



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Department of Earth Science System
Climate, Radiation and Remote Sensing
Meteorological Institute

Master's Thesis

Developing a new cloud cover product for the evaluation of an operational weather prediction model in the Arctic

primary supervisor: Prof. Dr. Stefan Bühler
secondary supervisor: Dr. Marvin Kähnert

Leonie Nellesen, 7399948
Gluckstraße 74c, 22081 Hamburg
leonie.nellesen@studium.uni-hamburg.de
M.Sc. Atmospheric Science

Submission:
Hamburg, 30.01.2026

Abstract

Accurate cloud coverage forecasts in the Arctic are of increasing importance as declining sea ice drives a rapid increase in human activity. However, cloud coverage observations required for forecast validations remain sparse and pose significant challenges.

This study utilizes images of an all-sky camera located in Svalbard to enhance the availability of cloud coverage observations during the polar night. An image-based machine learning algorithm is developed to extract the cloud coverage from these observations.

The resulting cloud coverage estimates are then used to evaluate the cloud coverage performance of the numerical weather prediction model AROME-Arctic. While AROME-Arctic captures the overall picture reasonably well, it exhibits notable difficulties in accurately predicting the cloud coverage during cold air outbreaks.

Contents

1. Introduction	5
2. Datasets	7
2.1. All-Sky Images	7
2.2. AROME-Arctic	7
3. Cloud Coverage Retrieval	11
3.1. Physical Approach	11
3.2. Machine Learning Approach	14
3.3. Combined Approach	17
4. AROME-Arctic Evaluation	21
4.1. Statistical Evaluation	21
4.2. Case Studies	28
4.2.1. Cold Air Outbreaks	28
4.2.2. Displacement Problem	34
5. Summary and Discussion	37
6. Conclusion and Outlook	41
A. Additional Figures	43
B. Additional Case	45
Bibliography	47

1. Introduction

Weather forecasts in the Arctic are of increasing importance with expected increase in human activity due to the significant reduction in sea ice extent caused by climate change (IPCC, 2021). New shipping routes emerge, natural resources are more accessible, and tourism expands.

All of these activities depend on reliable weather forecasts. In particular, the occurrence of clouds needs to be forecasted precisely as clouds strongly influence visibility, precipitation, icing, and near-surface temperature.

To ensure high forecast quality, observations are needed for validation (Bauer et al., 2015; Bengtsson et al., 2017). Observations remain particularly sparse in the Arctic due to its limited accessibility. Remoteness, complicated logistics, and severe weather conditions significantly complicate conducting measurements in the Arctic. Most measurement campaigns are restricted to summer months due to the extensive sea ice coverage in winter (Tjernström et al., 2004; Tjernström et al., 2012; Moser et al., 2023). Spatially, observations are limited by the flight path of airplanes or the route of the ship conducting the measurements. Long-term stationary measurements are constrained to coastal regions and areas near human settlements (Lee et al., 2022).

Cloud observations are especially difficult to conduct under these conditions. Among the first to measure the climatology of cloud coverage in the Arctic was Sverdrup (1933). In his work, he noted that cloud observations are difficult to quantify. Nowadays, several methods are available to observe cloud coverage. For example, measurement campaigns with airplanes, research vessels, or ground-based measurements on land, like ceilometers, radars, and lidars, and satellite observations. Synoptists, for example, at airports, report cloud coverage as well. However, the quantity of the existing observations is still too sparse to ensure sufficient data for the evaluation of weather forecasts.

This study aims to help close the gap of too sparse cloud coverage observations in the Arctic. A long-term deployed all-sky camera located at the Kjell Henriksen Observatory (KHO, 2026, KHO) near Isfjorden in Svalbard captures nighttime images (Figure 1.1 black pin and Figure 1.2). Until now, these images were only used for polar lights. This study explores their utility for cloud coverage observations as we develop a new way to gain cloud coverage information from these images.

To contribute to evaluating the accuracy of weather forecasts in the Arctic, the newly developed cloud coverage product is subsequently used to evaluate the cloud coverage prediction in the numerical weather prediction (NWP) model AROME-Arctic, a meso-scale, convection-permitting weather forecasting model for the European Arctic (Norwegian Meteorological Institute, 2025, Figure 1.1).

Chapter 2 introduces the images of the all-sky camera and the NWP model AROME-

1. Introduction

Arctic. Chapter 3 extracts the cloud coverage from the images, and in Chapter 4 the model's cloud coverage prediction is validated both statistically and on a case studies bases.

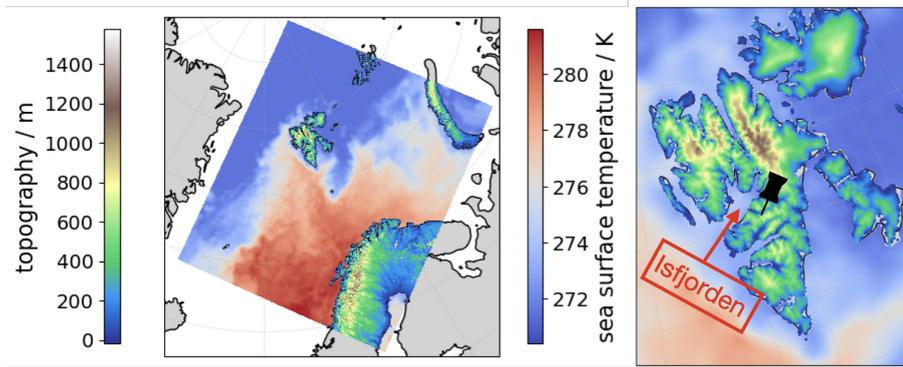


Figure 1.1.: Spatial domain of AROME-Arctic on the 2023-11-29 00:00 UTC. Landmasses show the topography, and sea areas show the sea surface temperature. The black pin marks the location of KHO (78.148° N, 16.043° E).



Figure 1.2.: KHO looking towards Isfjorden. The all-sky camera sits below one of the domes on the roof of KHO (2026).

2. Datasets

This study utilizes two datasets. One of which is an all-sky camera, and the other is the NWP model AROME-Arctic.

2.1. All-Sky Images

The main protagonists of the study are the all-sky images (Figure 2.1) captured by the all-sky camera located at KHO in Svalbard (Sony A7s, 2026, Figure 1.1). The primary purpose of the camera is to observe polar lights. Thus, the camera only operates during the polar night, which is approximately between mid-October and mid-March.

The camera is a Sony A7s all-sky camera with a Sigma 8mm f/3.5 EX DG Circular Fisheye lens. Minutely images are taken during the polar night. The images have a resolution of 480 x 480 pixels. For this study, the field of view of the equidistant camera is reduced from 180° to 169° to remove surrounding instruments and mountains from the horizon of the images. The new size of the images is a circle with a radius of 210 pixels (Figure 2.2). Until now, the image has been solely used to study polar lights. Clouds were only a byproduct and were filtered out. This study now utilizes the captured clouds. Images taken between 2019 and 2023 are manually labeled with cloud coverage in okta. Chapter 3 discusses a physical approach and develops a machine learning approach to extract the cloud coverage from these images. Images from 2019 and 2020 are later used as training data, while images from 2021 - 2023 are used as testing data.

2.2. AROME-Arctic

AROME-Arctic is a convection-permitting NWP model for the European Arctic (Figure 1.1) and is developed and operated by the Norwegian Meteorological Institute (Müller et al., 2017). It has been running operationally since November 15th, 2015. AROME-Arctic is based on the NWP model HARMONIE-AROME (Gleeson et al., 2024) with a horizontal spatial resolution of 2.5 km extending vertically up to 24 km over 65 vertical layers in hybrid sigma-pressure coordinates. A semi-Lagrangian spectral advection scheme serves as the dynamical core (Seity et al., 2011)

AROME-Arctic uses multiple parametrization schemes: A one-moment bulk microphysics scheme (Pinty and Jabouille, 1998) with special adaptations for cold conditions (Gleeson et al., 2024), a statistical cloud and condensation scheme (Bougeault, 1982; Bechtold et al., 1995; Rooy et al., 2022), a radiation scheme (Fouquart and Bonnel, 1980; Mlawer et al., 1997) employing six spectral bands for shortwave and 16 for longwave radiation, a surface scheme (Le Moigne, 2009), a 1.5 order turbulence scheme based on turbulent kinetic energy (Lenderink and Holtslag, 2004) and a dual-updraft shallow convection scheme (Neggers

2. Datasets

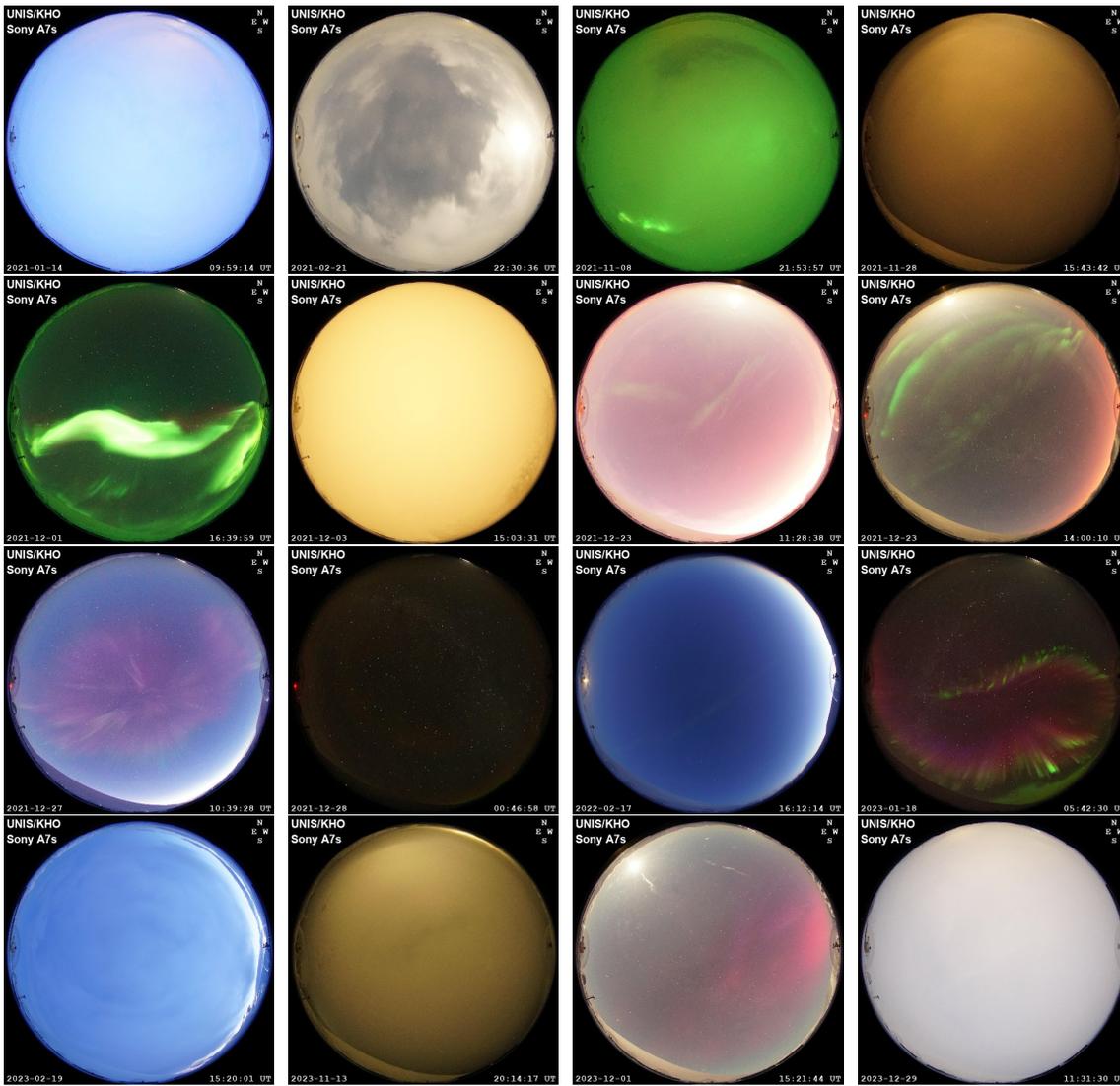


Figure 2.1.: Example of unprocessed camera images taken by the all-sky camera at KHO.

