Norwegian
Meteorological
Institute

No. 9/2024
METreport
ISSN 2387-4201
Meteorology

Impact of improved physiography data on hectometric weather forecasts in Svalbard

Torbjørg Venstad, Yurii Batrak and Teresa Remes

Norwegian Meteorological Institute

METreport

Title	Date
Impact of improved physiography data on hectometric weather forecasts in Svalbard	March 26, 2024
Section	**Report no.**
Meteorology	9/2024
Author(s)	**Classification**
Torbjørg Venstad, Yurii Batrak and Teresa Remes	● Free ○ Restricted

Abstract

This report investigates the impact of enhanced physiography data on the performance of hectometric numerical weather prediction simulations over Svalbard. We examine three versions of physiography data: the older ECOCLIMAP2, ECOCLIMAP SG, and a manually corrected version of ECOCLIMAP SG, focusing on their effects on simulations of surface and near-surface atmospheric variables. This assessment includes direct comparisons of numerical experiments and validations against observational data from satellites and weather stations. The results highlight that modifications in physiography data have a pronounced effect on modelled temperatures, especially in areas with updated glacier and sea cover masks, and topographical changes. However, for model forecasts evaluated against in situ observations of 2m temperature, 10m wind speed, and 2m specific humidity, using the modified physiography data did not lead to significant improvements. Nonetheless, a modest improvement was found for surface temperature, likely due to the satellite observations available over a wider area affected by the updated physiography data. The study also highlighted the significant role of adjusting the turbulent exchange between the surface and the model atmosphere in stable conditions. The results suggest further research focusing on different seasons, and separating the effects of modifying topography and cover types, for clearer insights. The findings support continued use of the enhanced physiography dataset in the future.

Keywords

numerical weather prediction, Arctic

Disciplinary signature

Responsible signature

Norwegian Meteorological Institute
Org.no 971274042
post@met.no
www.met.no / www.yr.no

Oslo
P.O. Box 43, Blindern
0313 Oslo, Norway
T. +47 22 96 30 00

Bergen
Allégaten 70
5007 Bergen, Norway
T. +47 55 23 66 00

Tromsø
P.O. Box 6314, Langnes
9293 Tromsø, Norway
T. +47 77 62 13 00

Contents

1 Introduction

Numerical weather prediction (NWP) systems depend on the accurate representation of land surface characteristics at their lower boundary. These characteristics include information about topography and the land-cover classifications, such as presence of glaciers, water bodies, urban areas, vegetation amount and type. In NWP systems, these surface properties are used when computing parameters for estimating turbulent and radiative fluxes between the surface and the model atmosphere which are important for correctly representing the energy exchange in the system, thus significantly influencing the model skill and forecast quality (*Walsh et al.*, 2021).

In practice, the slow-evolving surface characteristics are usually represented in static datasets known as physiography data. Creation of such datasets is a demanding and laborious task, which is nowadays often facilitated by applying machine learning techniques. An example of such contemporary data set providing global land-cover classification is ECOCLIMAP Second Generation (SG) (*Lovat et al.*, 2019). However, the methods relying on remote sensing and machine learning have shown limitations, particularly in the Arctic regions. Svalbard has been identified as a region where the quality of automatically generated land cover classifications faces significant challenges. To address these shortcomings, manual corrections based on locally-produced data sets have been applied, in collaboration between the Norwegian Meteorological Institute and the Icelandic Meteorological Office.

The assessment of how modified physiography impacts the forecasting of near-surface atmospheric parameters has not been extensively studied for these datasets. This is especially true for hectometric NWP systems, which put higher demands on the spatial resolution and quality of the physiography datasets compared to contemporary kilometre-scale forecasting system. On the other hand, resolution of the land cover classification maps is not an ultimate metric of their resolving capabilities. When producing these maps, different approaches might be applied in representing cover types within each grid cell of a product. The classic solution applied in older versions of ECOCLIMAP used a so-called mosaic approach where each grid cell provides information about fractions of represented cover types. The new ECOCLIMAP-SG abandons the mosaic approach and show only the most prevalent cover type in each grid cells. Benefits and limitation of each of these two approaches in providing the cover type data are also yet to be thoroughly investigated within a framework of hectometric NWP.

This study explores the impact of modified physiography data over the Svalbard archipelago on simulations conducted using a state-of-the-art hectometric Arctic NWP system, with a 500-meter horizontal grid spacing. This investigation involves a comparative analysis of three physiography versions: the older mosaic-based ECOCLIMAP2, the ECOCLIMAP SG, and a manually corrected version of ECOCLIMAP SG. This assessment was carried out through a combination of direct experimental comparisons and validation of the NWP model's performance, using observational data from remote sensing and weather stations within the model domain.

2 Data and Methods

2.1 Study Area

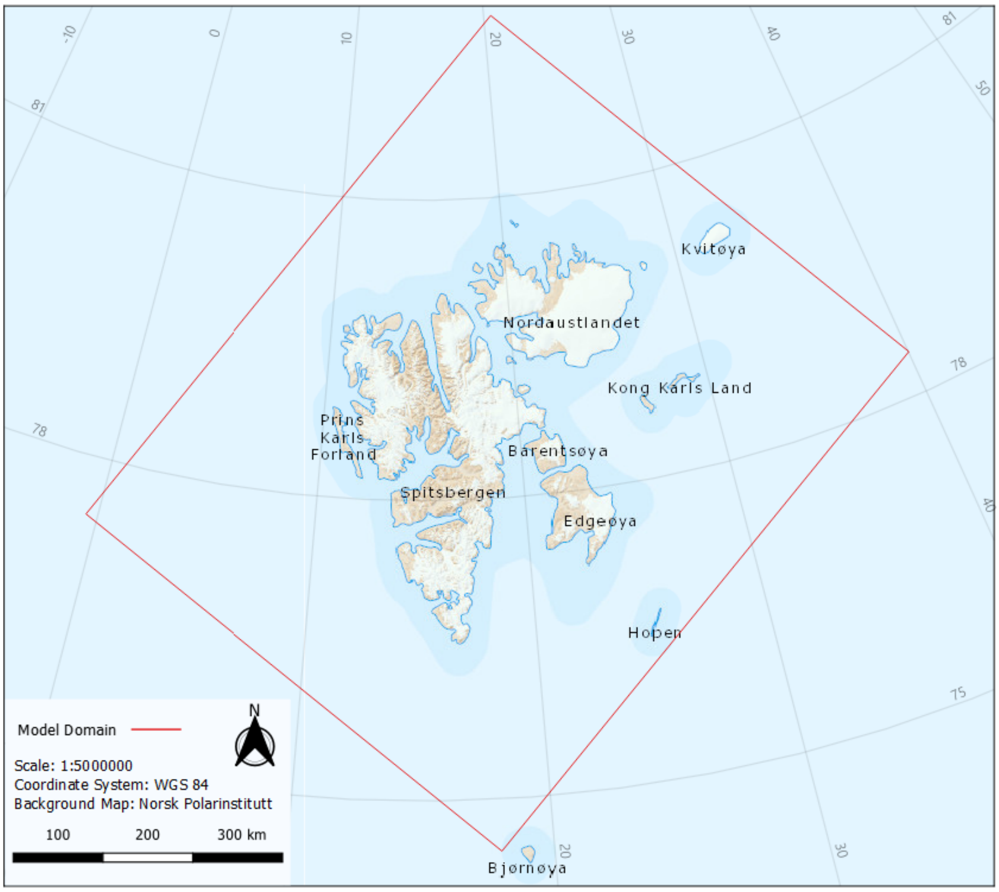

Figure 1: Study Area and overview map over Svalbard Archipelago. The red line contains the Model domain for the AS05 experiments.

The study area, containing the Svalbard archipelago, is shown in Figure 1. The Svalbard archipelago consists of several islands located between approximately 74°N – 81°N and 10°E – 35°E. Svalbard is known for its challenging weather and complex topography with valleys, fjords, steep mountains and glaciers. In addition, there is a large spatial variability in climate within the domain which includes areas with both maritime and continental conditions.

2.2 Numerical Data Sets

In this study, we assess the impact of external physiography data on the forecast accuracy of the HARMONIE-AROME NWP system, with a focus on surface and near-surface atmospheric variables. A series of numerical experiments were conducted using the HARMONIE-AROME NWP system (*Bengtsson et al.*, 2017), version cy46h, and various physiography data sets. The experiments are summarised in Table 1.

Specifically, these include:

1. The ECO2 experiment, which uses land cover data from the ECOCLIMAP2 database with a 1 km spatial resolution (*Faroux et al.*, 2013) and topography data from GMTED2010 terrain elevation dataset with a 250 m spatial resolution (*Danielson and Gesch*, 2011).

2. The SG experiment, which is based on ECO2 but replaces the ECOCLIMAP2 dataset with the newer ECOCLIMAP-SG with a spatial resolution of 300 m.

