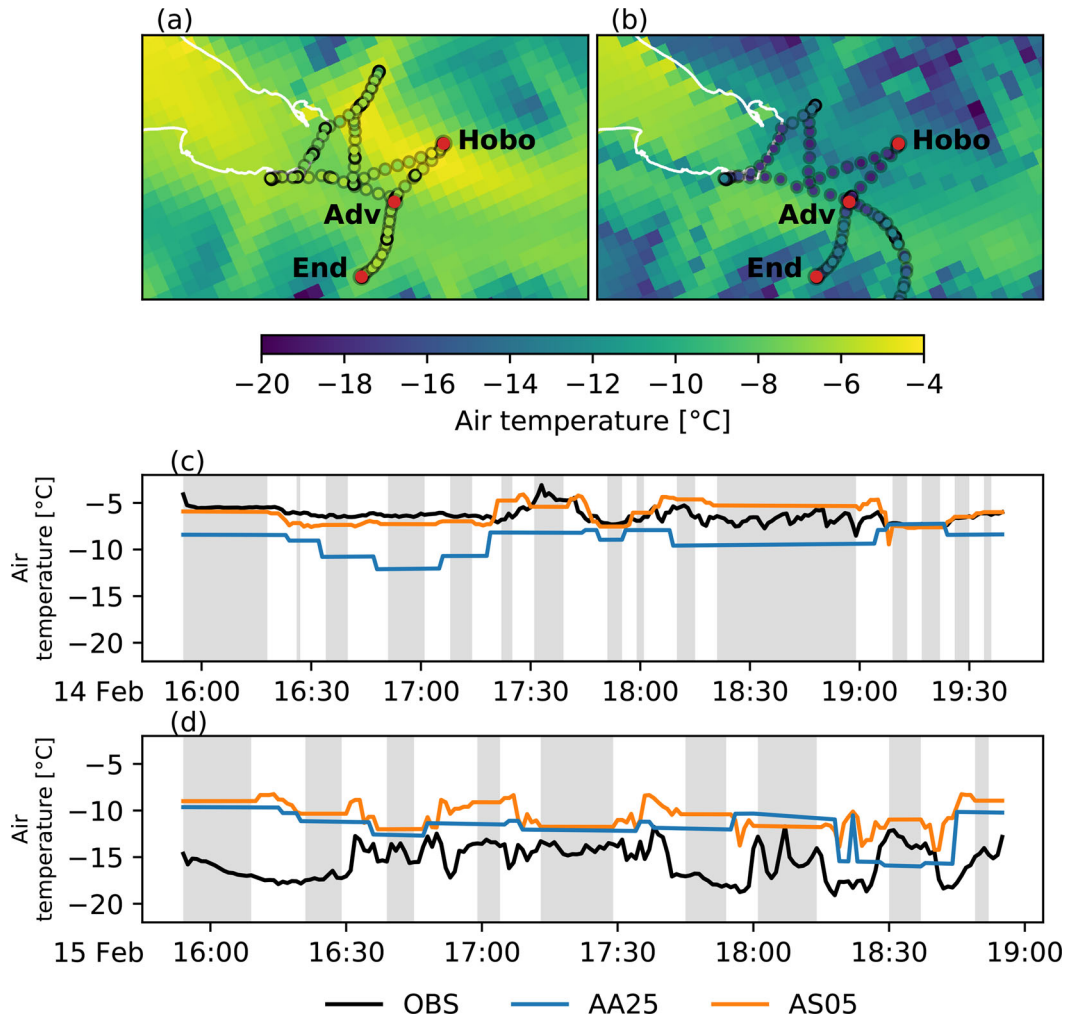


Fig. 10. Two-metre temperature field from AS05 compared to measurements from snowmobile. (a) AS05 forecast from 14 February 18 UTC and air temperature measured along the snowmobile track; (b) same as (a) but for 15 February 18 UTC; (c) time series of snowmobile measurements shown in the panel (a) compared to AS05 forecast, shaded areas correspond to stops of the snowmobile; (d) same as (c), but for the series of snowmobile measurements shown in the panel (b).

energy budget which is comparable to observations on average. However, the experiments do not capture the temporal changes in the surface energy budget. The delay in forecasting the wind speed decrease in the model experiments explains some of the temporal difference. In addition, as mentioned earlier, the restoring term towards deep ground temperature in the surface model is particularly high in AS05 at the time of the cold-air pool generation which leads to only a small surface temperature drop in AS05.