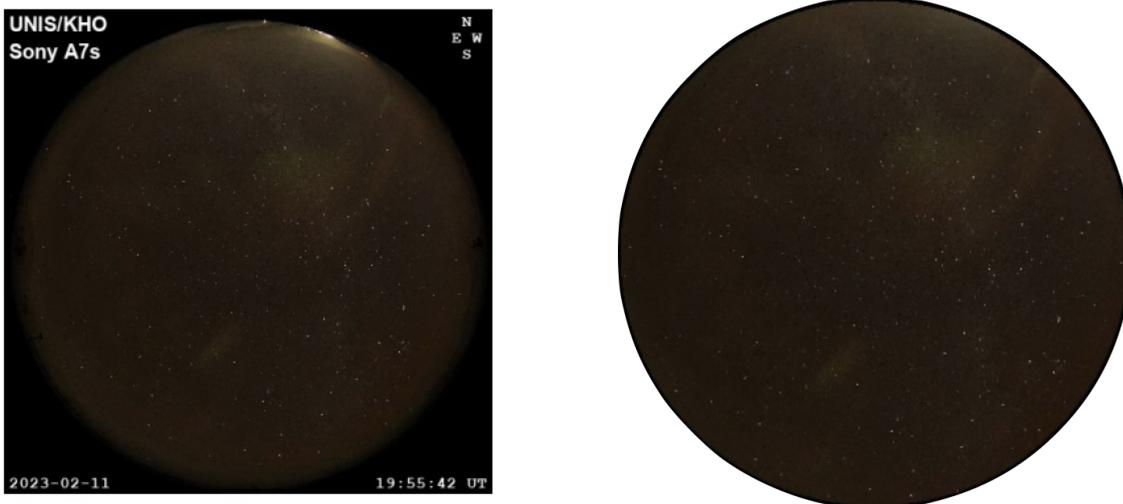


Figure 2.2.: Differences between unprocessed and processed image.

et al., 2009). The turbulence and shallow convection scheme are combined in an eddy-diffusivity mass-flux framework (Soares et al., 2004; Siebesma et al., 2007).

All operational model data of the Norwegian Meteorological Institute can be freely accessed

via [Norwegian Meteorological Institute \(2024\)](#). For the years 2021 - 2023, the periods of January - March and October - December are extracted and used for this study.

3. Cloud Coverage Retrieval

Extracting cloud information from all sky cameras has already experienced various approaches, of which a substantial portion are based on machine learning (Zhang et al., 2018; Kim et al., 2021; González-Fernández et al., 2024). González-Fernández et al. (2024) utilizes a convolutional neural network (CNN) to determine cloud coverage from all-sky cameras in Spain. Allowing for an offset of ± 1 okta in predicting cloud coverage, they gained an accuracy of above 93%. However, the camera providing the images for their study only produces daytime images, which is different from this study since only nighttime images exist.

Studies that utilize nighttime images of all-sky cameras to optimize the usage of telescopes by detecting clouds include Mommert (2020) and Zhong et al. (2024). The former compares a residual neural network (ResNet) with a gradient-boosted tree-based model (lightGBM) with an accuracy of 85% and 95%. The latter developed an eXtreme Gradient Boosting (XGBoost) with particle-swarm optimization (PSO) and achieved an accuracy of 96.91%.

In this study, we utilize two different methodologies to extract cloud cover information from nighttime all-sky images. The first one was developed by Alessandro Marocco and utilises a quasi-physics-based model to detect cloud coverage in okta (Marocco, 2025, personal communication Prof. Noora Partamies, UNIS). We refer to this as the "physical approach" (Section 3.1). The second one is based on machine learning, we refer to this as the "machine learning approach" (Section 3.2). Finally, both approaches are combined to achieve the best possible results, yielding the "combined approach" (Section 3.3).

The approaches mentioned above are benchmarked against each other. The cloud coverage distributions of all approaches are benchmarked with Makhotina et al. (2021), who finds that cloud coverage in the Arctic during polar night underlays a bimodal distribution of clear and completely cloudy. The intermediate cases occur much less than the edge cases, clear (0/8) and completely cloudy (8/8).

3.1. Physical Approach

The first approach to extract cloud coverage from all-sky camera images, by Alessandro Marocco, can be divided into two steps. First, a support vector machine (SVM) is applied to classify all images into 3 classes: clear, cloudy, and intermediate. SVMs work by finding boundaries in the dataset of images to separate these images into classes (Cortes and Vapnik, 1995). Only the intermediate cases get analyzed further, as clear and cloudy are classified as 0/8 and 8/8 (Figure 3.1). The intermediate cases are analyzed further by accounting for the image's brightness and colors, or by counting stars within a customized grid box.

The training dataset for the SVM ranges from November 2019 to February 2020 (polar

3. Cloud Coverage Retrieval

night only) and is labeled by me into clear, cloudy, and intermediate. The dataset consists of 2409 clear, 3496 cloudy, and 2177 intermediate images for the training (using 80%) and the testing (using 20%).

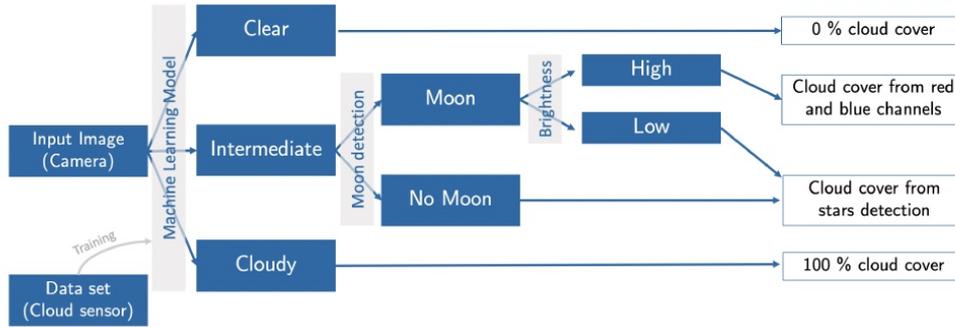


Figure 3.1.: The flow chart of the physical approach is created by Alessandro Marocco.

In a second step, the intermediate cases are analyzed based on an algorithm that first detects the moon’s position based on date and time. This gives indirect information about the image’s brightness. If the moon is 5° below the horizon, the image is considered to be dark, and a star detection algorithm is applied (Figure 3.1). The star detection algorithm lays a grid over the images. If there are two or more stars within a grid box (later changed to one), this part of the image is considered to be clear. Each grid box gets assigned cloudy or clear, and the sum of all grid boxes provides the cloud coverage of the image.

In the case of the moon being higher than 5° below the horizon, different steps are required to determine the cloud coverage. First, the overall brightness of the image is calculated by analyzing the mean of every pixel’s type of color, color intensity, and brightness of the color. If the brightness is lower than a threshold of 170 (later changed to 120), the star detection method is chosen. If the brightness is above the threshold, the image is too bright for stars to be seen. The cloud coverage is then determined by analyzing the red and blue channels of the image. Further details can be found in Marocco (2025).

Performance physical approach

All results and evaluations presented here are based on my own test cases and manually labeled references. First, the performance of the SVM is analyzed separately. Figure 3.2 visualizes the SVM results based on the manually labeled images from 2021 - 2023 (Chapter 2). The SVM can reliably discriminate among the three categories. Only 7 images are misclassified as clear or cloudy when they are labeled as cloudy or clear, respectively. When it comes to the intermediate cases, 9.9% of the clear images and 11.6% of the cloudy images are misclassified. Vice versa, 24.9% of the images depicting intermediate cloud cover get misclassified as clear, and 18.7% get misclassified as cloudy. The overall accuracy of the SVM based on equation

$$\text{Accuracy} = \frac{\sum_{i=0}^n C_{ii}}{\sum_{i=0}^n \sum_{j=0}^n C_{ij}} \quad (3.1)$$

n = number of classes

i, j = truth, prediction $\in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$

is 77.6%.

Figure 3.3a visualizes the performance of the physical approach. Displayed is a heatmap with the truth in okta on the x-axis and the prediction on the y-axis. The color intensity and the annotated values in each bin represent the number of images classified into that bin, i.e., 925 images are 8/8 and are predicted as such, while 147 images are 7/8 but are predicted as 8/8. If the physical approach were perfect, only the bins on the diagonal would be filled. On the top and the right are the marginal distributions.

The heatmap shows a recognizable diagonal, indicating that the physical approach manages to predict the cloud coverage well. However, there is an overprediction of 0/8.

The accuracy (Equation 3.1) is 63%. However, allowing for an offset of ± 1 okta

$$\text{Accuracy}_{\pm 1} = \frac{\sum_{i=0}^n C_{ii} + \sum_{i=0}^{n-1} C_{i,i+1} + \sum_{i=1}^n C_{i,i-1}}{\sum_{i=0}^n \sum_{j=0}^n C_{ij}} \quad (3.2)$$

n = number of classes

i, j = truth, prediction $\in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$

increases the accuracy to 86%. The mean distance to the diagonal

$$\text{Distance to diagonal} = \frac{\sum_{i=0}^n \sum_{j=0}^n C_{ij} |i - j|}{\sum_{i=0}^n \sum_{j=0}^n C_{ij}} \quad (3.3)$$

n = number of classes

i, j = truth, prediction $\in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$

is ± 0.66 okta. All values are listed in Table 3.1. Regarding cloud coverage in okta being subjective, as it is not a measurement, this offset is deemed acceptable. The marginal distribution of the prediction aligns with the U-shaped distribution of the observations, which is also found by Makhotina et al. (2021).

Nevertheless, there is room for improvement as González-Fernández et al. (2024), Momert (2020), and Zhong et al. (2024) achieved more accurate results.

3. Cloud Coverage Retrieval

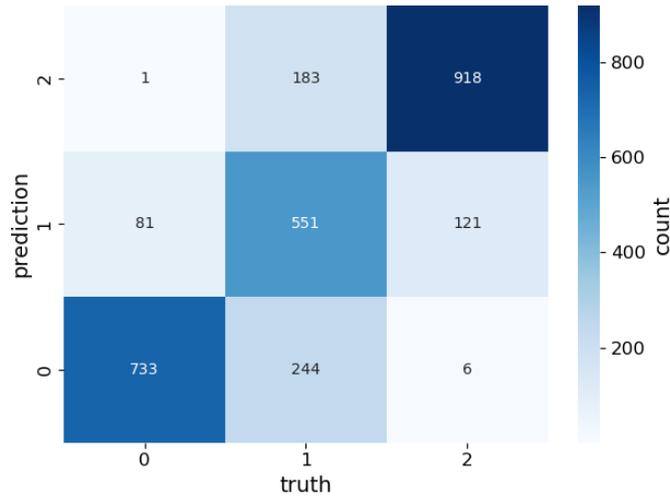


Figure 3.2.: Results of the SVM with the truth on the x-axis and the prediction on the y-axis. 0 = clear, 1 = intermediate, 2 = cloudy. The colors and numbers indicate the number of classified images.

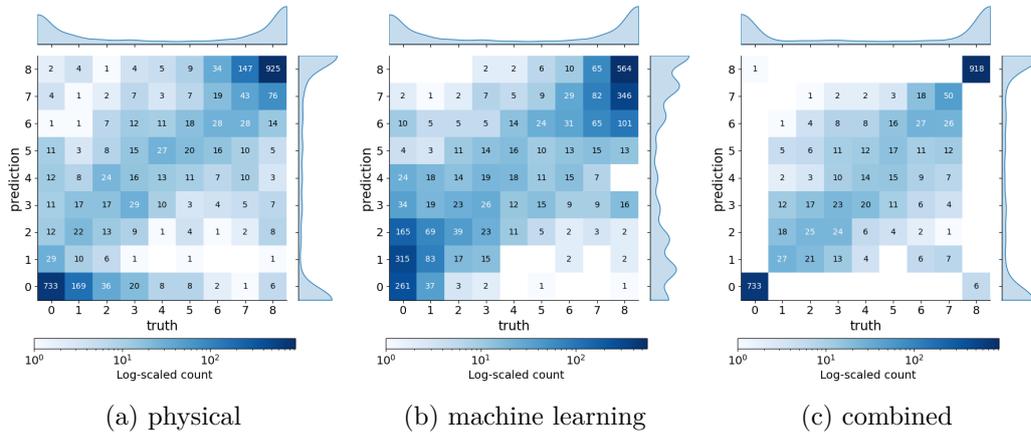


Figure 3.3.: Results of all approaches to extract cloud coverage represented as a heat map with the truth in okta on the x-axis and the prediction in okta on the y-axis. The colors and numbers indicate the number of classified images. At the top and the right of each heat map are the corresponding marginal distributions.