3. The SG-MOD experiment, which uses updated ECOCLIMAP-SG data with spatial resolution of 30 m and topography data from the digital surface model ArcticDEM with the same spatial resolution of 30 m.

In addition to the SG and SG-MOD experiments, two more simulations, namely, SG0 and SG-MOD0, were performed. These were configured in the same way as SG and SG-MOD, respectively, but with a modification in the turbulent exchange process between the surface and the model atmosphere to prohibit stable stratification within the constant flux layer. This modification involved adjusting the model's internal variable, XRIMAX, as described by *Homleid* (2022). These additional experiments were conducted to better assess the impacts of runaway cooling found in the original SG and SG-MOD configurations. Finally, we used operational archive data from the AROME-Arctic NWP system, maintained and operated by MET Norway (*Müller et al.*, 2017), as an extra reference dataset. The restrictions applied to the processes within the constant flux layer are the same as those used in experiments SG0 and SG-MOD0.

Numerical experiments were configured as follows: For all experiments listed in Table 1 covering the Svalbard archipelago, a dedicated model domain with a 500 m horizontal grid spacing was used. The vertical grid for this domain uses 90 model levels, with the

Experiment name	Domain and horizontal grid spacing	Physiography data	XRIMAX
ECO2	Svalbard 0.5 km	ECOCLIMAP2	0.2
SG	Svalbard 0.5 km	ECOCLIMAP-SG	0.2
SG-MOD	Svalbard 0.5 km	ECOCLIMAP-SG modified	0.2
SG0	Svalbard 0.5 km	ECOCLIMAP-SG	0.0
SG-MOD0	Svalbard 0.5 km	ECOCLIMAP-SG modified	0.0
AROME-Arctic	AROME-Arctic 2.5 km		0.0

Table 1: Experiments

lowest level at 5 m above the model surface. This is in contrast to the coarser vertical grid of the AROME-Arctic dataset, which uses 65 levels and places the lowest model at a 12 m height. Initially, all experiments used a time step of 15 seconds, but this was reduced to 10 seconds due to numerical instabilities developed at the 00 UTC cycle on 21 September, to complete the last two days of simulations. HARMONIE-AROME simulations for this study were forced by IFS-HRES at the model boundaries and did not include atmospheric nor surface data assimilation. This approach was chosen due to the relatively small size of the model domain and the advanced land-surface parameterization scheme, preferred over a simplified land-surface model constrained by observations from a few stations located over Svalbard. Additionally, a large-scale mixing procedure was applied to relax large-scale features of the prognostic fields towards IFS-HRES, preserving the fine-scale state of the hectometric NWP system. A six-hourly cycling was used for all experiments, and 48-hour forecasts were performed at 00 UTC cycles, with output variables, both surface and atmospheric, produced with hourly temporal resolution.

The experimental period of the runs covers 7 days, from 15 September to 22 September 2022. The initial model cold start for all experiments was on 1 September 2022, at 00 UTC. At the cold start, the surface variables within the model domain were interpolated from the 9 km state of IFS-HRES. This state does not adequately resolve all fine-scale features of the Svalbard land cover, requiring an additional initialisation procedure for the land grid cells. For these cells, model fields of the prognostic soil variables were replaced with values interpolated from the 2.5 km grid of the pre-operational version of the AROME-Arctic system. All snow states interpolated from IFS-HRES were removed from the grid cells, and snow cover was re-initialised only for glacier cells with an equivalent of 6 m of snow water. A 14-day spin-up period from 1 September to 14 September yielded the initial model state for each run. The additional runs (SG0 and SG-MOD0) were performed using the initial states prepared for the corresponding SG and SG-MOD runs without repeating the initialisation and spin-up procedures.

2.3 Observations

2.3.1 Satellite Observations

Surface temperature observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument onboard the Terra and Aqua satellites were utilised. This observational product has a resolution of 1 km and only the pixels with a maximum error below 2°C, according to the associated quality control data, were used for model validation. To compare the observations to the model output, firstly, spatial alignment was conducted; for high-resolution 500 m experiments, the MODIS observations were projected onto the model grid, whereas, for the operational 2.5km system, the process was reversed, and the model data were projected onto the observation grid. Secondly, temporal synchronisation was used to ensure that the observations were compared with the model output at the closest possible forecast time. Additionally, a cloud mask was applied to the satellite observations to include only the pixels where the model cloud fraction was less than 0.2, aligning with the MODIS data availability during clear sky conditions. Finally, certain areas, notably a significant portion of Nordaustlandet including the Austfonna glacier in the northeastern part of the archipelago, were excluded from the comparison due to a pronounced cold bias found in all high-resolution model runs, thereby focusing the comparison on more reliable data regions. No height adjustment was done between the satellite observations and model simulations.

2.3.2 Weather Stations

Weather station data were obtained from two sources: the Frost Application Programming Interface (API) (https://frost.met.no/) provided by the Norwegian Meteorological Institute, and stations at lighthouses managed by The University Centre in Svalbard (UNIS) (*Frank et al.*, 2023). These sources provided data for 25 stations for 2m temperature, 21 for 2m humidity, and 19 for 10m wind speed. Data from Frost included a quality code for each measurement; observations flagged as bad or assumed bad were excluded. Sensor heights were manually verified, with any temperature and wind speed measurements not at the specified heights of 2m and 10m, respectively, being removed. Additionally, non-physical values or behaviours were manually filtered out. UNIS data, lacking a quality code, underwent manual review. Observations were compared to the nearest grid point in the model experiments. The location of each station is shown in Figure 2. Most stations are located along the west coast of the main island, Spitsbergen, while four stations are located inland, and four others on smaller islands in the east or north of the main island.

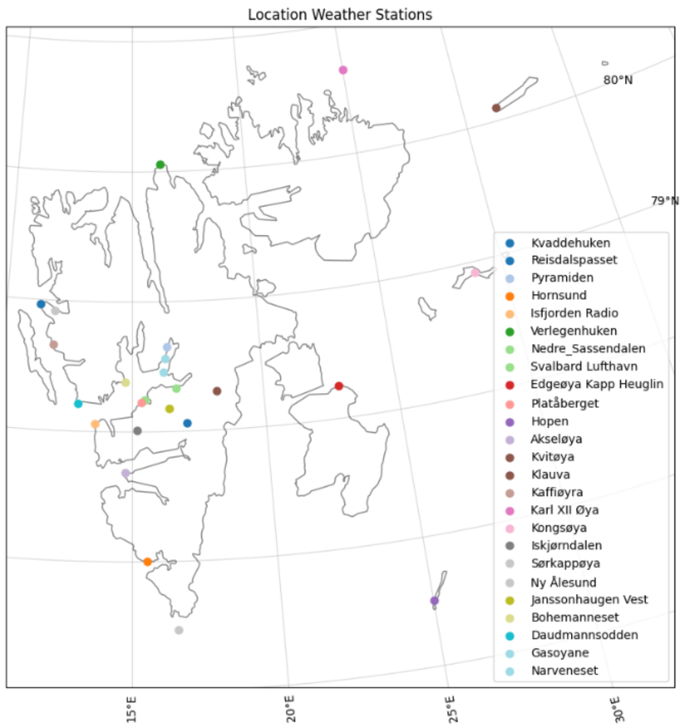

Figure 2: Locations for the weather stations used in the report.

3 Results

3.1 Comparative Analysis of Model Experiments

This section presents the results of the comparative analysis between model experiments without observations. The focus is on assessing differences in temperature and topography.

The most significant 2m temperature differences between SG and ECO2 were found near the coast (Figure 3). SG simulated higher temperatures in these areas, except for three eastern coastal sites where ECO2 was warmer. Additionally, in the western inland regions, ECO2 had lower temperatures than SG. Comparing SG-MOD and SG, the most notable temperature differences were found along the coast and over the Austfonna glacier in Nordaustlandet, northeastern part of the archipelago. SG-MOD generally showed higher temperatures along the coast, while it was cooler over Austfonna compared to SG.