Another factor influencing cold-air pool formation and persistence is known to be downvalley drainage of the cold air. At the Endalen weather station, located in one of the side valleys of Adventdalen, wind blew down the valley perpendicular to the wind direction in Adventdalen

with a strength of 3 m/s during the cold pool event. The model simulations do not capture the drainage flow in Endalen. AA25 does not represent Endalen at all and AS05 contains only 3–4 grid cells across the valley, which is below the effective resolution of the model simulation, generally around 7 times the grid cell length (Skamarock, 2004).

On 16 February both experiments were too warm and windy in Adventdalen from the beginning of the forecasts (Fig. 4a and b). The reason for the offset is related to a different development between the cycling of 3-h model forecasts during 15 February, which is used as a model first guess field for the analysis at 16 February 00:00, and the 24-h forecast initialised at 15 February 00:00 shown in Fig. 4. The temperature and wind forecast on 16

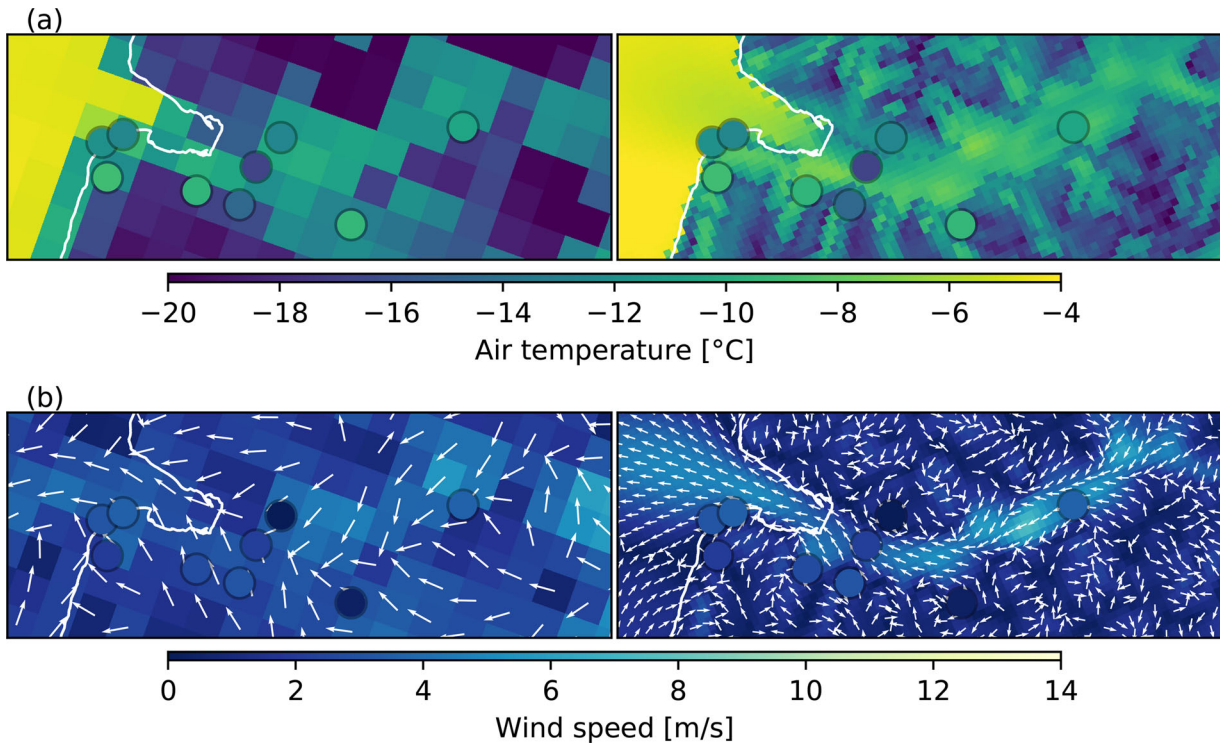


Fig. 11. As Fig. 9 but for 15 February 18:00 UTC.

January do not show any improvements during the duration of the forecast. This emphasises the importance of a good initial state for a successful forecast.