3.2. Machine Learning Approach

The following section introduces a machine learning approach with the idea of outperforming the physical model. When it comes to image classification, a machine learning model learns to recognize patterns in images and assigns them to categories like types of animals or, in this case, cloud coverage in okta. The images are converted into numbers representing pixel values, which are then processed through layers of connected nodes. These layers are, among other things, responsible for finding patterns. During training, the model adjusts the connections of the nodes to improve the prediction. Once trained, it can classify new images that are independent of the training dataset. Special types of machine learning models that are often used for image classification are CNNs and ResNets. For the machine learning approach, a ResNet (personal communication Are Frode Helvig Kvanum, MET Norway) is chosen.

Details about ResNets

Before going into the details of a ResNet, a CNN is explained first. A CNN consists of multiple layers. There are convolutional layers that apply filters to find patterns, activation functions allowing the CNN to learn more complex patterns, pooling layers reducing the spatial resolution to make the CNN more robust, and fully connected layers to combine all the learned features (LeCun et al., 1998). A ResNet, which is used in this study, is a special form of a CNN. In addition to all these layers of a CNN, a ResNet includes the so-called skip connections. These skip connections have the advantage that, unlike in a CNN, the input is not directly mapped to the output. Instead, the ResNet only learns the residual, thus, the difference between the input and the output. This provides the opportunity to make a ResNet much deeper and more trainable than a CNN (He et al., 2015). The deeper a CNN, the more likely it can be subject to the vanishing gradient problem, which the skip connections in a ResNet can alleviate (Bengio et al., 1994; He et al., 2015). The gradients are responsible for adjusting the connections of the nodes. If these gradients converge to zero, the network is not trainable anymore.

The architecture, the combination of the different layers, is taken from the outstanding work of He et al. (2015). Additionally, there is a set of important parameters influencing the performance of the ResNet. These parameters are not directly taken from He et al. (2015). They are subject to a hyperparameter tuning to find their best combination. Parameters of interest are the number of epochs (number of training loops), which is tested between 20 and 400 epochs. As a loss function the cross-entropy loss (pyt, 2025c, Equation 3.4) and the Focal loss (Lin et al., 2017, Equation 3.5) are tested. The Adam optimizer (pyt, 2025a, Equation 3.6), and the AdamW optimizer (pyt, 2025b, Equation 3.7) are tested, as well as different learning rates (10^{-5} , 10^{-4} , 10^{-3}). The usage of learning rate decay was also tested. Additionally, different batch sizes of 16, 32, and 64 were tested. The loss function measures how far off the ResNet's prediction is from the truth, whereby the type of loss function can be chosen. The Cross-Entropy loss function

$$L = - \sum_{i=1}^N y_i \log(p_i)$$

N = number of classes (3.4)

y = truth

p = predicted probability for class i

and the Focal loss function

$$L = - \sum_{i=1}^N y_i \alpha_i (1 - p_i)^\gamma \log(p_i)$$

N = number of cases (3.5)

y = truth

p = predicted probability for class i

α = class weighting factor

γ = focusing parameter

3. Cloud Coverage Retrieval

are two possible options. The Focal loss is a modified version of the Cross-Entropy loss, which takes imbalances in a dataset into account. This is done by weighting classes according to their prediction difficulty (Equation 3.5, $\alpha = 1$ in this setup). γ (Equation 3.5) equals 2 in this setup and can be interpreted as reducing the loss contribution of easily classified examples. It draws the focus to difficult cases.

The optimizer adjusts the ResNet’s weights according to the loss, using gradients to update the weights in the direction of decreasing loss. In the case of the Adam optimizer, the gradient descent is calculated as follows

$$\begin{aligned}
 \theta_{t+1} &= \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}} \\
 \hat{m}_t &= \frac{m_t}{1 - \beta_1^t} \\
 \hat{v}_t &= \frac{v_t}{1 - \beta_2^t} \\
 m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\
 v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \\
 g_t &= \nabla_{\theta} L_t.
 \end{aligned} \tag{3.6}$$

α is the learning rate, m_t and v_t denote the first and second moment (exponentially weighted mean of past and square gradients). β_1 and β_2 are the exponential decay rates. ε is a small constant for numerical stability, and Θ_t denotes the model parameters at iteration t . The AdamW optimizer is calculated following

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}} - \alpha \lambda \theta_t \tag{3.7}$$

with λ being the weight decay coefficient. Compared to the Adam optimizer, AdamW is more stable during the training process and improves generalization by applying weight decay separately from the gradient update, ensuring proper regularization of the model’s weights (Loshchilov and Hutter, 2019).

The learning rate quantifies how much the weights and biases in the ResNet are adjusted to continue decreasing the loss. The decaying learning rate works as follows. If the validation loss does not improve for 5 consecutive epochs, it is multiplied by 0.5.

The batch size controls how many training examples are seen by the ResNet before the weights and biases are updated.

The hyperparameter tuning revealed that the best set of parameters is 200 epochs, the Focal loss, the AdamW optimizer, a decaying learning rate with an initial value of 10^{-4} , and a batch size of 16.

It is important to note that this ResNet needed a substantial amount of memory and was trained on the supercomputer from Deutsches Klimarechenzentrum (DKRZ) (2025) using GPU acceleration.

The training of the ResNet

The images from 2019 to 2023 are split into testing, training, and validation datasets. Images covering the period from 2019 to 2020 are taken as training and validation data, and images covering the period from 2021 to 2023 are taken as testing data. The split is chosen to make the results comparable to the physical approach. This is one reason why the split is not chosen to be random, as is usually done. Another reason is that some images are taken shortly after one another. A random split could cause the testing data to be closely related/correlated to the training data, which would falsify the results. The training and validation datasets are balanced over all oktas to avoid the ResNet favoring certain cases, which leaves 80 images per case for the training and validation datasets. The data is randomly split into training and validation subsets in an 80:20 ratio, while maintaining a balanced distribution of cloud coverage cases in both subsets.

Performance machine learning approach

Figure 3.3b shows the results of the fully machine learning based approach. A striking aspect is the marginal distribution of the prediction shown on the right of the panel. It disagrees with the common marginal distribution of the truth that both edge cases occur most frequently (Makhotina et al., 2021, marginal distribution of the truth Figure 3.3b). While 8/8 okta is still predicted most frequently, the occurrence of 0/8 okta is strongly reduced, with 1/8 okta being predicted more often. Furthermore, the intermediate part is predicted more often compared to the testing data (truth). While the main diagonal is clearly visible, a substantial distance to the diagonal of ± 1.2 okta is present (Equation 3.3). The predicted values exhibit a reduced slope relative to the diagonal, resulting in an overestimation of low oktas and an underestimation of high oktas. Clear-sky conditions (0/8) are systematically overpredicted: when the true cloud coverage is 0/8, predictions of 1/8 occur more frequently than correct 0/8 classifications. Conversely, fully overcast conditions (8/8) are underpredicted. This leads to a performance decrease in predicting the edge cases compared to the physical approach (Figure 3.3a).

The accuracy of the machine learning based approach is 39% and thus, worse than for the physical approach. Allowing for an offset of ± 1 okta increases the accuracy to 77%, which is still lower than the physical one (Table 3.1).

3.3. Combined Approach

The machine learning approach does not outperform the physical method. However, since the physical approach's SVM performs well and the ResNet appears to grasp the underlying relationship (see the heatmap in Figure 3.3b), a combination of the two methods may yield the most accurate results and potentially outperforms the physical model. This approach might not seem promising, since the physical approach performs better than the machine learning approach. I gave it a shot regardless and ended up outperforming the physical approach.

In this combined approach, the SVM from the physical approach is applied first to classify the images into clear, intermediate, and cloudy. If an image is classified into clear or

3. Cloud Coverage Retrieval

cloudy, the cloud coverage extraction is done with being set to 0/8 or 8/8 okta, respectively. However, if an image is classified as intermediate, a modified version of the ResNet is used. This ResNet classifies the intermediate images into cloud area fractions between 1/8-7/8. Here, the new ResNet differs from the previously employed version (see Section 3.2) as it now only classifies into 7 classes instead of 9. Hence, the ResNet needs to be retrained.

For the retraining, the SVM's predictions are prioritized over the ResNet's, such that images labeled as intermediate by the SVM are excluded from being classified as either clear or cloudy. This differs from the physical model, where images initially classified as intermediate by the SVM could still be subsequently labeled as clear or cloudy. Due to the ResNet focusing on only the intermediate cases (1/8 -7/8), cloud-free and fully cloud-covered cases are omitted from the training. This is why 81 images (clear but predicted as intermediate) and 121 images (cloudy but predicted as intermediate, Figure 3.2) are excluded from the testing data of the 7-class ResNet. Another reason why a 7-class and not a 9-class ResNet is chosen for the combined approach is that not many images labeled as 0/8 and 8/8 are expected to be part of the images that are predicted as intermediate by the SVM. This would cause the training dataset for a 9-class ResNet to be unbalanced and too small.

Another practical reason to remove images that are neither labeled nor predicted as intermediate is the whole idea of combining both methods and testing a new approach to outperform the physical model. The physical approach that is not the SVM also only deals with the intermediate images since the edge cases are handled by the SVM (Figure 3.1).

Performance of the combined approach

The performance of the combined approach is depicted in Figure 3.3c. As already found by Makhotina et al. (2021) in the Arctic during polar night, the marginal distribution of the prediction favors the edge cases 0/8 and 8/8. Furthermore, the diagonal is nicely visible. By giving the SVM more weight, the overall performance of the combined approach outcompetes the physical approach (Figure 3.3a). Overpredictions of 0/8 are only possible if the SVM predicts 8/8, and underpredictions of 8/8 are only possible if the SVM predicts 0/8. Another interesting aspect is that the 7-class ResNet in the combined approach does not struggle as much with its edge cases 1/8 and 7/8 as the 9-class ResNet in the machine learning approach. These improvements are reflected by an increased accuracy compared to the physical and machine learning approaches. The accuracy with zero tolerance is 83%, and the accuracy for a tolerance of ± 1 okta is 92%. The distance to the diagonal is with ± 0.32 okta smaller than for the other approaches.

All accuracies and distances to the diagonal are listed in table 3.1. The combined approach performs best with a considerable margin, followed by the physical and machine learning approach. Hence, the combined approach is used in the following chapter to generate a cloud product used for validating the operational AROME-Arctic NWP model.

Table 3.1.: Accuracies and distances to the diagonal of the different cloud detection methods.

	accuracy		distance to diagonal
	truth = pred	truth = ± 1 okta pred	
physical	63%	86%	0.66 okta
machine learning	39%	77%	1.2 okta
combined	83%	92%	0.32 okta

4. AROME-Arctic Evaluation

We now use the developed combined approach to evaluate the operational NWP model AROME-Arctic. If not stated otherwise, all model investigations use the first model output step of AROME-Arctic (lead time of 0h). Furthermore, only the cloud coverage above 500 m is considered due to the camera being located 520 m.a.s.l..

The evaluation is split into two parts. First, a general statistical overview of the model's performance is generated. Then, a selection of cases where model errors exceeded 5 okta are discussed individually.

Keeping in mind that the combined approach is not ideal, especially not in the intermediate part, the model evaluation in the second part focuses mainly on problematic edge cases, completely cloudy forecasted as completely clear and vice versa. These are also the forecast misses that can be considered most severe for operational weather forecasting.

4.1. Statistical Evaluation

Before the performance of AROME-Arctic can be evaluated, an actual cloud coverage value comparable to the estimated truth needs to be calculated. The cloud coverage in AROME-Arctic is defined between 0 and 1. Every grid box in every level contains a cloud area fraction value. To estimate total cloud cover in a way that matches what a meteorologist or a ground-based cloud camera would see looking upward, a cloud overlap scheme needs to be chosen to calculate the total cloud cover for a model grid column. The most common schemes are the maximum overlap -, random overlap -, and maximum-random overlap scheme (Figure 4.1), which are used by i.e. [Arbizu-Barrena et al. \(2015\)](#) and [Illingworth et al. \(2007\)](#). They are defined as follows:

- **Maximum cloud overlap scheme:** Assuming maximum overlap of all cloud fractions in one column by simply taking the largest cloud area fraction of a column (Figure 4.1a). In the example, the highest cloud coverage is 5/8, thus the total cloud coverage is 5/8.
- **Random cloud overlap scheme:** Assuming the vertical overlap to be random (Figure 4.1b) with the following equation

$$C_{\text{rand}} = 1 - \prod_{i=1}^n (1 - c_i) \quad (4.1)$$

n = number of layers

c = cloud fraction $[0, 1]$.