The Probability Density Functions (PDF) of the temperature differences between the experiments are shown in Figure 4. Both SG-ECO2 and SG-MOD - SG have a mode slightly shifted on the positive values. The SG-MOD versus SG comparison shows a broader distribution of temperature differences, with a more pronounced tail indicating lower tem-

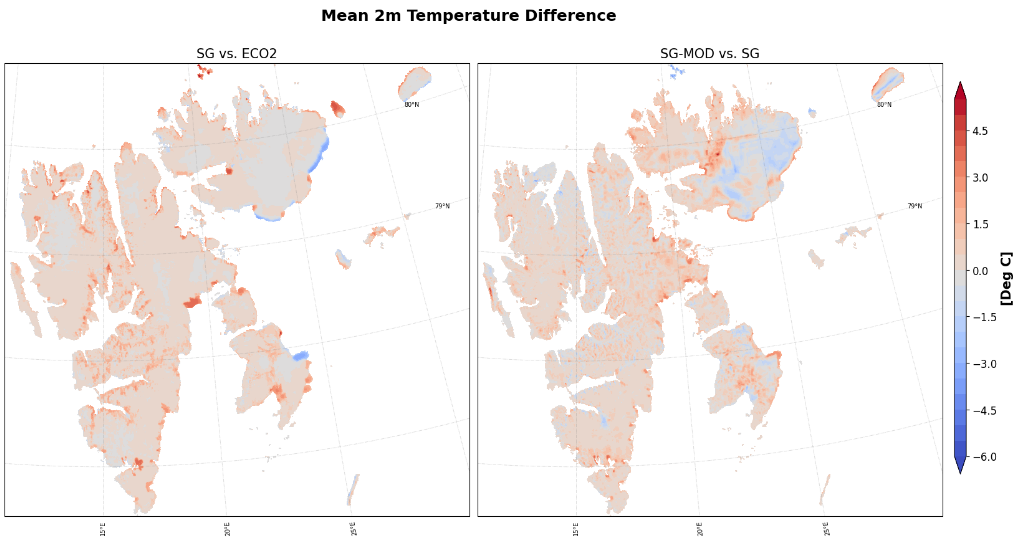

Figure 3: Mean 2-meter temperature difference over land between SG and ECO2 experiments (left), and SG-MOD and SG (right) during 15-21 September 2022.

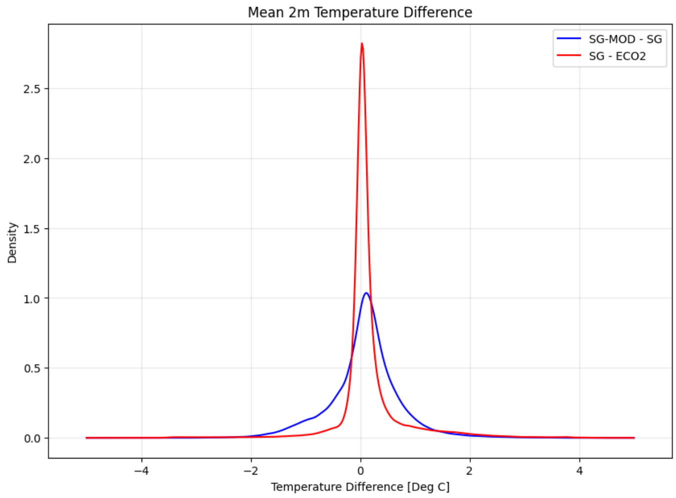

Figure 4: Probability Density Functions (PDF) illustrating the distribution of mean 2-meter temperature differences for SG versus ECO2 (red) and SG-MOD versus SG (blue) for the study period 15-21 September 2022.

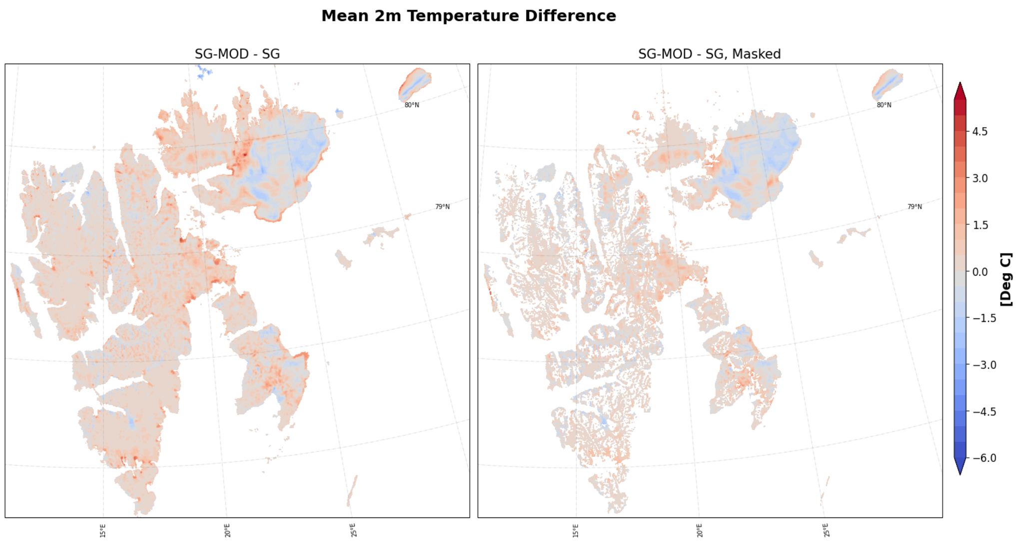

Figure 5: Mean 2-meter temperature difference between SG-MOD and SG experiments when sea mask is applied (left) and when sea and permanent snow mask is applied (right)

peratures for SG-MOD than SG.

In Figure 5, both the left and right plots show the mean temperature differences between SG-MOD and SG. However, the plot on the right excludes the areas with changes in sea and permanent snow cover types, leading to the removal of grid cells along the coast and inland from the visualisation.

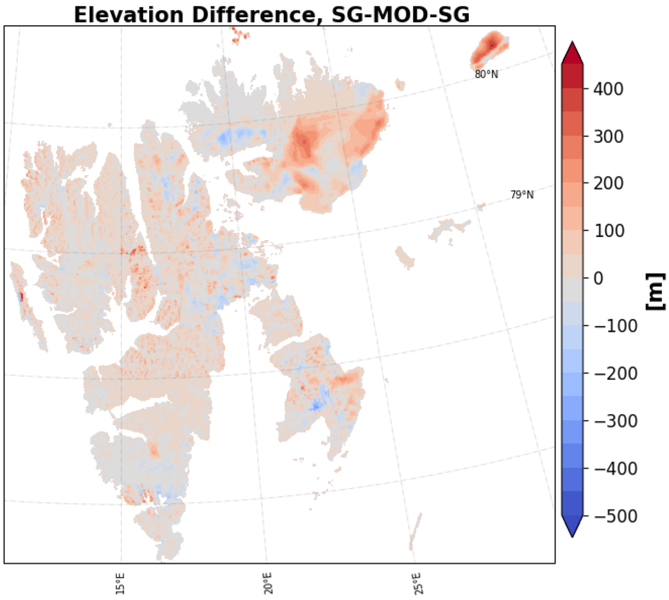

Figure 6: Elevation differences between SG-MOD and SG experiments.

The largest elevation differences between SG-MOD and SG are found in Nord-Austlandet, with elevations up to 404 m higher in SG-MOD than SG, as illustrated in Figure 6. In some regions of Svalbard, changes in topography have led to shifts in the positions of slope and valleys. One such example is the Nordenskiöld Land area, located in the central part of Spitsbergen island.

3.2 Model Evaluation Against Observations

In this subsection, we evaluate the model experiments by comparing their output with observational data from remote sensing and weather stations, focusing on key meteorological variables such as temperature, wind speed, and humidity.

3.2.1 Satellite Observations

Comparison of model surface temperature against MODIS surface temperature observations (Figure 7) shows that ECO2, SG and SG-MOD have too low temperature when the observed temperature is below -6 °C. In contrast, the operational AROME-Arctic, SG0 and SG-MOD0, show less deviation from the observations at these low temperatures, with SG-MOD0 deviating the least. During the daytime, SG and SG0 tend to overestimate temperatures by 1-2°C when the observed temperature is higher than -4°C. The operational

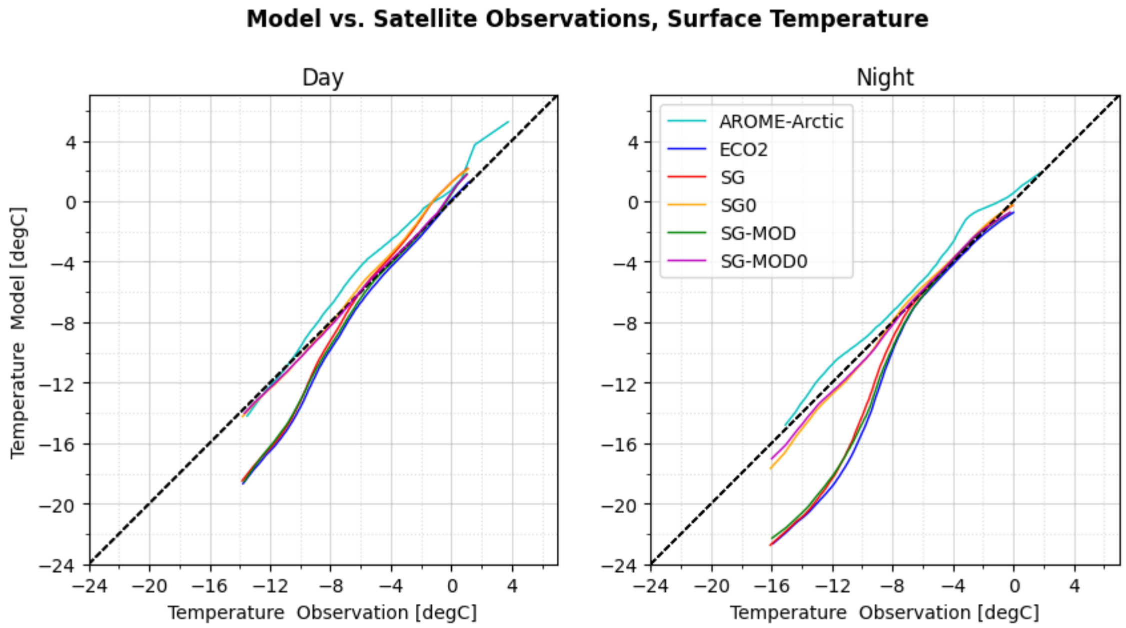

Figure 7: Quantile-Quantile plot of the model surface temperature vs. the observed surface temperature from MODIS during day (left) and night (right).