Vertical profiles were retrieved from a tethersonde at the Adventdalen weather station on 16 February between 17:34 and 19:12 UTC (Fig. 12). The temperature profile reveals a strong inversion near the surface with the temperature increasing from about -23°C at the surface to -13°C at 50 m height (Fig. 12a). Both model experiments create a shallow temperature inversion near the surface, but AA25 and AS05 overestimate temperatures by 5 and 9°C at the surface and both by 4°C at 50 m height compared to observations. AS05 partly reproduces the shape of the temperature inversion at the lowest model levels, however, the vertical temperature gradient is underestimated. A vertical cross section of potential temperature across the Adventdalen valley from AS05 (Fig. 13b) illustrates that the model simulates the cold pool at the valley bottom, but possibly displaced and with overestimated temperatures. Unlike AS05, AA25 lacks a clear indication of the surface-based temperature inversion at the model levels (Fig. 12a). Instead, the surface scheme computes an excessive temperature inversion between the surface and the lowest model level. Even though AA25 captures the surface and 2-m temperature closer to the observations than AS05 at this time and location, it is evident that the unrealistic temperature profile between the surface and

the lowest model level improves the 2-m air temperature forecasts for the wrong reason.

10-m wind speeds of around 2 m/s were observed at the Adventdalen weather station (Fig. 12b). On top of the shallow inversion, at about 70 m height, a low-level jet was present with wind speeds of about 5 m/s. AS05 captures a realistic shape for the wind profile in the lowest 500 m, including the low-level jet, but underestimates the wind speed by up to 2 m/s above 500 m. AA25, in comparison, represents the vertical profile of wind poorly in the lowest 300 m and does not simulate any low-level jet. Above 300 m, however, AA25 matches the observed wind speed very well. The vertical cross section through Adventdalen (Fig. 13d) shows that the relatively strong, channelled flow, including the low-level jet, was banked up against the southern side of the Adventdalen valley. It was strongest between 50 and 150 m above the surface, with wind speeds of about 8 m/s. In AS05 the near-surface temperature is lowest at the northern side of Adventdalen, where the low-level jet is absent. This indicates that the jet causes too much vertical mixing in the temperature.

5. Discussion and conclusions

Terrain-induced atmospheric features pose challenges for prediction of near-surface atmospheric conditions, even in