In the example n is 8 and c_1 to c_8 are $3/8$, $4/8$, $0/8$, $2/8$, $4/8$, $1/8$, $5/8$, and $1/8$, respectively. This results in $C_{\text{rand}} = 0.97$, which translates to $8/8$.

- **Maximum-random cloud overlap scheme:** Assuming maximum overlap if the cloud fraction is non-zero for adjacent vertical layers. If there is more than one contiguous block of clouds, these blocks are assumed to overlap randomly (Figure 4.1c; Equation 4.1, n being the number of cloud blocks and c being the cloud fraction of a cloud block based on maximum overlap). In this example, $n = 2$, $c_1 = 4/8$, and $c_2 = 5/8$. This results in $C_{\text{rand,max}} = 0.8125$, which translates to $7/8$.

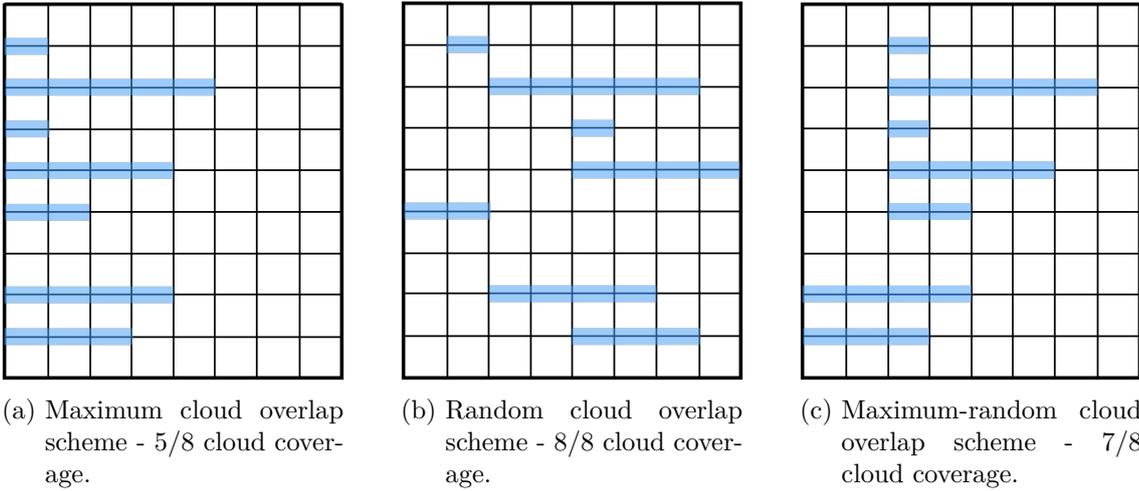


Figure 4.1.: Sketches of different cloud overlap schemes of the same modeled situation.

The different cloud overlap schemes can be sorted by the amount of cloud coverage they see

$$C_{\text{max}} \leq C_{\text{rand,max}} \leq C_{\text{rand}}. \quad (4.2)$$

Hogan and Illingworth (2000), who conducted a radar-based cloud overlap study in southern UK during winter, found that vertically continuous clouds overlap more randomly than assumed. Moreover, the degree of overlap depends on the type of clouds.

Since our study is conducted in the Arctic during polar night, it is, to some extent, comparable to Hogan and Illingworth (2000). Thus, the random cloud overlap scheme is chosen for this study. The impact on the forecast performance of using the other overlap schemes is investigated as well.

Similar to the overlap schemes, a spatial domain that maps the camera's field of view into AROME-Arctic needs to be chosen. One can find a multitude of approaches for mapping

the observation of cloud coverage into NWP models in recent studies. Hogan et al. (2001), who use ground-based radar and lidar data, and Arbizu-Barrena et al. (2015), who use ceilometer and all-sky camera data, both map the observed value on the nearest pixel within the model. Hogan et al. (2009) and Illingworth et al. (2007) use a more sophisticated method taking into account the horizontal extent of the cloud and its movement, which are both mapped into the model of interest. Crocker and Mittermaier (2013) compare using a pixel with using a cylinder of 30 km radius in der model.

Since the all-sky camera in this study has a cone-shaped equidistant field of view, at an altitude of 5 km the radius already exceeds 50 km (Chapter 2). This and Crocker and Mittermaier (2013) cylinder approach motivates the use of a cylinder with a radius of 30 km to map the camera’s field of view into AROME-Arctic (Figure 4.2 red circle). The impact on the forecast performance of using only the pixel closest to the camera’s location to map the observations into AROME-Arctic is investigated as well (Figure 4.2 black pin).

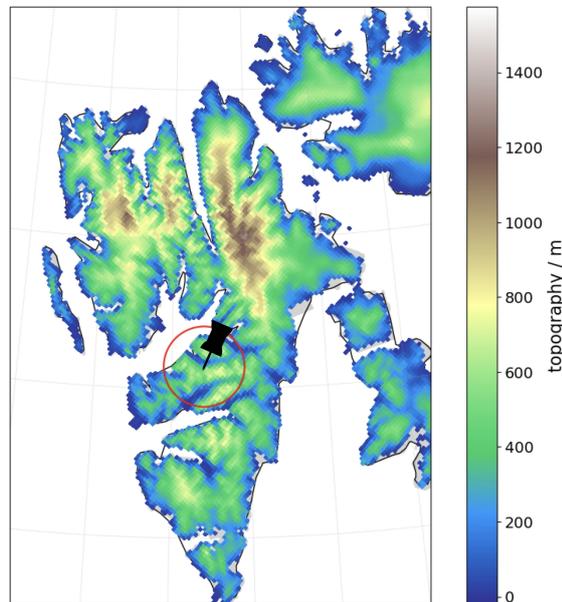


Figure 4.2.: Sketch to show pixel (black pin) and cylinder (red circle) in AROME-Arctic.

Overlap Scheme and Spatial Domain Analysis

The usage of a cylinder together with the random overlap scheme yields the results seen in Figure 4.3a. A perfect model only has hits on the diagonal, which is not the case for AROME-Arctic. Nevertheless, there are accumulation points at 0/8 (99 cases) and 8/8 (262) that show that AROME-Arctic manages to predict the majority of the edge cases correctly. However, the truth of 0/8 is sometimes overpredicted as intermediate and even 8/8 (left column in Figure 4.3a), vice versa, the truth of 8/8 can be underpredicted (right column in Figure 4.3a). AROME-Arctic tends to overpredict the intermediate part as it favours 8/8 for the truth being 1-7/8 (upper row/frame in Figures 4.3a). Based on this analysis, the performance of AROME-Arctic only shows little skill in predicting cloud coverage. Later, the case studies (Section 4.2) reveal further details in favor of AROME-Arctic.

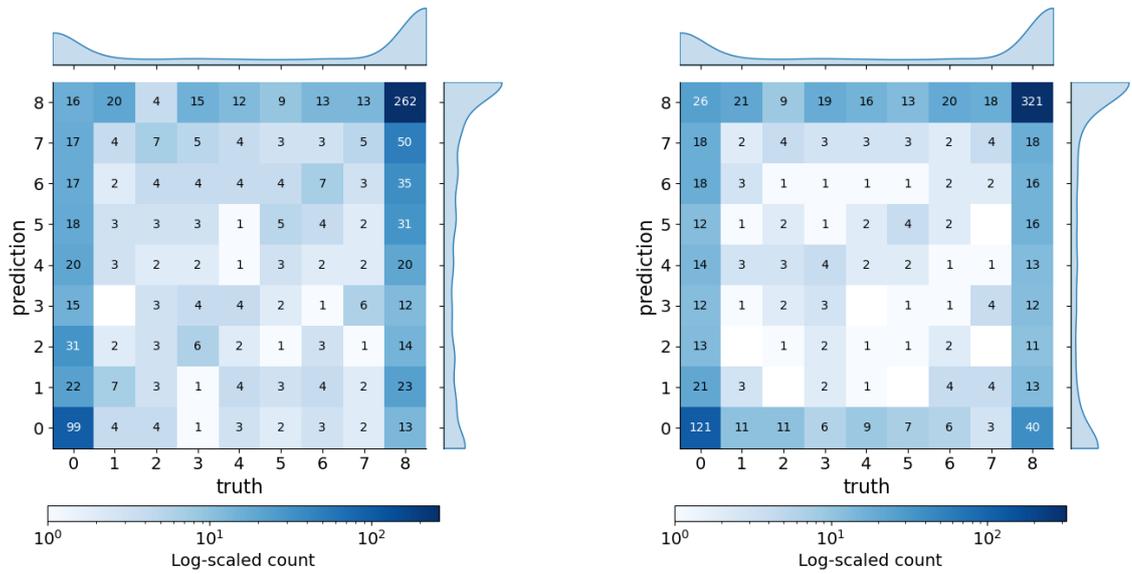
4. AROME-Arctic Evaluation

As mentioned before, the impact on AROME-Arctic's forecast performance using a pixel as a different spatial domain and using different cloud overlap schemes is analyzed in the following paragraph. Figure 4.3b shows the random overlap scheme in combination with a pixel. The frame-like distribution favouring the edge cases is even more pronounced than for the cylinder (Figure 4.3a). Furthermore, the intermediate cases are subject to error compensation. If the truth is intermediate, AROME-Arctic favours 0/8 or 8/8 and hardly predicts 1-7/8. This can be explained by the pixel being blind to its surroundings, which is especially problematic if the overall situation is intermediate, as the part seen by the pixel is either cloudy or clear. If the truth is 0/8, AROME-Arctic favours 8/8 as the second priority and predicts 1-7/8 equally often. Vice versa, if the truth is 8/8, AROME-Arctic predicts 0/8 as the second priority and 1-7/8 equally often as the third priority.

Interestingly, considering the marginal distributions (Figure 4.4) the pixel with the random cloud overlap scheme seems to perform best compared to all other compositions of spatial domain and overlap scheme. This is a direct consequence of the error compensation done by the pixel. Thus, only considering the mean values over the investigated period can hide very important details and falsify the believed results.

Figure 4.4 also includes the marginal distributions of the cylinder and the pixel in combination with the maximum and the maximum-random overlap scheme. It is striking that the cylinder with the maximum and the maximum-random overlap scheme, compared to the cylinder with the random overlap scheme, systematically overestimates the intermediate cases and 0/8, and drastically underestimates 8/8. This is not surprising as the maximum, and the maximum-random overlap scheme show lower values than the random overlap scheme (Figure 4.1, Equation 4.2). Similar arguments hold for the pixel with the maximum and the maximum-random overlap scheme compared to the pixel with the random overlap scheme. They overestimate the intermediate cases and 0/8 and underestimate 8/8.

To sum up the different methods to calculate cloud coverage in AROME-Arctic, it can be said that the cylinder with the random overlap scheme provides the most realistic results. However, choosing a different composition of spatial domain and overlap schemes does not favour the prediction quality of AROME-Arctic as they all show a similar trend in overpredicting the intermediate cases and underpredicting the edge cases.



(a) Cylinder in AROME Arctic in which the cloud coverage is calculated.

(b) Pixel in AROME Arctic in which the cloud coverage is calculated.

Figure 4.3.: Joint probability distribution of the true cloud coverage okta on the x axis and the cloud coverage prediction by AROME-Arctic on the y axis. Both the color and the numbers in the joint distribution indicate the number of corresponding cases. At the top and the right are the marginal distributions of the truth and the prediction. Figure a) uses a cylinder with a radius of 30km around KHO, and Figure b) uses the model grid box closest to KHO.