Variable/Model	ECO2	SG	SG0	SG-MOD	SG-MOD0	Arome Arctic
Surface Temp Day	0.687	0.682	0.726	0.702	0.752	0.798
Surface Temp Night	0.753	0.757	0.780	0.749	0.785	0.858

Table 2: Correlation between the models and the observed surface temperature for both nighttime and daytime observations.

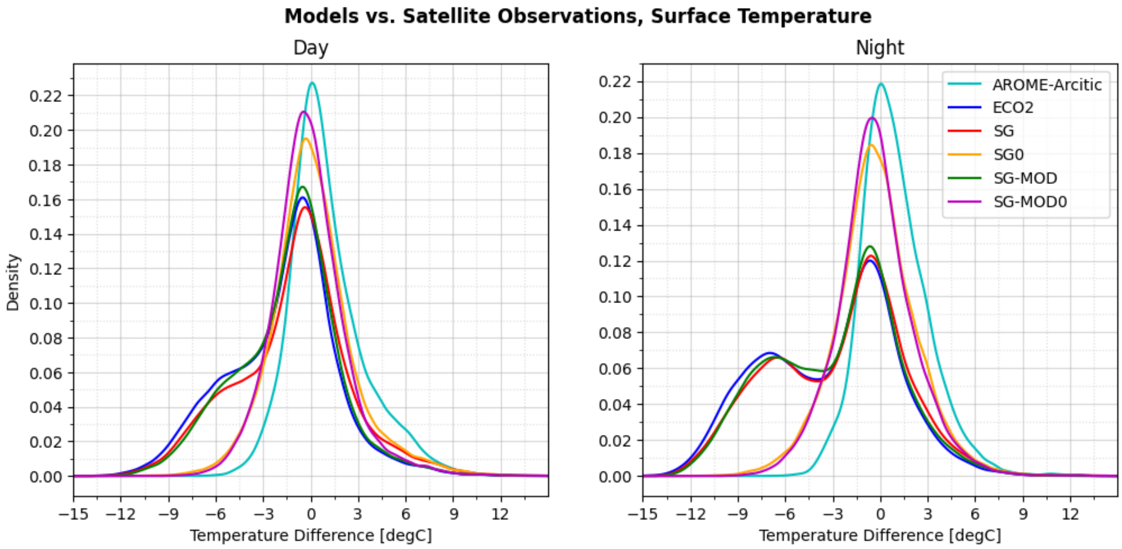

Figure 8: Probability density functions of the differences in surface temperature between the experiments and the MODIS observations during day (left) and night (right).

AROME-Arctic consistently shows a warm bias of up to 2°C for all temperature ranges, both day and night, except at observed temperatures below -10°C during nighttime.

Figure 8 shows the distributions of surface temperature differences between our experiments and the observational MODIS data. Across all experiments, the most prominent peak is observed near zero. The high-resolution experiments ECO2, SG, and SG-MOD tend to exhibit a slight cold bias, while the operational AROME-Arctic model shows a minor warm bias, both during day and night. Notably, ECO2, SG, and SG-MOD show a bimodal distribution. This is particularly evident at night with a secondary peak around -7°C, and a similar, but less pronounced, pattern during the day. In contrast, the operational AROME-Arctic, SG0, and SG-MOD0 demonstrate more symmetric distributions around the mean peak, without any bimodal characteristics.

Throughout the study period, the root mean square errors (RMSE) of the surface temperature in ECO2, SG, and SG-MOD are notably higher compared to those of the operational AROME-Arctic, SG0 and SG-MOD0 for both daytime and nighttime values, as shown in Figure 9. It is noteworthy that the RMSE values are larger at the beginning of the study period and decrease towards the end. Correspondingly, the number of observations con-

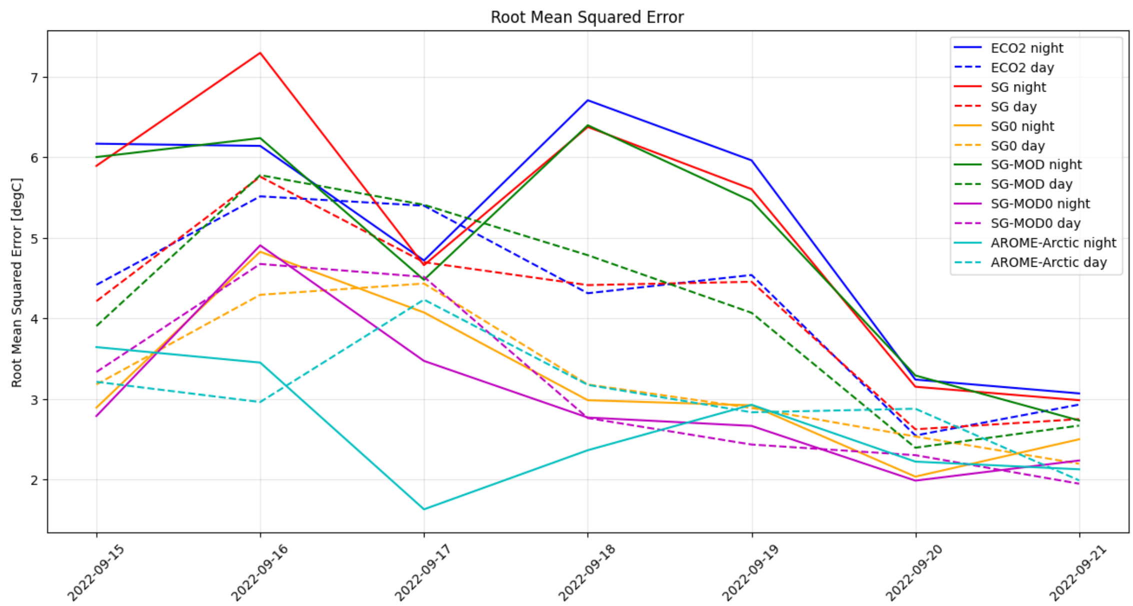

Figure 9: Evolution of root mean square errors (RMSE) of surface temperature for the experiments compared to the MODIS observations during daytime (solid lines) and night-time (dashed lines).

tributing to the RMSE calculation is initially low during the first three days but it increases towards the end, as shown in Figure 10.

3.2.2 Weather Stations

Before evaluating the model performance against station observations, we first present a comparison of how station elevation and cover type vary among the experiments.

Figure 11 illustrates the elevation differences between the surface elevation in model experiments and the actual height of the weather stations. The operational AROME-Arctic has the largest mean absolute elevation difference of 58.1 meters. In comparison, the high-resolution experiments ECO2, SG, and SG-MOD have mean elevation differences of 18.8 meters, 23.5 meters, and 23.8 meters, respectively. Among these, SG-MOD generally has the smallest elevation differences at most stations. An exception is Hopen, where ECO2 has the smallest elevation difference, while SG and SG-MOD show larger discrepancies. It is, however, important to note that the overall topography of Hopen Island, including

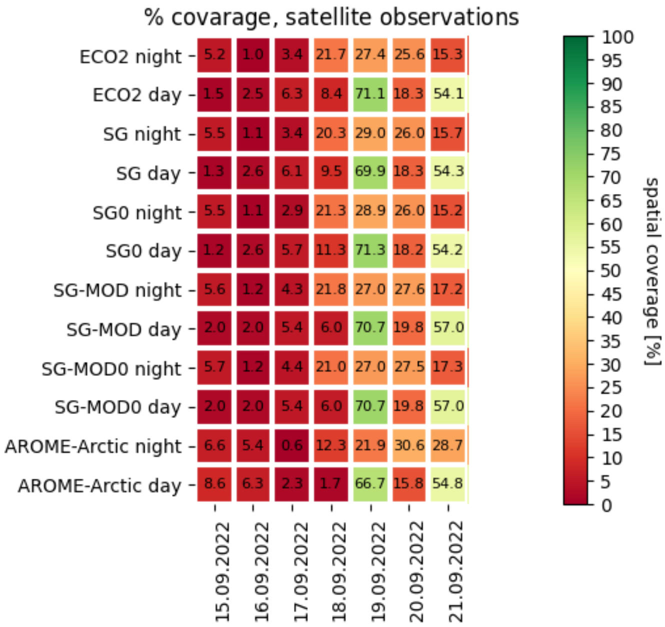

Figure 10: Percentage spatial coverage over all land masses in the domain for day- and nighttime observations in the different experiments.

its elevation maximum and shape, is more accurately captured in SG and SG-MOD than in ECO2. Despite this, the relative positioning of the station on the island is not as accurately represented in SG-MOD. When data from Hopen are excluded, SG-MOD aligns closest to the actual station elevations with a mean difference of 17.3 meters, followed by SG at 18.3 meters, and ECO2 at 18.6 meters.