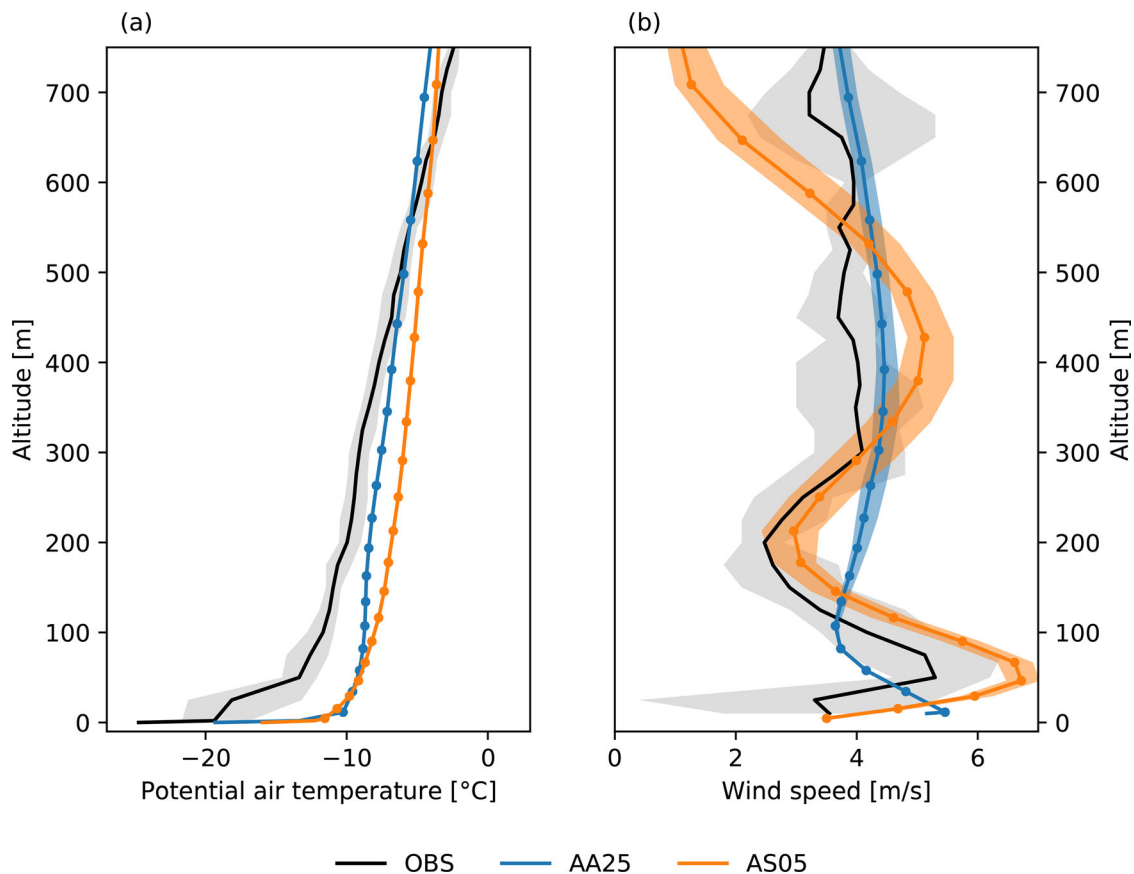


Fig. 12. Observed (tethersonde) and modelled (AA25 and AS05) vertical profiles of (a) potential temperature and (b) wind speed at the Adventdalen weather station. The tethersonde data are from two profiles, each with two sensors, obtained on 16 February between 17:34 and 19:12 UTC. The model data are from 17:00 to 19:00 UTC. Solid lines indicate the average of the profiles and the range is shown with shading. The potential temperature profiles have been extended to surface by including 2-m and surface temperature and the AA25 wind speed profile has been extended by including the 10-m data.

kilometre-scale NWP models. In this study, the ALADIN-HIRLAM NWP system with 0.5 km horizontal grid spacing and an increased number of vertical levels was compared to the 2.5-km model system similar to the current operational NWP system at MET Norway. The impact of the increased resolution on the model's ability to represent and realistically simulate boundary-layer processes was investigated for the period from 12 to 16 February 2018 in an Arctic fjord-valley system in the Svalbard archipelago. Model simulations were compared to a wide range of observations conducted during a field campaign. The studied period is a part of the YOPP SOP-NH1 and this study is a contribution to YOPP.

The operational 2.5-km model system is able to forecast the synoptic-scale temporal changes in near-surface temperature and wind speed. It has earlier been shown that the operational model system performs well compared to other NWP systems with similar or coarser resolution for this area (e.g. Køltzow et al., 2019). However,

results presented here show that it does not resolve local topographic effects in the narrow Adventdalen valley and therefore is not able to accurately predict the local variability in the weather.

We found that the sub-kilometre experiment improved both the spatial structure and overall verification scores of the near-surface temperature and wind forecasts compared to the 2.5-km experiment. The sub-kilometre experiment successfully captured the wind channelling through the Adventdalen valley and the temperature field associated with it. The added value of higher resolution on wind speed is in line with previous kilometre-scale studies made in Svalbard (Kilpeläinen et al., 2011, 2012; Kim et al., 2019) and on sub-kilometre Arctic systems (Vionnet et al., 2015; Yang, 2019).

Despite the overall added value, the sub-kilometre model system does not resolve certain aspects. The mean error of 2-m temperature forecasts at low elevation stations was particularly high in the sub-kilometre