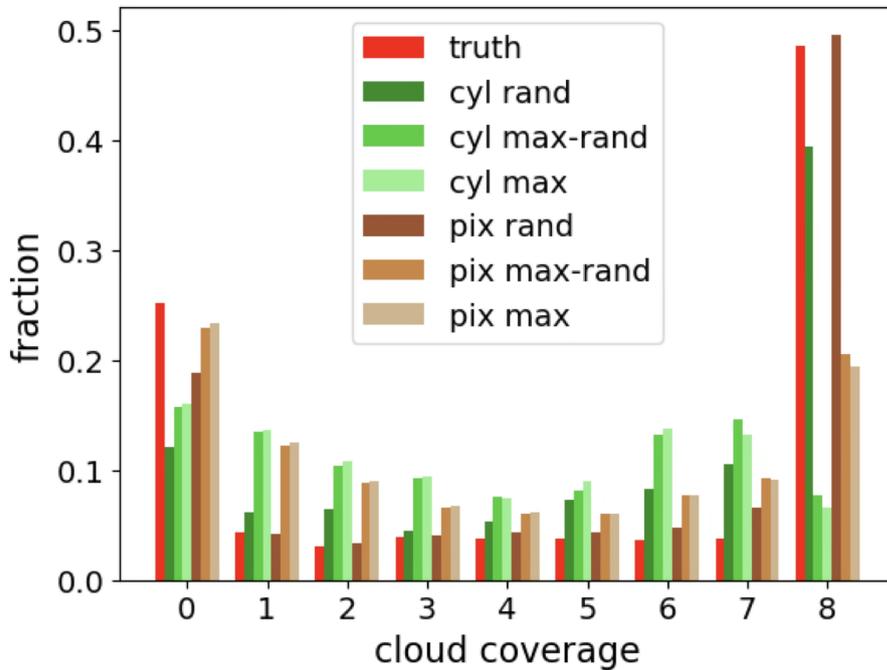


Figure 4.4.: Marginal distributions of the truth (red), the prediction by AROME Arctic based on a cylinder (green), and a pixel (brown). Both spatial domains are considered for all three cloud overlapping schemes. The x-axis shows the cloud coverage in okta, and the y-axis shows the occurrence probability for the different okta.

Lead Time Analysis

This section utilizes the cylinder in combination with the random cloud overlap scheme to analyze the cloud cover predictions by AROME-Arctic across multiple lead times. Figure 4.5 gives insights into the forecast performance with respect to increasing lead times. The Pierce Skill Score

$$PSS = H - F \quad (4.3)$$

H = hit rate

F = false alarm rate,

which is the difference between the hit rate (H)

$$H = \frac{a}{a + c} \quad (4.4)$$

a = hits "almost clear"

c = overprediction

and the false alarm rate (F)

$$F = \frac{b}{b + d} \quad (4.5)$$

b = underprediction

d = hits "clouds present",

is chosen to quantify the skill. The binary definition of H and F results in a 2x2 contingency table. Therefore, the 9 value-based cloud coverage value is simplified and separated into "almost clear" (0-1/8) and "clouds present" (2-8/8) (Table 4.1). Other separations were tested as well, but did not provide any additional information.

The results in Figure 4.5 show that there is no trend visible for increasing lead times. AROME-Arctic shows some skills across all lead times as all points in Figure 4.5 are located left of the diagonal. This shows that AROME-Arctic has some skill even if it is not very high, as discussed with Figure 4.3a. However, the PSS might not be the most suitable skill score to evaluate AROME-Arctic's forecast quality.

The findings from this section are set into perspective by the next Section 4.2, which reveals further insights into AROME-Arctic's forecast quality.

Table 4.1.: 2×2 contingency table for cloud coverage prediction in AROME-Arctic.

	Observation (0–1/8)	Observation (2–8/8)
AROME-Arctic (0–1/8)	a) hits "almost clear"	b) underprediction
AROME-Arctic (2–8/8)	c) overprediction	d) hits "clouds present"

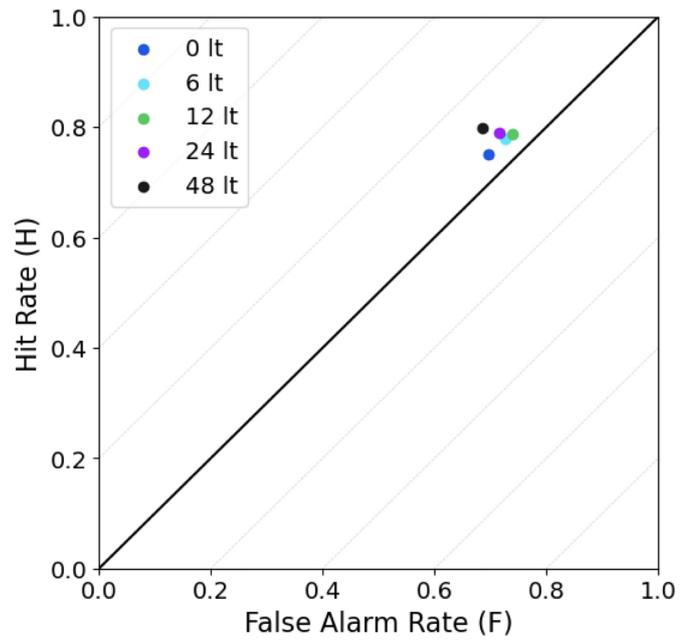


Figure 4.5.: Pierce Skill Score (PSS) for increasing lead times, separated between clear (0-1/8) and cloudy (2-8/8). The best skill would occur if the values were located in the top left corner. The diagonal itself has no skill, and to the right of the diagonal stands for worse than a random model.

4.2. Case Studies

The second part of the AROME-Arctic evaluation analyzes case studies. These case studies focus on situations where deviations exceeding 5 oktas between the estimated truth from the all-sky images (combined approach, Section 3) and the modeled cloud area fraction occurred. Analyzing these cases revealed three particular situations that led to large systematic errors with high deviation in AROME-Arctic (prediction) and the cloud camera (truth). One being cold air outbreaks with high cloud overprediction (case 1). Second, there are cold air outbreaks with too effective cloud shadowing downstream of Isfjorden or with too weak cloud formation over Isfjorden (case 2). Case 3 shows a displacement problem during which AROME-Arctic predicts the cloud coverage well, but in the wrong location. An additional Case can be found in the Appendix B.

4.2.1. Cold Air Outbreaks

More than 50% of the cases that depict a difference of at least 5 okta between the observed and modeled cloud cover occur during cold air outbreaks (CAO). CAOs can be connected to high-impact weather events in the Arctic associated with huge waves, maritime icing, and polar lows (Businger and Reed, 1989; Papritz and Pfahl, 2016; Samuelsen and Graversen, 2019). A CAO occurs when cold, dry air, originating over cold surfaces like sea ice or snow-covered land, is advected over a much warmer ocean surface. The strong air-sea temperature contrast leads to increased upward sensible and latent heat fluxes of up to 1000 Wm^{-2} (Grossman and Betts, 1990). As a result, the initially cold, dry, and stably stratified boundary layer rapidly transforms into a warmer, more humid, and unstable boundary layer. This instability enhances convection and vertical mixing, promoting cloud formation that often organizes into characteristic cloud streets (Hartmann et al., 1997; Etling and Brown, 1993). Once these air masses flow over land, they are usually subject to cloud shadowing. Clouds are blocked by topography, typically rain out, or dissolve over land due to increased stability.

A typical method to quantify the intensity of a CAO is the CAO index, which is defined as the potential temperature difference between the surface and 900 hPa. The CAO index is also used by related studies (Papritz and Sodemann, 2018; Kähnert et al., 2021).

Cold Air Outbreaks - high cloud overprediction

The first case we discuss here occurred on 2022-12-25 12:00 UTC and depicts a problem of overpredicting high clouds during a CAO. The majority of the CAO cases, during which the cloud coverage is overpredicted shows this behaviour. During the event, Svalbard is completely covered with clouds in AROME-Arctic (Figure 4.6a) while the camera evidences clear-sky conditions (Figure 4.7). The low cloud area fraction (below 500 m, Figure 4.6b) indicates weak cloud streets west of Svalbard, which are typical for CAOs (Etling and Brown, 1993; Hartmann et al., 1997; Kähnert et al., 2021). The cloud coverage above 500 m west of Svalbard intensifies directly after the air is blown across the sea ice edge (Figure 4.6a red line indicating the sea ice edge and arrows indicating the wind speed and direction). Simultaneously, the CAO index increases, already before the sea ice edge

is passed, due to leads in the marginal ice zone. The CAO index reaches a maximum of 16 K. Latent and sensible fluxes increase to more than 800 W m^{-2} downstream of the sea ice edge (Figure 4.6c/d). Boundary layer clouds do not overspill Svalbard during the CAO as the colder land masses hempen convection and cause cloud dispersion. Figure 4.6b shows that the low clouds quickly dissolve over land.

A north-to-south cross section (yellow line in Figure 4.6a) reveals insight into the vertical cloud structure (Figure 4.8a). The clouds responsible for the overprediction are not boundary layer clouds but relatively thin, high-level clouds at around 9 km. These clouds are not connected to the boundary layer processes triggered by the CAO. An animation (Nellesen, 2026a) showing the next hourly time steps of the same model run reveals that the thin layer of clouds at a height of 9 km vanishes in the following time steps and gets replaced by a thicker layer of lower clouds, which is also independent of the CAO as they do not originate from north of Svalbard. Moreover, they are at heights above 4 km which would be very high for Arctic boundary layer clouds. This finding might hint towards an initialization problem in AROME-Arctic of the analyzed model run. However, the model run three hours prior shows that these high clouds are already predicted (Figures A.1). Figure A.1d shows the model run of 2022-12-25 09:00 UTC with a lead time of 3 hours, coinciding with the time of the initial model run. The same thin cloud layer at 9 km is present. This earlier model run partially feeds the model run 3 hours later at 12:00 UTC. Another explanation for these high clouds could be an error in the European Center for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS), which serves as boundary conditions for AROME-Arctic (Müller et al., 2017). The source for these high clouds can either be found in AROME-Arctic at even earlier lead times or in the IFS. Despite the high clouds, AROME-Arctic captures the boundary layer processes above Svalbard during the CAO adequately.

While not relevant for the found overestimation during this case another interesting model error is revealed by the west to east cross section (green line in Figure 4.6a, Figure 4.8b) showing convective cells west of Svalbard which are the cloud streets seen in Figure 4.6b. These convective cells are surrounded by a homogeneous cloud layer masking the mesoscale organization when evaluating, e.g., cloud cover or outgoing longwave radiation. The exact cause of this phenomenon in the model is not yet clear. Entrainment of water vapor from the cloud cells into the cloud-free areas takes place. Instead of being transported downwards with the vertical velocity, this water vapor is converted to cloud ice instead by the microphysics. As a consequence, the downdraft regions are being filled by clouds as well. Hence, while the model captures the atmospheric dynamics correctly (cloud streets here), investigating the cloud area fraction would give the impression that the model would erroneously simulate a stratus layer instead (personal communication Marvin Kähnert).

The brightness temperature based on remote sensing by satellites (Figure 4.9) underlines the existence of the typical cloud streets west of Svalbard. This shows that AROME-Arctic shows the correct dynamical processes for low clouds. At the location of the camera, it is difficult to differentiate whether the low brightness temperature originates from a cold surface or from clouds. However, given that the entire area of Svalbard shows the same homogeneous brightness temperature that differs from the surrounding ocean, it is likely that no clouds are present. However, this has to be taken with care as remote sensing

4. AROME-Arctic Evaluation

observations have known weaknesses like detecting clouds in high latitudes during polar night (Liu et al., 2004).

To sum up the first case. Overall, AROME-Arctic gets the CAO above Svalbard correctly but predicts too many high clouds within the whole domain (Figure 4.6a and 4.8a), which causes the erroneous cloud coverage at the camera's location. Although the high clouds coincide with the CAO in time, they seem to be independent of the CAO and potentially originate from the IFS model.