Cover type fractions at the closest grid cells to weather station locations show considerable changes between the SG and SG-MOD experiments, as presented in Figure 12. The most notable shift is found at Kvitøya station, where the cover type changes entirely from 100% sea to 100% bare land. Additionally, substantial differences in permanent snow cover are evident at Hornsund and Klauva stations when comparing the SG and SG-MOD experiments.

Next, we compare the performance of the model experiments against observations from weather stations. This comparison includes an overall assessment of the experiments and a station-wise analysis of 2m temperature, 10m wind speed and 2m specific humidity. Additionally, we explore the influence of factors like elevation and cloud cover on the accuracy of 2m temperature predictions.

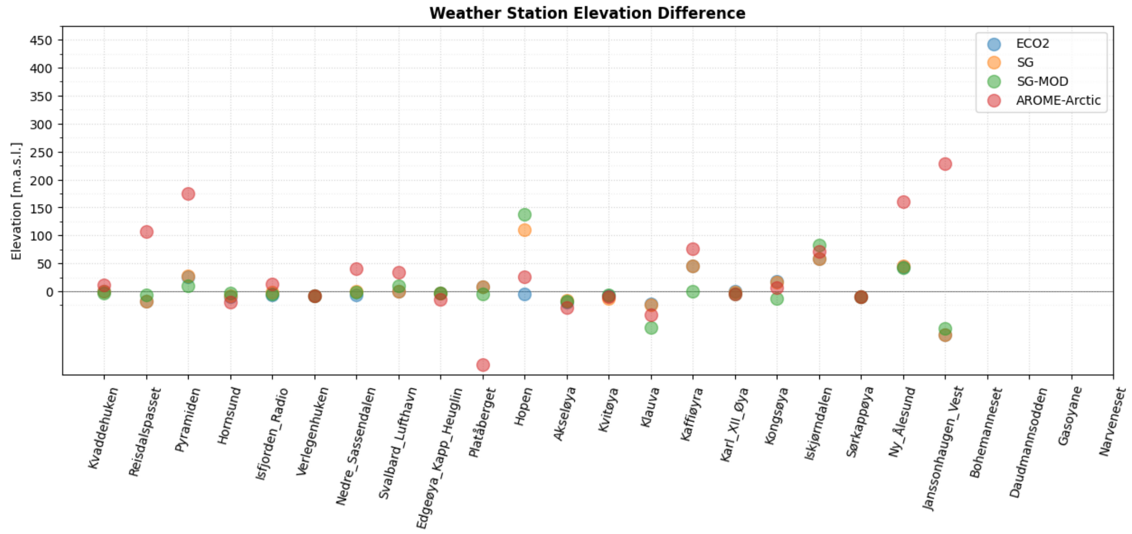

Figure 11: Elevation difference between the experiments and the weather stations.

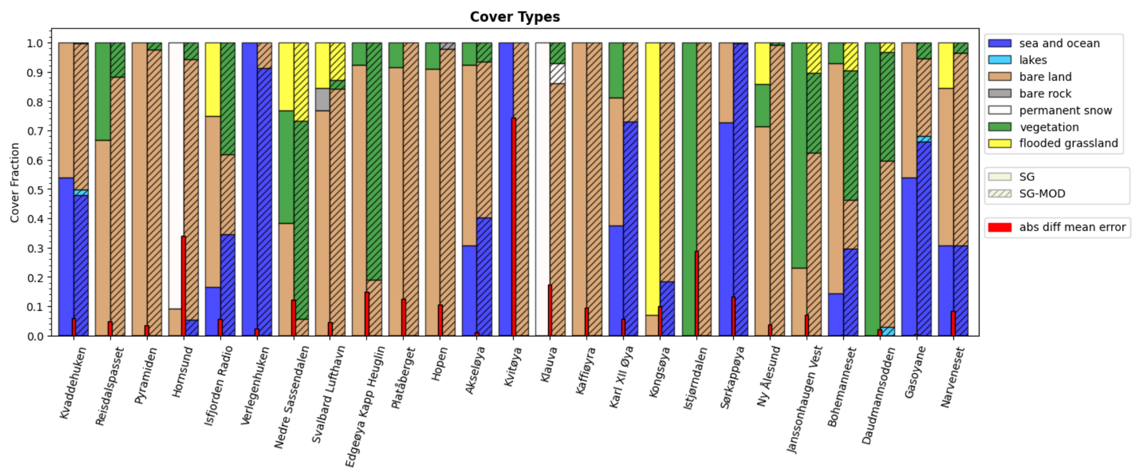

Figure 12: Fraction of the different cover types in the grid cell corresponding to the weather stations in SG and SG-MOD. The red bars are the absolute difference in the mean 2m temperature error in the experiments.

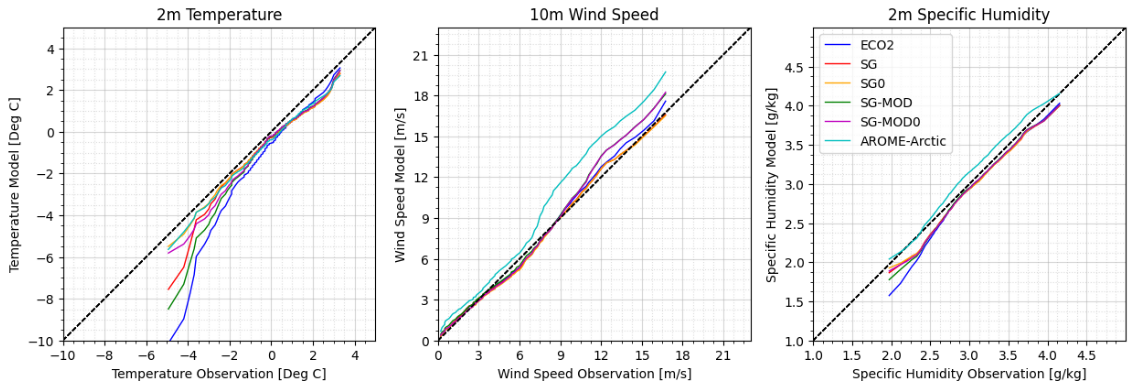

Figure 13: Quantile-Quantile plots for 2m temperature(left), 10m wind speed (middle) and 2m specific humidity (right).

Table 3: Correlation between the experiments and the observations for 2m temperature, 10m wind speed and 2m specific humidity.

Variable/Model	ECO2	SG	SG0	SG-mod	SG-MOD0	Arome Arctic
2m Temperature	0.750	0.777	0.805	0.773	0.802	0.778
10m Wind Speed	0.798	0.802	0.803	0.801	0.802	0.786
2m Specific Humidity	0.703	0.714	0.715	0.708	0.712	0.754

Mean Error, 2m Temperature

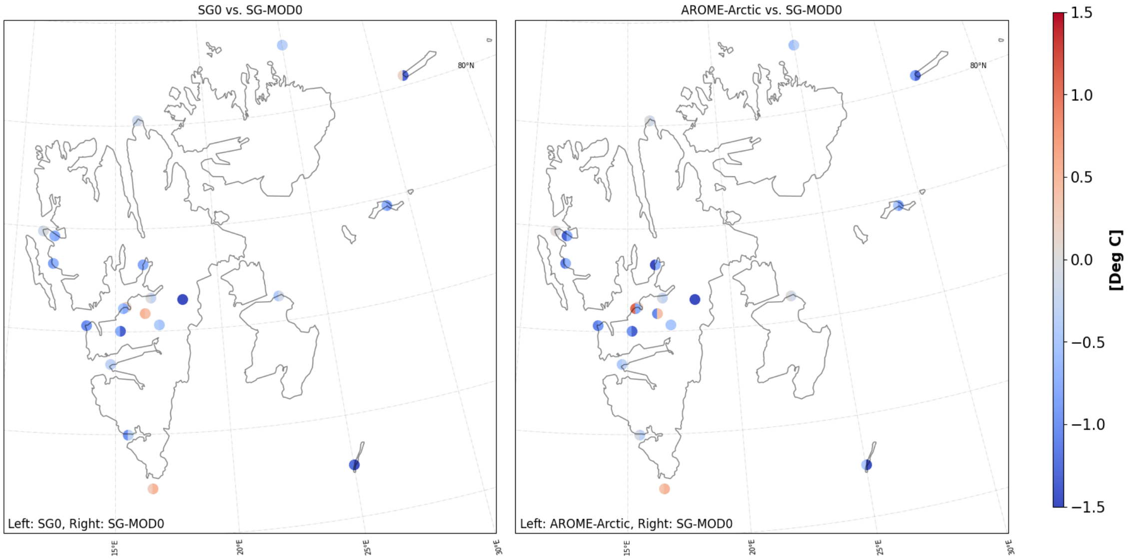

Figure 14: Mean 2m temperature error at the different weather stations. Left: SG0 vs. SG-MOD0. Right: operational model vs. MOD0