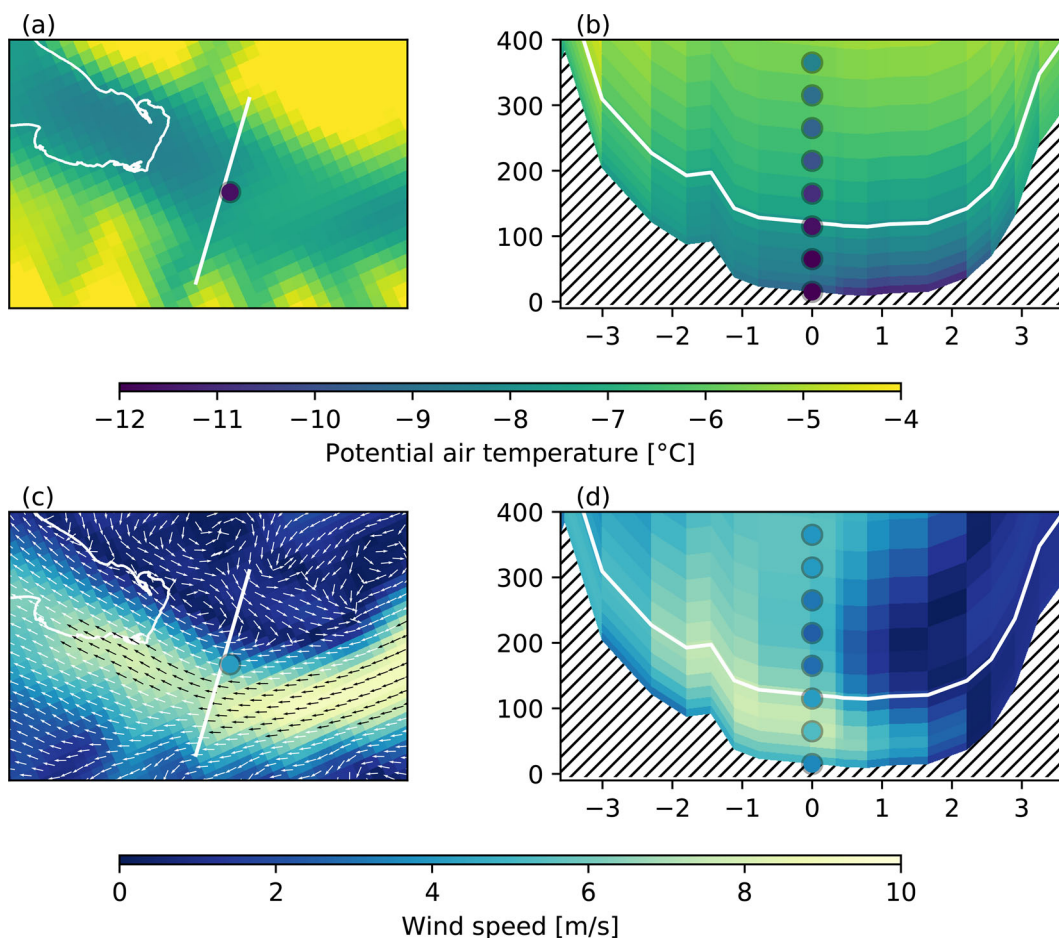


Fig. 13. Vertical cross-section of the AS05 potential air temperature and wind speed compared to the mean profile from tethered sondes. The tethered sonde data are from two profiles, each with two sensors, obtained on 16 February between 17:34 and 19:12 UTC. The model data are from 18:00 UTC. (a) Potential air temperature at 110 metre height above the model topography in AS05; (b) vertical cross-section of the potential air temperature in AS05 across the Adventdalen valley; (c) same as (a) but for wind speed; (d) same as (b) but for wind speed. Also on (a–d) observations from the tethered sonde are shown with filled circles.

experiment. Like Vionnet et al. (2015) for sub-kilometre experiments in the Canadian Arctic and Kilpeläinen et al. (2011, 2012), and Mayer et al. (2012) for kilometre-scale experiments at Svalbard, we found an inability of the sub-kilometre model to capture the intensity of the valley cold-air pool. However, the use of measurement campaign data provides some encouraging results. For example, the sub-kilometre system had a more realistic vertical temperature gradient and wind profile in the lowest part of the atmosphere, and the surface temperature sensitivity to the net surface energy was closer to the observations. It appears to be beneficial to have more model layers close to the surface to resolve the vertical profile of the atmosphere and the surface-atmosphere sensitivity properly. Mayer et al. (2012) also pointed out that increased vertical resolution potentially can lead to

improved representation of inertial-gravity waves under stable conditions.

Deficiencies in different parts of the model system, not necessarily related to the resolution, hamper the full potential of the sub-kilometre system. The way towards operational sub-kilometre simulations should therefore also address these issues. Large model biases were found in both driving and response terms of the surface energy budget. The downwelling longwave radiation had a large negative bias and was strongly affected by cloud cover and possibly cloud properties. The simulated surface turbulent fluxes were overestimated and further investigations are ongoing for the turbulent flux formulations in stable conditions.