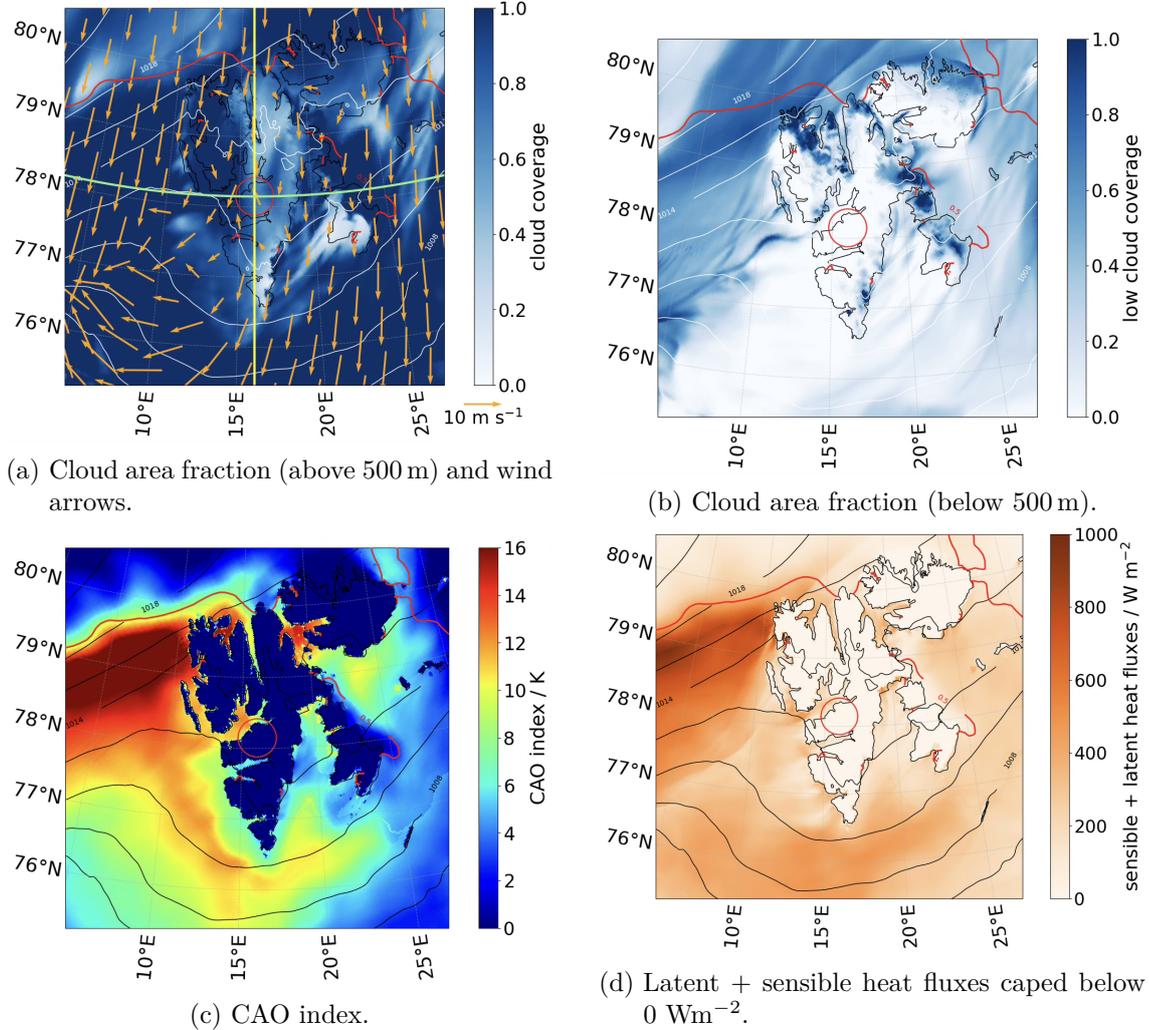


Figure 4.6.: Case 1: 2022-12-25 12:00 UTC. The red line indicates the area where the sea ice concentration exceeds 50% of the grid box, and the red circle indicates the camera's area of sight in AROME Arctic. The pressure lines at sea level are shown in white (upper panels) or in black (lower panels).

Cold Air Outbreaks - missing cloud formation over Isfjorden

The second case (Figure 4.10) on the 2023-11-15 03:00 UTC shows a CAO with similar boundary processes as in case 1. The cloud coverage intensifies after being advected over the ocean, and the clouds disappear again with landfall (Figure 4.10a and 4.10b). Compared to case 1, the boundary layer needs more time to build up after the air has flown over the sea ice edge. This can be seen by the cloud-free gap downstream of the sea ice edge (Figure 4.10a). Figure 4.10b shows the, for a CAO, typical cloud streets that are



Figure 4.7.: Case 1: 2022-12-25 12:00 UTC. All-sky camera at KHO at 11:34 UTC showing clear sky conditions, and stars are visible.

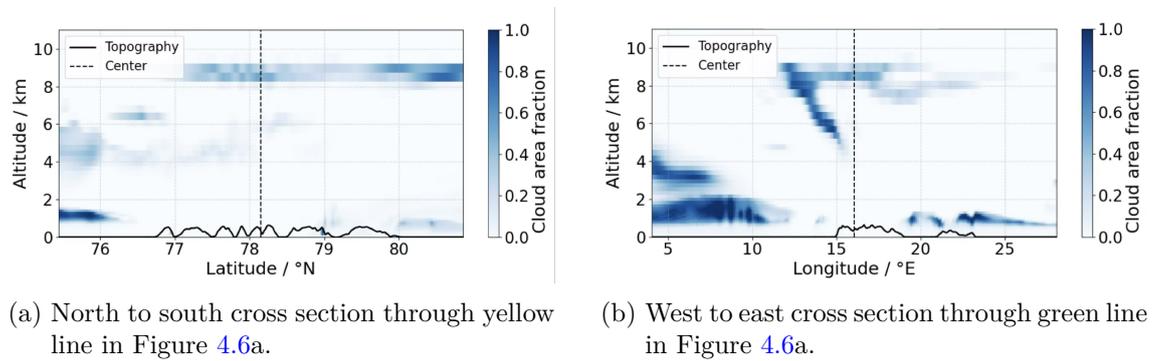


Figure 4.8.: Case 1: 2022-12-25 12:00 UTC. The cross sections show the vertical structure of the clouds. The dotted vertical line shows the location of the camera, and the solid line at the bottom indicates the topography.

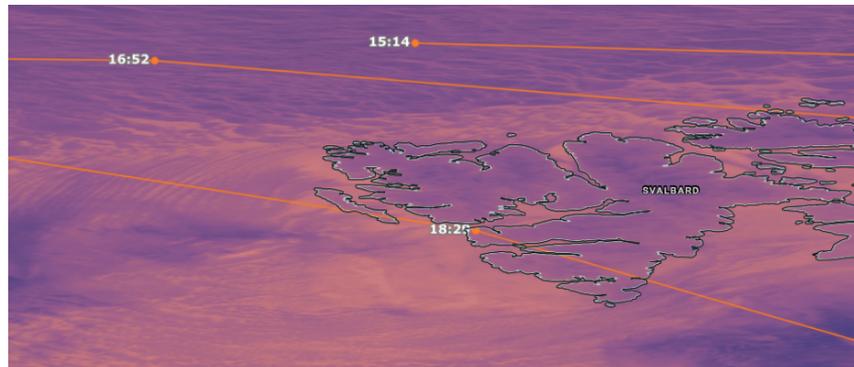


Figure 4.9.: Case 1: 2022-12-25 12:00 UTC. Brightness temperature (?) visualized by [NASA Earth Observing System Data and Information System \(EOSDIS\) \(2024\)](#) at 2022-12-25 16:52 and 18:20 UTC. The darker the color, the colder the brightness temperature. Purple indicates cold surfaces like clouds, sea ice, and snow.

also seen in case 1. The CAO index only increases downstream of the sea ice edge and not, like in case 1, before it is passed (Figure 4.10c). The latent and sensible heat fluxes (maximum $< 700 \text{ Wm}^{-2}$, Figure 4.10d) indicate that the CAO is not as strong as in case 1.

Despite the similarities to case 1, the error in cloud coverage during this case is opposite. AROME-Arctic predicts the CAO and the boundary layer conditions fairly well. Only

4. AROME-Arctic Evaluation

the cloud formation over Isfjorden is not captured. The truth (Figure 4.13) shows complete overcast, while AROME-Arctic predicts cloud-free conditions over Svalbard (Figure 4.10a). As the air flows over Isfjorden, the situation resembles a miniature CAO, with cold air from Svalbard advected over the warmer waters of the fjord. (see CAO-index Figure 4.12b). Likewise, clouds start to form over Isfjorden just before they get advected towards its southern coast (Figure 4.12a). They vanish shortly after landfall. The all-sky camera (Figure 4.13) shows overcast conditions and thus suggests that clouds form over Isfjorden and do not dissolve immediately after landfall.

A possible explanation is that Isfjorden is not resolved in IFS, which, among others, initializes AROME-Arctic's sea surface temperature. This could cause an excessively cold Isfjorden, which hampers convection and thus cloud formation. Another possible explanation might be channeling through the valley away from the camera towards Isfjorden. This potentially prevents the clouds in AROME-Arctic from being transported to the camera's location. A possible solution could be the 500 m resolution model AROME-Svalbard (Valkonen et al., 2020). Mack et al. (2025) has shown that AROME-Svalbard represents channeling through valleys better than AROME-Arctic.

Similar to case 1, a stratus layer is seen west of Svalbard. The lower layers show cloud streets (Figure 4.10b) while the upper layer shows a homogeneous cloud layer (Figure 4.10a). Figure 4.11b also depicts this phenomenon. The satellite image confirms the existence of cloud streets (Figure 4.14).

To sum up case 2, it can be said that, similar to case 1, AROME-Arctic manages to predict the CAO correctly. Only over Isfjorden cloud formation is too weak, and the clouds disperse too soon before they get into the camera's field of view.

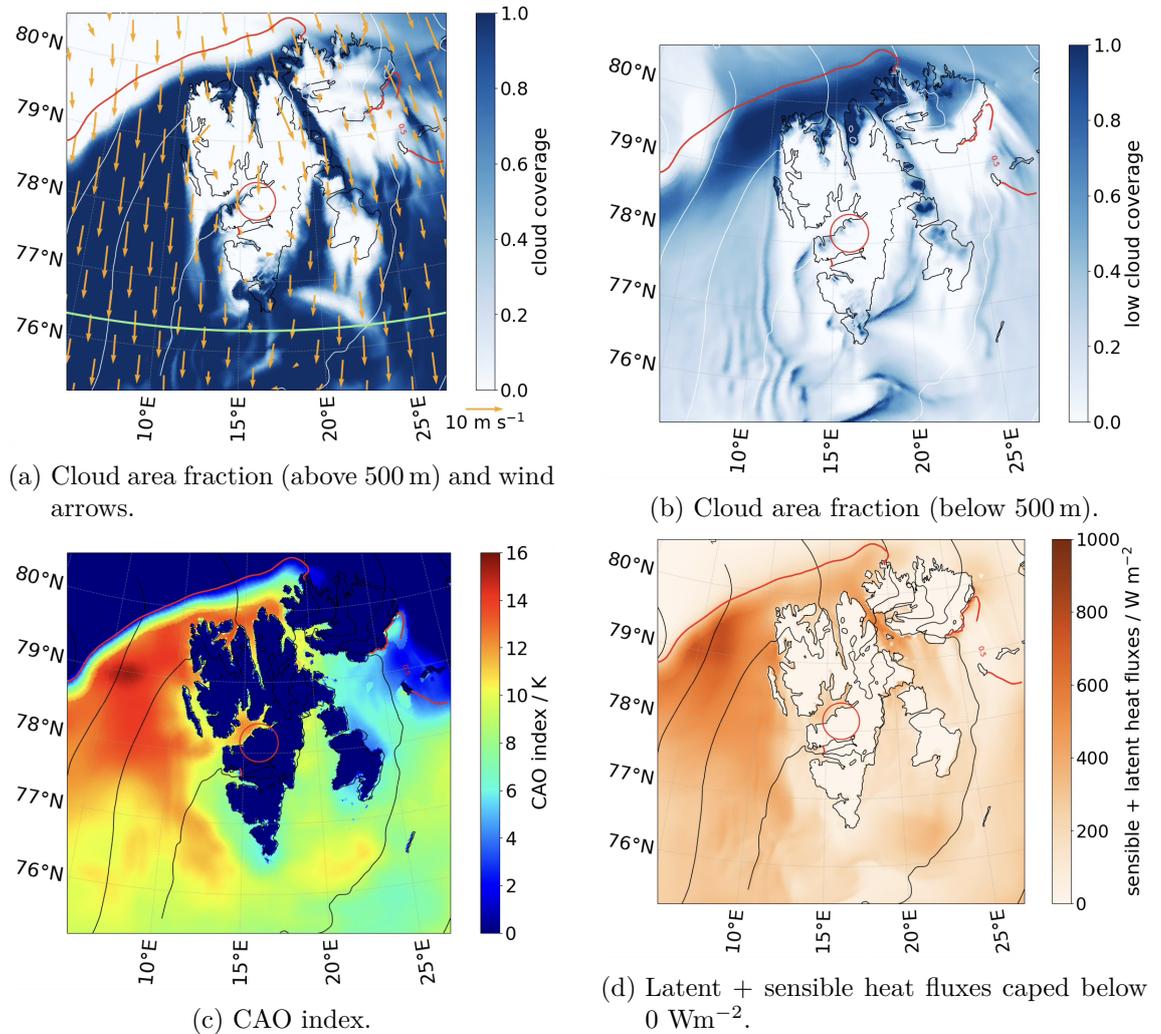


Figure 4.10.: Case 2: 2023-11-15 03:00 UTC. The red line indicates the area where the sea ice concentration exceeds 50% of the grid box, and the red circle indicates the camera's area of sight in AROME Arctic. The pressure lines at sea level are shown in white (upper panels) or in black (lower panels).