Figure 13 shows Quantile-Quantile plots between model experiments and observations for 2m temperature, 10m wind speed, and 2m specific humidity, while correlations are shown in Table 3. For 2m temperature, all model experiments have a cold bias at all temperature ranges. ECO2 has a large cold bias on low temperatures, and the same behaviour is visible for SG-MOD and SG, though less pronounced. Experiments with restrictions applied to the processes within the constant flux layer, SG0 and SG-MOD0, have significantly smaller biases on low temperatures than the equivalent experiments without the restriction. SG0 and SG-MOD0 have also the highest correlations for 2m temperature. For temperatures above 0°C, all experiments perform nearly similarly with a slight cold bias in all experiments.

Figure 14 shows the mean errors for 2m temperature at each station for the SG0, SG-MOD0, and operational AROME-Arctic. Generally, these models tend to simulate lower temperatures than observed at most stations. However, at Sørkappøya, located at the southern tip of Svalbard, all experiments show a positive temperature bias. This is particularly strong in SG-MOD0, likely due to its grid point being completely covered by sea, as

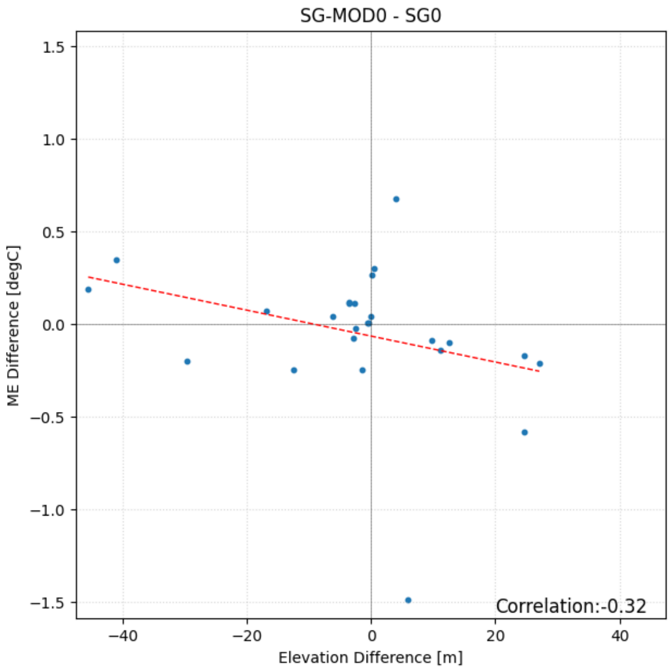

Figure 15: Elevation difference between models plotted against the difference in the mean error in 2m temperature, comparing SG-MOD and SG.

illustrated in Figure 11. The Istjørndalen station, situated in central inland Svalbard, and Kvitøya station, the easternmost station, show the most significant differences between SG0 and SG-MOD0. At Istjørndalen, SG-MOD0 has a stronger cold bias compared to SG0. At Kvitøya, SGG0 has a positive mean error, while the SG-MOD0 has a negative one. When comparing the operational AROME-Arctic with SG-MOD0, the Platåberget station stands out with the largest difference, with the operational being too warm and the SG-MOD0 being too cold.

Figure 15 illustrates how the mean error (ME) in 2m temperature correlates with the elevation differences between SG0 and SG-MOD0. The correlation coefficient between these elevation differences and the ME is -0.306. This suggests that there is a connection with simulated elevation errors and temperature errors. However, this pattern is not consistent across all stations. For example, seven stations simulated lower temperatures in the SG-MOD model compared to SG, despite having lower elevations in SG-MOD.

To better understand the temperature biases found at Istjørndalen, we present the simulated cloud cover fraction at the station, as shown in Figure 16. From September 17th

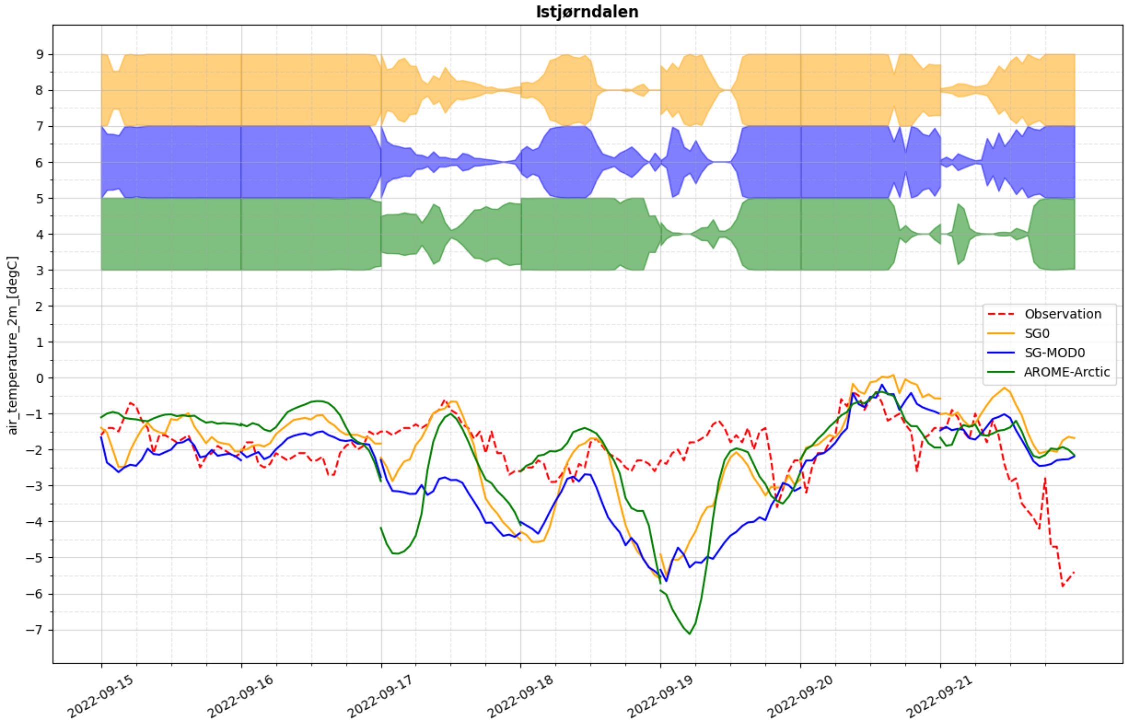

Figure 16: 2m temperature in SG0, SG-MOD0 and the operational model. Red dotted line is the observed temperature at Istjørndalen station. In the upper part of the Figure, the total cloud fraction in ECO2, SG and SG-MOD is plotted.

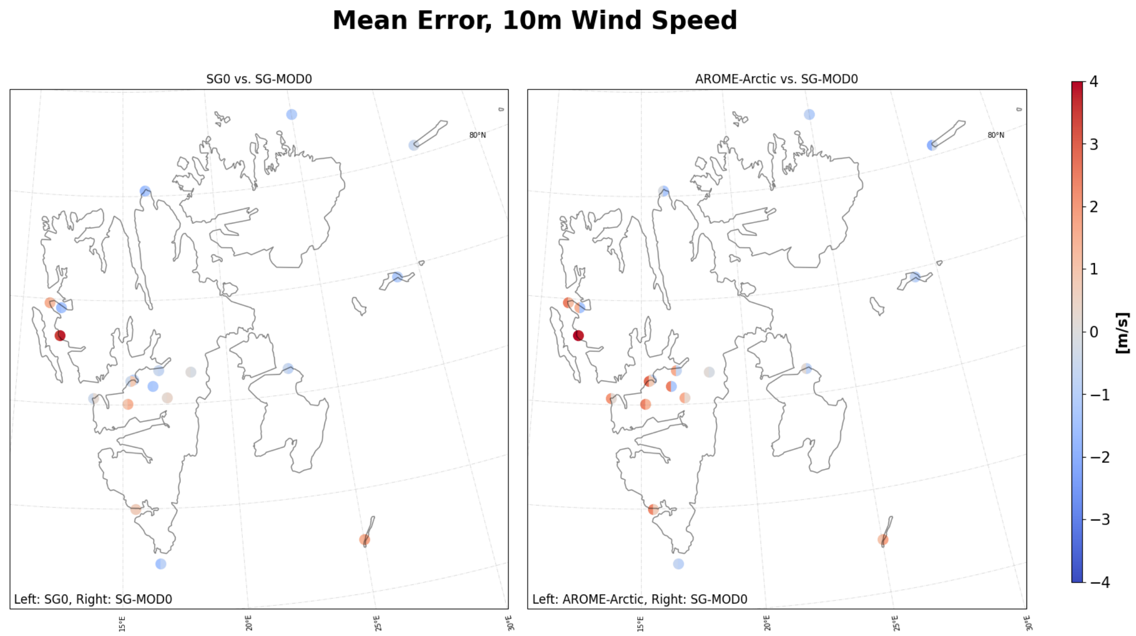

Figure 17: Mean wind speed error at the different weather stations. Left: SG0 vs. SG-MOD0. Right: operational model vs. mod0