Another factor limiting the forecast quality is simplistic treatment of the land surface and soil. The force-restore

method for the surface temperature evolution provides a simplified representation of the surface and soil processes including ground snow. The force-restore method might be a limitation for taking the full advantage of sub-kilometre resolutions especially for high latitude locations without diurnal forcing. In our results, it contributed substantially to a larger warm bias during a cold-pool event. Day et al. (2020) demonstrated that usage of a multi-layer snow scheme improves the temperature forecasts in two high-Arctic sites compared to a single-layer snow model. A recent version of the SURFEX model with the adequate description of soil, surface and snow processes might improve the sub-kilometre system and should be investigated.

Having accurate initial conditions is important in general and also reported on in a number of sub-kilometre and Svalbard kilometre-scale studies (Mayer et al., 2012; Hagelin et al., 2014; Vionnet et al., 2015; Kim et al., 2019). To avoid a long spin up in the simulations, it is required that the small scales represented in the sub-kilometre simulations are present from the start of the forecast. Similarly, an accurate initialization of the surface is a prerequisite for the NWP system to be able to forecast the highly local processes in the stable boundary layer important for the Arctic region. The cold-pool event discussed here is another example showing how initial surface properties potentially contribute to larger forecast errors.

A question for future studies is the impact of drainage flows on the development of the cold-air pools. Both model experiments were too warm in the lowest part of the atmosphere. This could be due to inadequate representation of vertical mixing or advection. Although the sub-kilometre experiment simulates down-sloping winds on some valley slopes, it does not capture the cold-air drainage in Endalen at the time of the cold-air pool, and the lack of drainage flows from side valleys might limit the model's ability to simulate cold-air pools accurately in Adventdalen.

As suggested by Casati et al. (2017), additional observations and process-based diagnostics have been important to better understand the differences between kilometre-scale and sub-kilometre-scale simulations. Our use of independent campaign observations, such as tether-sondes and snowmobile observations, drastically broadened the understanding of sub-kilometre model performance compared to the use of synoptic weather stations only.

Even though the studied period is short, our work demonstrates the potential of sub-kilometre NWP systems for forecasting weather in complex Arctic terrain. We show that the sub-kilometre system performs well compared to the operational 2.5-km system even when no

specific adjustments or adaptations have been made for the model system besides the improved horizontal and vertical resolution. However, our results also suggest that the increase in resolution should be accompanied with further development of other parts of the model system. The sub-kilometre model system for Svalbard is currently being further investigated for longer periods and for an optimal setup also considering the computational costs.

Acknowledgements

The study 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.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The study 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)' and the European Union's Horizon 2020 Research and Innovation programme through Grant 727862 APPLICATE (Advanced Prediction in Polar regions and beyond: modelling, observing system design and Linkages associated with a Changing Arctic climate).

References

- Accadia, C., Mariani, S., Casaioli, M., Lavagnini, A. and Speranza, A. 2003. Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Weather Forecast.* **18**, 918–932. > 2.0.CO;2. doi:10.1175/1520-0434(2003)018<0918:SOPFSS>2.0.CO;2
- Ataskin, E. and Vihma, T. 2012. Evaluation of NWP results for wintertime nocturnal boundary-layer temperatures over Europe and Finland. *Q. J. R. Meteorol. Soc.* **138**, 1440–1451. doi:10.1002/qj.1885
- Batrak, Y., Kourzeneva, E. and Homleid, M. 2018. Implementation of a simple thermodynamic sea ice scheme, SICE version 1.0-38h1, within the ALADIN-HIRLAM numerical weather prediction system version 38h1. *Geosci. Model Dev.* **11**, 3347–3368. doi:10.5194/gmd-11-3347-2018