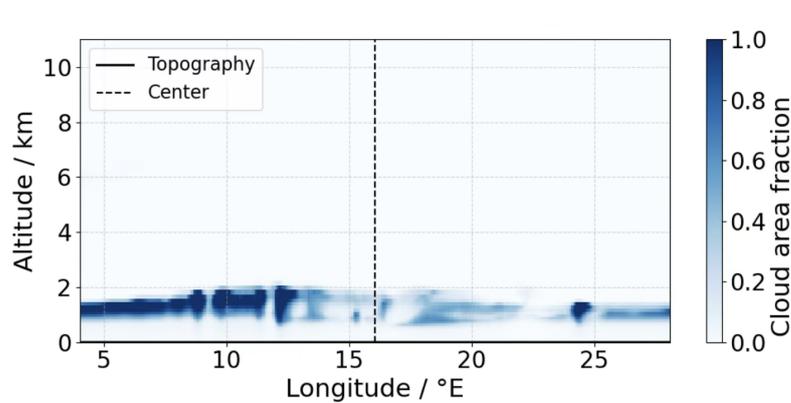


Figure 4.11.: Case 2: 2023-11-15 03:00 UTC West east cross section through the green line in Figure 4.10. The dotted vertical line shows the location of the camera, and the solid line at the bottom shows the topography.

4. AROME-Arctic Evaluation

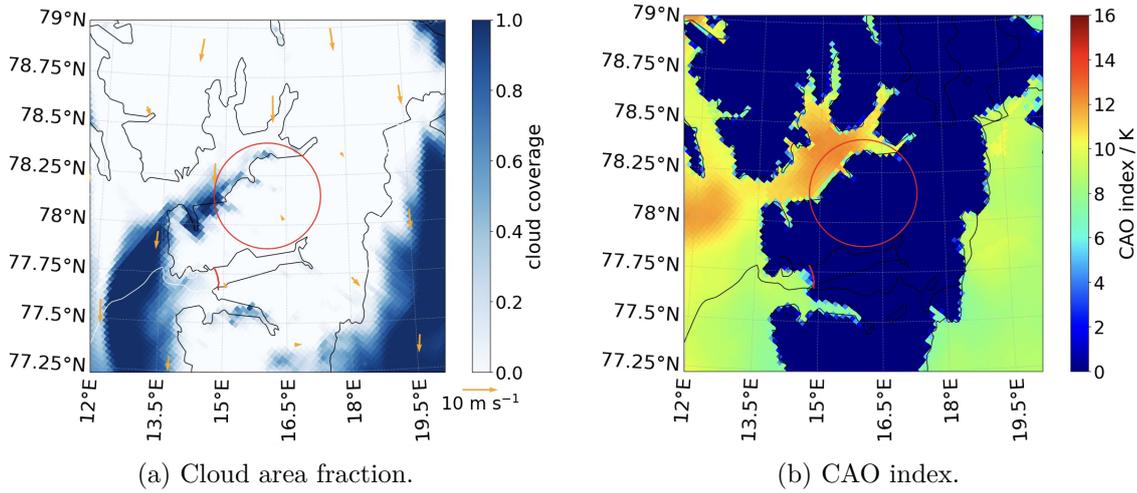


Figure 4.12.: Case 2 zoomed.



Figure 4.13.: Case 2: 2023-11-15 03:00 UTC. All-sky camera at KHO at 03:12 UTC showing overcast conditions.

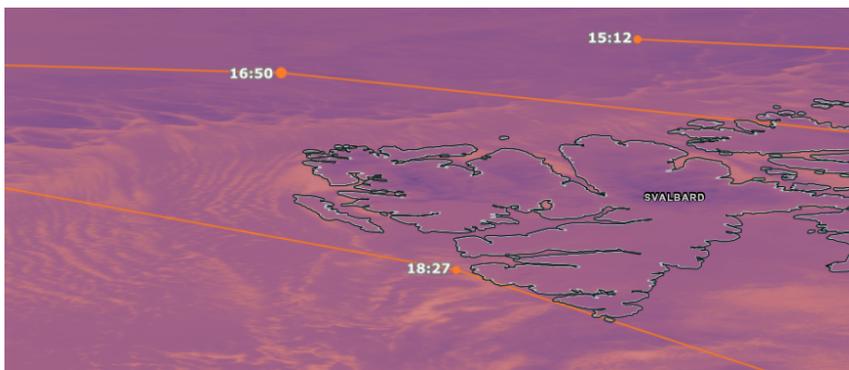


Figure 4.14.: Case 2: 2023-11-15 03:00 UTC. Brightness temperature (?) visualized by [NASA Earth Observing System Data and Information System \(EOSDIS\)](#) (2024) at 2023-11-14 16:50 and 18:27 UTC. The darker the color, the colder the brightness temperature. Purple indicates cold surfaces like clouds, sea ice, and snow.

4.2.2. Displacement Problem

The following case addresses a displacement problem. It discusses a larger synoptic regime with displaced pressure systems. AROME-Arctic correctly predicts the situation, but does not manage to get the spatial location right. Among the high offset cases, displacement

problem cases were found for approximately 20% of them.

Displacement Problem - wind divergence

Case 3 (Figure 4.15) on the 2021-12-01 18:00 UTC shows the displacement problem. A diverging wind field is located in the southern part of Svalbard (Figure 4.15a), which causes sinking motions and thus cloud dissipation. The diverging wind field itself is caused by synoptic conditions. A high-pressure system sits east of Svalbard, and a second, weaker, high-pressure system sits west of Svalbard, a bit more to the south than the first one. A low-pressure system is located north west of Svalbard (Figure 4.15b pressure lines). The two high-pressure systems form a saddle region in between them with high pressure but no strong gradients, which causes low and diverging wind speeds. The cloud-free regions in AROME-Arctic are located on the saddle south-west of Svalbard and below the high-pressure system east of Svalbard (Figure 4.15a). At the camera's location, AROME-Arctic sees overcast conditions, while the camera shows that the conditions are clear (Figure 4.16). Since the brightness temperature measured by a satellite is not very good at showing clouds over polar land masses during nighttime, the calculated cloud area fraction is chosen for this case study as an additional source for validation (Figure 4.17). Known weaknesses of remote sensing cloud detection should be kept in mind (Liu et al., 2004). The satellite sees the cloud-free region further north than AROME-Arctic and aligns with the all-sky camera (Figure 4.16).

It seems like AROME-Arctic predicts the synoptic situation and the interplay of the pressure systems fairly well and only misses the exact location. The origin of misplaced synoptic situations might come from the boundary conditions fed into the AROME-Arctic by the IFS (Müller et al., 2017). A possible solution could be ensemble model runs as they vary the initial conditions. This might lead to a different cloud formation behavior. Furthermore, the wind divergence zone might be slightly more north, which could cause the camera's location to be cloud-free.

To sum up case 3, it can be said that AROME-Arctic predicts the situation fairly well except for the wrong location of the pressure systems, which is likely caused by the IFS.

4. AROME-Arctic Evaluation

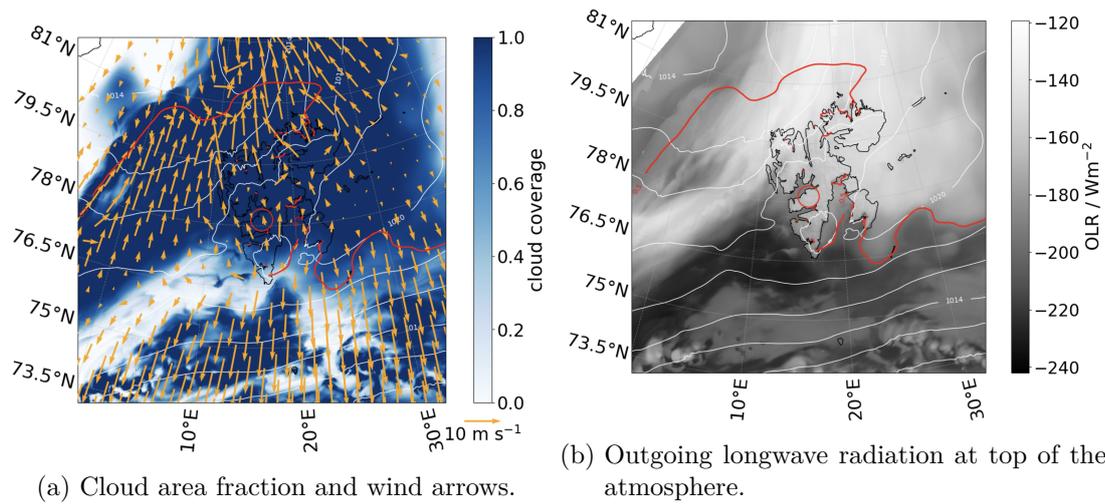


Figure 4.15.: Case 3: 2021-12-01 18:00 UTC. The red line indicates the area where the sea ice concentration exceeds 50% of the grid box, and the red circle indicates the camera's area of sight in AROME Arctic. The pressure lines at sea level are shown in white (upper panels) or in black (lower panels).



Figure 4.16.: Case 3: 2021-12-01 18:00 UTC. All-sky camera at KH0 at 18:01 UTC showing clear sky conditions. Stars and aurora are visible.

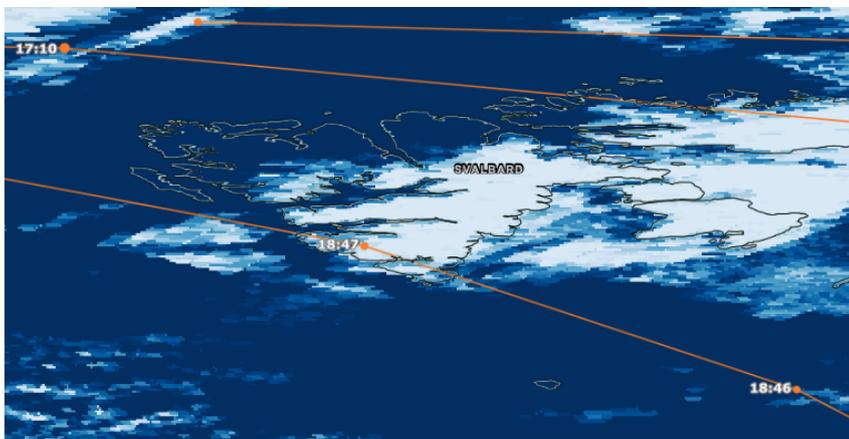


Figure 4.17.: Case 3: 2021-12-01 18:00 UTC. Cloud area fraction (?) visualized by [NASA Earth Observing System Data and Information System \(EOSDIS\) \(2024\)](#) at 2021-12-01 17:10 and 18:47 UTC. Blue colors indicate clouds.

5. Summary and Discussion

This chapter summarizes the main findings of this study and discusses its limitations and uncertainties.

To summarize Chapter 3, it can be said that among the three investigated approaches (physical, machine-learning, and combined), the combined approach provides the best results. However, there is still potential to improve the combined approach further. First, one could increase the number of manually labeled images to improve the training and validation dataset, both for the SVM and the 7-class ResNet. Most of the earlier studies utilize a larger training dataset than this study and achieved higher accuracies (Mommert, 2020; González-Fernández et al., 2024; Zhong et al., 2024). Due to a lack of time only the years 2019-2023 were manually labeled for this study. Doubling the dataset might already improve the performance of the combined approach.

A downside of machine learning compared to physical models is that their reasoning is often outside of our comprehension. They do not rely on explicit physical laws, which makes their internal decision processes difficult to interpret. A physical model can be improved by understanding the physics and adding them to the model, while a machine learning model is more like a black box that can only mimic the training dataset. In this particular case, the ResNet is only as good as I, the person labeling the training dataset. The ResNet learns to classify the images the way I classify the training dataset.