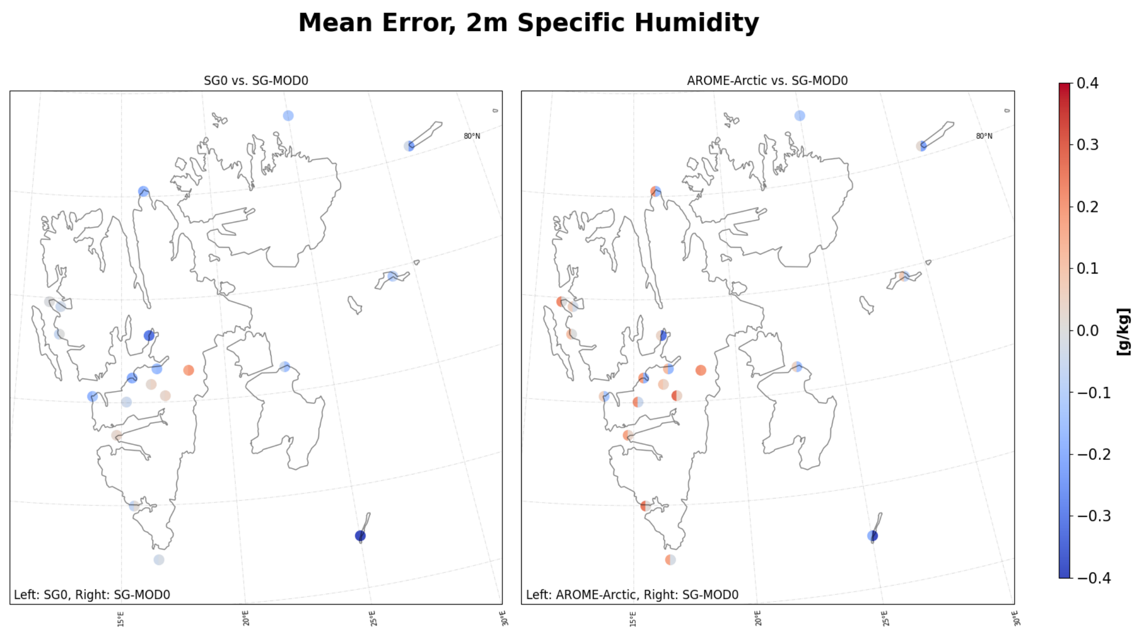

Figure 18: Mean error in specific humidity at the different weather stations. Left: SG0 vs. SG-MOD0. Right: operational model vs. SG-MOD0

to 19th, and on the evenings of September 20th and 21st, the simulated cloud cover varied between full coverage and clear skies. The most notable differences between SG0, SG-MOD0 and AROME-Arctic occur on September 18th. On this day, the operational AROME-Arctic shows a 100% cloud cover for a longer period than the high-resolution experiments. Satellite images from September 17-18 confirm consistent full cloud cover over the Istjørndalen station during all available satellite overpasses.

The observed temperature evolution at Istjørndalen show little to no diurnal variation throughout the week until a drop on the evening of September 21st. In contrast, all model experiments have a diurnal pattern from September 17th to 19th, aligning with the cloud-free periods in the simulations. The operational AROME-Arctic has the most pronounced diurnal cycle during this period. It also differs the most from the observed temperatures at night. During the day, SG0's peak temperatures align closely with observed values. SG-MOD0, while matching SG0's nighttime temperatures, fails to rise as much during the day, resulting in consistently lower temperatures than observed.

For 10m wind speed the operational AROME-Arctic on average simulates higher speeds than observed at all wind speed ranges, as shown in Figure 13. This discrepancy increases with higher wind speeds. Among the high-resolution experiments, SG0 is the most accurate with a correlation of 0.803, while ECO2 is the least accurate at 0.798. The operational AROME-Arctic has the lowest correlation at 0.786.

When examining the mean errors in 10m wind speed at different stations, as shown in Figure 17, we see varying accuracy with some stations simulating wind speeds that are too high, and others too low. The operational AROME-Arctic tends to overestimate wind speeds at most stations, resulting in a higher mean absolute error of 1.6m/s and a peak error of 5.3m/s across all stations. In comparison, the high-resolution models SG0 and SG-MOD0 are more accurate, both having a mean absolute error of 1.0m/s. However, they differ slightly in their maximum errors, with SG0 reaching 3.4m/s and SG-MOD0 going up to 3.8m/s.

For 2m specific humidity averaged over all stations, the operational AROME-Arctic tends to show higher (wetter) values than the observations, while the higher resolution experiments show a tendency towards lower (dryer) values (Figure 18). The operational model has the highest correlation at 0.754. Among the high-resolution experiments, correlations range from 0.703 to 0.714.

For 2m specific humidity at different stations, we see a shift in bias between the high-resolution experiments and the operational model, as shown in Figure 18. The operational model tends to simulate too moist conditions along the coast and too dry inland. On the other hand, the high-resolution experiments show the opposite situation, with too moist conditions inland and drier conditions along the coast.

4 Discussion

4.1 Experiment Comparison

The experiments using different physiography databases reveal the most pronounced temperature differences mainly along the coast and in certain inland areas, as highlighted in Figures 3 and 5. Inland areas with reduced permanent snow cover also show a considerable 2m temperature increase, confirming the influence of adjusted glacier masks. Hovewer, for some areas, differences in the temperatures can not be explained only by the change in cover types. For example, over Austfonna, temperature differences can be attributed rather to the changes in topography than in land cover classification. This is also seen over the Edgeøya island in the southeastern part of the domain, where regions of higher elevation in SG-MOD are associated with lower temperatures. The Probability Density Functions (PDFs) in Figure 4 also demonstrate a bias towards lower temperatures in SG-MOD, likely due to these topographic changes over Austfonna. The improved topographic resolution in the experiments is further illustrated by temperature changes over Prins Karl Forland and in the valley structure of Norensköld Land. In Prins Karl Forlands land, a mountain ridge shifted westwards between SG and SG-MOD, resulting in higher temperature along the coast in SG-MOD and a slightly lower temperature further inland compared to SG. Similarly, in Norensköld Land, temperature changes along the valley sides align with topographic changes.

4.2 Observations

4.2.1 Satellites Observations

The most pronounced difference in temperature between our experiments arose from adjusting the turbulent exchange between the surface the model atmosphere in stable conditions. The XRIMAX set at 0.2 allows for a stable stratification within a layer between the lowermost model level and surface, whereas setting it to 0 enhances surface mixing by always keeping this layer unstable or neutral. This change had a more substantial impact on model performance than physiography modifications. Experiments with XRIMAX at 0.2 showed lower surface temperatures than observed, especially at night, due to a runaway cooling effect caused by surface-atmosphere decoupling. This effect is apparent when modelled surface temperature is compared against a satellite product since this comparison can be done only for clear-sky model grid cells which favour stable stratification and more pronounced decoupling.

The Root Mean Square Error (RMSE) decreased by 29% for ECOCLIMAP-SG and 32% for its modified version when XRIMAX was set to zero, particularly during calmer synoptic conditions from September 15th to 19th. However, fewer observations due to extensive cloud cover from September 15th to 17th increase uncertainty for this period. The approach of a storm system on September 20th and 21st resulted in increased wind speeds, which seem to prevent the decoupling between the surface and atmosphere and limited the cooling effect in experiments with XRIMAX at 0.2.

It is important to note that our analysis only compared grid cells with model cloud fractions below 0.2 to MODIS products and limited our observations to clear-sky conditions. This approach potentially over-represents critical situations for models with a positive XRIMAX value. Under cloud-free conditions, the small incoming long-wave radiation flux typically leads to strong radiative cooling of the surface and turbulent exchange between the surface and atmosphere is very important. While this analysis does not change the fundamental issue of decoupling in stable conditions, the satellite observations do not provide insight into model performance under cloudy conditions.