- Bauer, P., Magnusson, L., Thépaut, J. and Hamill, T. M. 2016. Aspects of ECMWF model performance in polar areas. *Q. J. R. Meteorol. Soc.* **142**, 583–596. doi:10.1002/qj.2449
- Belamari, S. 2005. Report on uncertainty estimates of an optimal bulk formulation for surface turbulent fluxes, Marine Environment and Security for the European Area – Integrated Project (MERSEA IP), Deliverable D4.1.2, 29 pp.
- Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J. and co-authors. 2017. The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system. *Mon. Wea. Rev.* **145**, 1919–1935. doi:10.1175/MWR-D-16-0417.1
- Boone, A., Calvet, J. and Noilhan, J. 1999. Inclusion of a third soil layer in a land surface scheme using the force–restore method. *J. Appl. Meteor.* **38**, 1611–1630. > 2.0.CO;2. doi:10.1175/1520-0450(1999)038<1611:IOATSL>2.0.CO;2
- Casati, B., Haiden, T., Brown, B., Nurmi, P. and Lemieux, J.-F. 2017. Verification of environmental prediction in polar regions: Recommendations for the Year of Polar Prediction. WWRP 2017-1, WMO, 44 pp.
- Clements, C. B., Whiteman, C. D. and Horel, J. D. 2003. Cold-air-pool structure and evolution in a mountain Basin: Peter Sinks, Utah. *J. Appl. Meteor.* **42**, 752–768. > 2.0.CO;2. doi:10.1175/1520-0450(2003)042<0752:CSAEIA>2.0.CO;2
- Danielson, J. J. and Gesch, D. B. 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011-1073, 26 pp.
- Dawson, J., Hoke, W., Lamers, M., Liggett, D., Ljubicic, G. and co-authors. 2017. Navigating weather, water, ice and climate information for safe polar mobilities. WWRP/PPP No. 5. World Meteorological Organization.
- Day, J., Arduini, G., Sandu, I., Magnusson, L., Beljaars, A. and co-authors. 2020. Measuring the impact of a new snow model using surface energy budget process relationships. *Earth Space Sci. Open Arch.* **12**, e2020MS002144. doi:10.1029/2020MS002144
- Douville, H., Royer, J. and Mahfouf, J. 1995. A new snow parameterization for the Météo-France climate model. *Clim. Dyn.* **12**, 21–35.
- Esau, I., Tolstykh, M., Fadeev, R., Shashkin, V., Makhorylova, S. and co-authors. 2018. Systematic errors in northern Eurasian short-term weather forecasts induced by atmospheric boundary layer thickness. *Environ. Res. Lett.* **13**, 125009. doi:10.1088/1748-9326/aaecfb
- Hagelin, S., Auger, L., Brovelli, P. and Dupont, O. 2014. Nowcasting with the AROME model: first results from the high-resolution AROME airport. *Weather Forecast.* **29**, 773–787. doi:10.1175/WAF-D-13-00083.1
- Isaksen, K., Nordli, Ø., Førland, E. J., Łupikasza, E., Eastwood, S. and co-authors. 2016. Recent warming on Spitsbergen—influence of atmospheric circulation and sea ice cover. *J. Geophys. Res. Atmos.* **121**, 11,913–11,931.
- Kilpeläinen, T., Vihma, T., Lafsson, H. O. and Karlsson, P. E. 2011. Modelling of spatial variability and topographic effects over Arctic fjords in Svalbard. *Tellus A* **63**, 223–237. doi:10.1111/j.1600-0870.2010.00481.x
- Kilpeläinen, T., Vihma, T., Manninen, M., Sjöblom, A., Jakobson, E. and co-authors. 2012. Modelling the vertical structure of the atmospheric boundary layer over Arctic fjords in Svalbard. *Q. J. R. Meteorol. Soc.* **138**, 1867–1883. doi:10.1002/qj.1914
- Kim, D.-H., Kim, H. M. and Hong, J. 2019. Evaluation of wind forecasts over Svalbard using the high-resolution Polar WRF with 3DVAR. *Arct. Antarct. Alp. Res.* **51**, 471–489. doi:10.1080/15230430.2019.1676939
- Költzow, M., Casati, B., Bazile, E., Haiden, T. and Valkonen, T. 2019. An NWP model intercomparison of surface weather parameters in the European Arctic during the year of polar prediction special observing period Northern Hemisphere 1. *Weather Forecast.* **34**, 959–983. doi:10.1175/WAF-D-19-0003.1
- Litt, M., Sicart, J.-E. and Helgason, W. 2015. A study of turbulent fluxes and their measurement errors for different wind regimes over the tropical Zongo Glacier (16° S) during the dry season. *Atmos. Meas. Tech.* **8**, 3229–3250. doi:10.5194/amt-8-3229-2015
- Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A. and co-authors. 2013. 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. doi:10.5194/gmd-6-929-2013
- Mauder, M. and Foken, T. 2004. Documentation and instruction manual of the eddy covariance software package TK2. Universität Bayreuth, Abt. Mikrometeorologie, Print, ISSN 1614-8916, Arbeitsergebnisse, 26, 44 pp.
- Mayer, S., Jonassen, M. O., Sandvik, A. and Reuder, J. 2012. Profiling the Arctic stable boundary layer in Advent Valley, Svalbard: measurements and simulations. *Boundary-Layer Meteorol.* **143**, 507–526. doi:10.1007/s10546-012-9709-6
- Miller, N. B., Shupe, M. D., Lenaerts, J. T. M., Kay, J. E., de Boer, G. and co-authors. 2018. Process-based model evaluation using surface energy budget observations in central Greenland. *J. Geophys. Res. Atmos.* **123**, 4777–4796. doi:10.1029/2017JD027377
- Moore, G. W. K., Renfrew, I. A., Harden, B. E. and Mernild, S. H. 2015. The impact of resolution on the representation of southeast Greenland barrier winds and katabatic flows. *Geophys. Res. Lett.* **42**, 3011–3018. doi:10.1002/2015GL063550
- Müller, M., Batrak, Y., Kristiansen, J., Költzow, M. A., Noer, G. and co-authors. 2017. Characteristics of a convective-scale weather forecasting system for the European Arctic. *Mon. Wea. Rev.* **145**, 4771–4787. doi:10.1175/MWR-D-17-0194.1
- Nawri, N. and Stewart, R. E. 2008. Channelling of high-latitude boundary-layer flow. *Nonlin. Processes Geophys.* **15**, 33–52. doi:10.5194/npg-15-33-2008
- Norwegian Polar Institute. 2014. *Terrengmodell Svalbard (S0 Terrengmodell) [Data set]*. Norwegian Polar Institute, Tromsø, Norway.
- Oltmanns, M., Straneo, F., Seo, H. and Moore, G. W. 2015. The role of wave dynamics and small-scale topography for downslope wind events in Southeast Greenland. *J. Atmos. Sci.* **72**, 2786–2805. doi:10.1175/JAS-D-14-0257.1
- Sandu, I., Beljaars, A., Bechtold, P., Mauritsen, T. and Balsamo, G. 2013. Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models? *J. Adv. Model. Earth Syst.* **5**, 117–133. doi:10.1002/jame.20013