The large impact of the tuning parameter XRIMAX can be expected to vary with season. In the Arctic, stable conditions occur more often during the autumn and winter seasons. A study period in the autumn, as in this report, will therefore enhance the impact of this parameter compared to an experiment period, for example, during the summer. Further, the enhanced mixing between the surface and the atmosphere during the summer season could have led to larger differences between the experiments than what the chosen study period provided.

4.2.2 Weather Stations

Most weather stations in our study are along the western coastal areas of the Svalbard archipelago, with little to no change in the permanent snow cover in the physiography datasets. This limits the detection of direct effects due to cover type changes in our model comparisons. Only three stations, notably Kvitøya, Klauva, and Hornsund, had significant changes in cover type. Kvitøya, with a shift from sea to land cover, shows the largest change in 2m temperature error. It was 0.2°C too warm in SG0, having all sea as the cover type, while in MOD0, it was -1.3°C too cold, having all land. Klauva and Hornsund have a lower permanent snow fraction in SG-MOD compared to SG. Further, Hornsund has the second largest change in the mean error between SG0 and SG-MOD0, with the cold bias is 0.7 degrees larger in SG0, corresponding with a larger fraction of permanent snow in the grid cell. The improvement in the performance at this station may, therefore, be directly related to the updated permanent snow cover in SG-MOD. Despite minor differences in correlation across variables, the performance of SG-MOD remains comparable to SG. This contrasts with satellite data results, particularly in temperature, likely due to the limited spatial coverage of the weather stations in areas affected by cover type changes.

When evaluating ECO2, SG and SG-MOD, both topograhpy updates and cover type changes are important. Generally, the shift from SG to SG-MOD improved forecast accuracy due to better topography. However, Hopen is an exception where SG and SG-MOD perform worst among the high-resolution experiments, and with SG-MOD performing even worse than SG. A closer look at Hopen reveals SG-MOD offers more accurate terrain representation than ECO2, which flattens the island's mountainous terrain to sea level. The location of the weather station close to the higher-elevation mountains surrounding the station creates a mean elevation for the coverage of the grid cell, higher than the elevation at the station itself for SG and SG-MOD, worsening their performance when compared to the station's elevation. ECO2, not being able to capture the topography of the mountainous areas, has a station height at sea level, which happens to be closer to the actual value. Thus, while SG and SG-MOD's elevation inaccuracies introduce a cold bias and degrade their performance at Hopen, they overall present a more realistic depiction of the area's topography.

At locations such as Kaffiøyra, the increased accuracy in the topography in SG-MOD likely contributed to its better performance compared to the other models. In this case, SG-MOD has a smaller elevation difference than the other experiments, leading to a smaller cold bias and indicating a more accurate forecast. This relationship between elevation adjustments and temperature accuracy is supported by the data, as illustrated in Figure 15. However, this correspondence is not found across all stations. Variability in model performance can partly be attributed to changes in cover types, yet some discrepancies remain unexplained. These may be due to the changing synoptic situation during the study period, affecting the influence of physiography updates on model accuracy.

At Istjørndalen station, SG-MOD0 performs worse compared to both the operational model and SG0, simulating colder temperatures than observed. This discrepancy can be attributed to the pronounced diurnal cycle at the station. Nighttime radiative cooling, intensified by a lack of cloud cover found in the experiments, resulted in a significant cold bias. SG0 and the operational model managed to counteract this effect with stronger daytime heating than SG-MOD0. Consequently, SG-MOD0 not only has a cold bias at night but also has a cold bias during the day, leading to its overall poorer performance in simulating 2m temperature at Istjørndalen.

The Quantile-Quantile plots and correlations for 2m temperature, 10m wind speed, and 2m specific humidity highlights the performance characteristics of each experiment. Adjusting the XRIMAX parameter to zero markedly improves the model's performance at lower temperatures, and the runaway cooling is reduced due to the enhanced mixing at lower atmospheric levels. Although this influence is less pronounced for 2m temperatures than for surface temperatures, it is still noticeable.

For wind speed, all high-resolution experiments perform better than the operational model, which tends to predict too high wind speeds, with errors increasing at higher wind speeds. SG and SG0 perform the best among the high-resolution models, while SG-MOD and SG-MOD0 consistently overpredict wind speeds for the higher wind speeds. An analysis of individual stations does not directly connect this to either topographical or cover type modifications, suggesting the influence of additional factors on these differences.

For specific humidity, the operational model tends to overestimate humidity, whereas the high-resolution experiments are too dry. Observations at individual stations reveal a geographic bias: the operational model is too moist along the coast and too dry inland, while for the high-resolution experiments show the opposite trend. The results do not highlight significant improvements from one ECOCLIMAP version to another. However, as discussed previously, there are individual differences within the stations, and the most pronounced changes in the physiography do not occur in most of the locations of the weather stations.

5 Conclusions

In this study, we evaluated the impact of improved physiography data over the Svalbard archipelago on simulations conducted using a hectometric NWP system for a period of one week in September 2022. The assessment was carried out through a combination of direct experimental comparisons and validation of the NWP model's performance using observational data from the MODIS satellite, and weather stations. Our analysis focused on understanding the impacts of the updated physiography dataset (experiment SG-MOD) against earlier versions of the datasets (experiments ECO2 and SG).

Our main conclusions are the following:

- Modifications in physiography data have a pronounced effect on modelled temperatures. The largest impact was seen in areas with updated permanent snow and sea cover, as well as with changes due to topography, as evident in regions along the coast and some inland areas.
- Verification of the experiments' performance against weather station observations showed no significant improvement with the modified physiography dataset in measurements of 2m temperature, 10m wind speed, and 2m specific humidity.
- Small improvements in SG-MOD's performance over SG were apparent when compared to satellite observations of surface temperature, likely due to the satellite observations available over a wider area affected by the updated physiography data.
- The tuning parameter for atmospheric stability significantly impacts model performance, especially in relation to temperature observations both from weather stations and satellites.

Caution should be given to the shown impact of physiography due to the chosen study period in mid-September. The impact of physiography is likely to be more pronounced during the summer, particularly due to the absence of snow cover. Additionally, the influence of the tuning parameter for atmospheric stability is expected to be less significant in the summer, given the lower occurrence of stable conditions compared to autumn and winter. More robust results could be achieved if a repeat analysis were conducted with a study period that allows for a greater impact of physiography, thereby allowing a better understanding of its influence. Lastly, understanding the separate impacts of topography and cover types remains a challenge. Future experiments that focus on updating only one of these aspects at a time could provide clearer insights into their individual impacts.

Given the modest yet positive impact of the improved physiography demonstrated in this study, we plan to continue utilising the enhanced physiography dataset in our future work.

References

Bengtsson, L., U. Andrae, T. Aspelien, Y. Batrak, J. Calvo, W. de Rooy, E. Gleeson, B. Hansen-Sass, M. Homleid, M. Hortal, K.-I. Ivarsson, G. Lenderink, S. Niemelä, K. P. Nielsen, J. Onvlee, L. Rontu, P. Samuelsson, D. S. Muñoz, A. Subias, S. Tijm, V. Toll, X. Yang, and M. Ødegaard Køltzow (2017), The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system, *Monthly Weather Review*, *145*(5), 1919 – 1935, doi:10.1175/MWR-D-16-0417.1.

Danielson, J. J., and D. B. Gesch (2011), Global multi-resolution terrain elevation data 2010 (GMTED2010), *Tech. rep.*, doi:10.3133/ofr20111073, report.

Faroux, S., A. T. Kaptué Tchuenté, J.-L. Roujean, V. Masson, E. Martin, and P. Le Moigne (2013), ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, *Geoscientific Model Development*, *6*(2), 563–582, doi:10.5194/gmd-6-563-2013.

Frank, L., M. O. Jonassen, T. Remes, F. R. Schalamon, and A. Stenlund (2023), Iwin: the isfjorden weather information network, *Earth System Science Data*, *15*(9), 4219–4234, doi:10.5194/essd-15-4219-2023.

Homleid, M. (2022), Improving model performance in stable situations by using a pragmatic shift in the drag calculations - XRISHIFT, *ACCORD Newsletter 2*, pp. 96–108.

Lovat, A., B. Vincendon, and V. Ducrocq (2019), Assessing the impact of resolution and soil datasets on flash-flood modelling, *Hydrology and Earth System Sciences*, *23*(3), 1801–1818, doi:10.5194/hess-23-1801-2019.

Müller, M., Y. Batrak, J. Kristiansen, M. A. O. Køltzow, G. Noer, and A. Korosov (2017), Characteristics of a convective-scale weather forecasting system for the European Arctic, *Monthly Weather Review*, *145*(12), 4771 – 4787, doi:10.1175/MWR-D-17-0194.1.

Walsh, E., G. Bessardon, E. Gleeson, and P. Ulmas (2021), Using machine learning to produce a very high resolution land-cover map for ireland, *Advances in Science and Research*, *18*, 65–87, doi:10.5194/asr-18-65-2021.