- Skamarock, W. C. 2004. Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.* **132**, 3019–3032. doi:10.1175/MWR2830.1
- Stigter, E. E., Litt, M., Steiner, J. F., Bonekamp, P. N. J., Shea, J. M. and co-authors. 2018. The importance of snow sublimation on a Himalayan Glacier. *Front. Earth Sci.* **6**, 108. doi:10.3389/feart.2018.00108
- Tjernström, M., Žagar, M., Svensson, G., Cassano, J. J., Pfeifer, S. and co-authors. 2005. Modelling the Arctic boundary layer: an evaluation of six arcmip regional-scale models using data from the Sheba project. *Boundary-Layer Meteorol.* **117**, 337–381. doi:10.1007/s10546-004-7954-z
- Vickers, D. and Mahrt, L. 1997. Quality control and flux sampling problems for tower and aircraft data. *J. Atmos. Oceanic Technol.* **14**, 512–526. > 2.0.CO;2. doi:10.1175/1520-0426(1997)014<0512:QCAFSP>2.0.CO;2
- Vionnet, V., Bélair, S., Girard, C. and Plante, A. 2015. Wintertime subkilometer numerical forecasts of near-surface variables in the Canadian Rocky Mountains. *Mon. Wea. Rev.* **143**, 666–686. doi:10.1175/MWR-D-14-00128.1
- Yang, X. 2019. TAS, an operational forecast model at hectometric scale. ALADIN-HIRLAM Newsletter. No 12, 42–49. Online at: <https://www.umr-cnrm.fr/aladin/IMG/pdf/nl12.pdf>.
- Yang, X., Palmason, B., Sattler, K., Thorsteinsson, S., Amstrup, B. and co-authors. 2018. IGB, the upgrade to the joint operational HARMONIE by DMI and IMO in 2018. ALADIN-HIRLAM Newsletter, No 11, 93–96. Online at: <https://www.umr-cnrm.fr/aladin/IMG/pdf/nl11.pdf>.