The labeling of the images introduces the first uncertainty to the final results. Cloud coverage in okta is not a measurement; it is a subjective value given by a meteorologist who eyeballs the cloud coverage of the whole visible sky. Another meteorologist might conclude on a different value. To assess my own subjectivity while labeling the images a 'labeling party' was conducted. Seven meteorology students from the University of Hamburg labeled the same subset of images. It appears that I have a small positive bias compared to the average value of each image labeled by the others (Figure A.2). Potential consequences for further studies might be that the labels could be double-checked by other meteorologists to reduce the subjectiveness of the datasets.

Furthermore, more time could be spent on tuning the SVM by Marocco (2025) to improve the performance of the prefixed model in the combined approach. The SVM and the ResNet could be tested with different or deeper architectures or even more loss functions and optimizers that have not been considered yet.

To sum up the evaluation of AROME-Arctic (Chapter 4), the used scores to evaluate AROME-Arctic statistically might be misleading. The case studies revealed that AROME-Arctic usually manages to predict the overall situation fairly well, but the clouds are spatially very variable, and for scores to be good, AROME-Arctic needs to capture the cloud at the right location and the right time.

5. Summary and Discussion

The statistical analysis shows that a cylinder together with the random cloud overlap schemes (Figure 4.3a) provides more realistic results compared to the pixel and the maximum and maximum-random cloud overlap schemes. This aligns with Hogan and Illingworth (2000) results that the vertical cloud overlap during winter in the UK is random. Furthermore, the cylinder is a more realistic than the pixel because it matches the cameras field of view better. The pixel in combination with the random cloud overlap scheme appears to perform well in Figure 4.4, however, this performance results from error compensation, as illustrated in Figure 4.3b. Error compensation means that in this case, if the true cloud coverage is clear or cloudy, the prediction is intermediate, while on the other hand, if the true cloud coverage is intermediate, the prediction is clear or cloudy. The consequences of error compensation are that it would have been fatal to only consider the marginal distribution (Figure 4.4) because the joint distribution (Figure 4.3a) might provide very important insights. The frame-like distribution of Figure 4.3a and 4.3b, which also aligns with the U-shaped nature of cloud coverage (Makhotina et al., 2021), is also found by Hogan et al. (2009), who compare the cloud fraction modeled by the Deutscher Wetterdienst (DWD) with radar and lidar-based cloud observations.

A lead time analysis shows that there is no dependency of cloud coverage forecast accuracy and lead times. The used skill score PSS shows that AROME-Arctic only has little skill in predicting cloud coverage. This is in contrast to the case studies in Chapter 4.2 that demonstrate an adequate representation of large, as well as smaller scale, synoptic features in AROME-Arctic. Thus, the PSS used in this study might not be the best choice to evaluate the cloud coverage forecast performance of an NWP model. Alternatively, multi-categorical skill scores like the Heidke Skill Score (HSS) or the Gerrity Skill Score (GSS) could be tested. However, as mentioned before, these skill scores have to be handled with caution. More detailed investigations like case studies potentially yield clearer results.

The case studies highlight the importance of considering details of specific situations so that one is not blindfolded by the falsify results of skill scores. The case studies focus on often-seen synoptic conditions with at least 5 oktas difference between the observed and the modeled cloud coverage. The boundary processes during CAOs (cases 1 and 2, Figure 4.6 and 4.10) were resolved adequately. Despite the huge prediction errors, only distinct and local processes (high clouds in case 1 and missing convection over Isfjorden in case 2) can be responsible for these offsets. This underlines the challenges in evaluating cloud coverage. The origin of the increased high cloud coverage (case 1) might originate from the IFS, which provides the boundary conditions for AROME-Arctic (Müller et al., 2017). The lack of cloud formation over Isfjorden (case 2) might be caused by the fact that Isfjorden is not well resolved in AROME-Arctic. Thus, the sea surface temperature in Isfjorden might be too cold, leading to too weak convection. Further investigations of the sea surface temperature in Isfjorden could be conducted based on a buoy system deployed in Isfjorden by UNIS or weather stations deployed on ships sailing in Isfjorden (Frank et al., 2023).

Another interesting aspect is the fact that both case 1 and case 2 show a very similar CAO in AROME-Arctic but case 1 shows clear sky conditions in the all-sky camera while case 2 shows overcast conditions. This might be explainable by local processes that are not fully resolved by AROME-Arctic. It might be that in reality case 1 has a lot less cloud

formation over Isfjorden which impacts the camera's location. Another possible reason for the different cloud coverages could be local wind channeling through surrounding valleys that either blow the clouds away from the camera (case 1) or towards the camera (case 2). It is difficult to isolate the exact reasons since these local processes are not resolved in AROME-Arctic. The 500 m resolution model AROME-Svalbard (Valkonen et al., 2020) might yield further insights as Mack et al. (2025) has shown that channeling through valleys is represented better in AROME-Svalbard than in AROME-Arctic.

Case 3 shows a displacement problem during which AROME-Arctic correctly predicts the synoptic situation, but is shifted in space. It is likely that the displacement happens due to the boundary conditions of the IFS. Ensemble runs could be a good way to approach this problem since small deviations in the initial conditions might lead to a different location of the wind divergence zone and thus of the cloud free area.

6. Conclusion and Outlook

This study introduced a methodology to gain cloud coverage estimates from all-sky images taken during the polar night. Furthermore, this cloud cover product was used to evaluate the NWP model AROME-Arctic.

The methodology is based on a machine learning approach which combines an SVM to classify the images into 0/8, 8/8, and intermediate cases, and a ResNet to classify the intermediate cases into 1-7/8 (Chapter 3). Compared to a testing dataset spanning over 3 years, an accuracy of 83% could be achieved.

The statistical evaluation of AROME-Arctic leads to an ill-conceived conclusion that it is not able to represent cloud coverage (Section 4.1). A central result of this study is that average scores need to be used with care. The case studies show that AROME-Arctic can predict to overall situation fairly well. Only a few processes in AROME-Arctic, of which not all are completely understood yet, can mess up its cloud coverage.

Cloud coverage, per se, is a lot more complicated to validate in an NWP model than, for example, temperature. The first reason is that cloud coverage in okta is subjective and not based on exact measurements. Another reason is the uncertainties with the cloud overlap schemes and the mapped spatial domain in the NWP model. Additional difficulties, especially for the location of KHO, are the complex terrain around it consisting of mountains, valleys, and a fjord close by. Local differences in e.g. channeling through valleys can lead to spatial displacement of cloud and thus huge misses between the all-sky camera and AROME-Arctic.

AROME-Arctic does an acceptable job, keeping in mind that the location of the camera is rather difficult to model due to the complex topography.

There is a lot of potential to continue this study. The transferability of the combined approach to extract cloud coverage from all-sky images can be explored with additional cameras at different locations (Figure 6.1). Not only can the extension to other camera locations be tested, but also the extension to daytime images, and thus all seasons can be explored. This is where the combined approach, which is mainly based on the 7-class ResNet, has a huge advantage compared to the physical approach. It is likely that the ResNet only needs a new training dataset to adapt to daytime images, whereas the physical approach simply does not work during daytime since it is, among other things, based on the moon's position and the number of visible stars. If the combined approach is reusable for other locations and daytime images, there is a huge potential in extending the cloud coverage validation of AROME-Arctic. How well does AROME-Arctic predict the cloud coverage at different locations? Are there daily or seasonal differences in cloud coverage accuracy in AROME-Arctic? Are high offsets usually related to CAOs? Is the initialization by the IFS better at lower latitudes?

6. Conclusion and Outlook

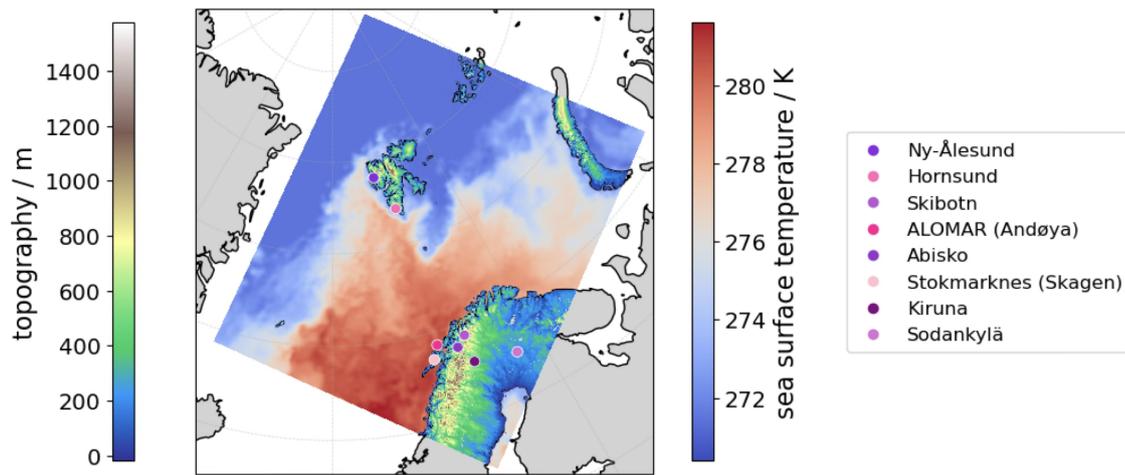


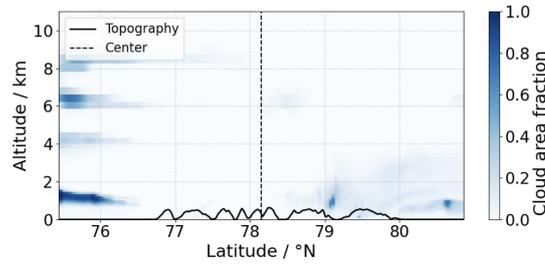
Figure 6.1.: Spatial domain of AROME-Arctic 2023-11-29 00:00 UTC with potential camera locations.

Even the current setup can be used to continue evaluating the cloud coverage prediction of AROME-Arctic, as the time series can easily be extended up to the current date.

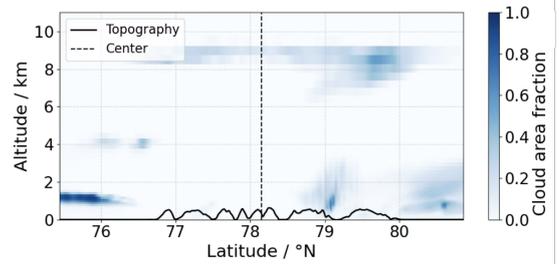
The findings about the cloud coverage prediction in AROME-Arctic can be used to further tune and improve the model. It is possible to isolate, and thus detect, the single processes that are responsible for the (non)existence of clouds. This will give insights into potential ways to improve AROME-Arctic.

While not the main purpose of this study, the weaknesses of remote sensing cloud coverage in high latitudes at night are shown. The cloud coverage retrieved from the all-sky camera can be further used to validate the cloud coverage detection of satellites.

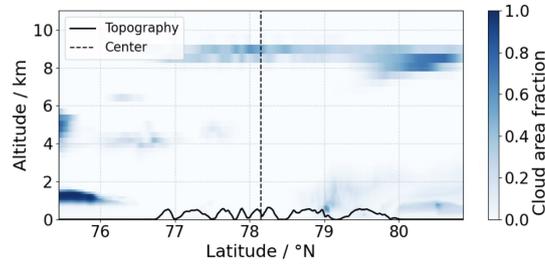
A. Additional Figures



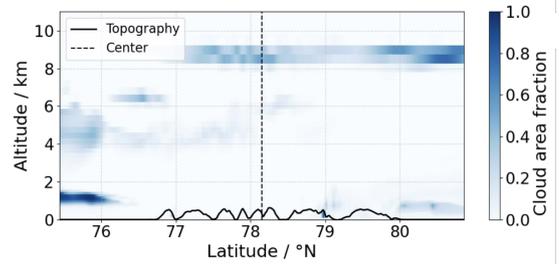
(a) 0 hour lead time -> 2022-12-25 09:00 UTC.



(b) 1 hour lead time -> 2022-12-25 10:00 UTC.



(c) 2 hours lead time -> 2022-12-25 11:00 UTC.



(d) 3 hour3 lead time -> 2022-12-25 12:00 UTC.

Figure A.1.: Case 1: North-south cross-section of AROME Arctic model run of 2022-12-25 09:00 UTC for different lead times.

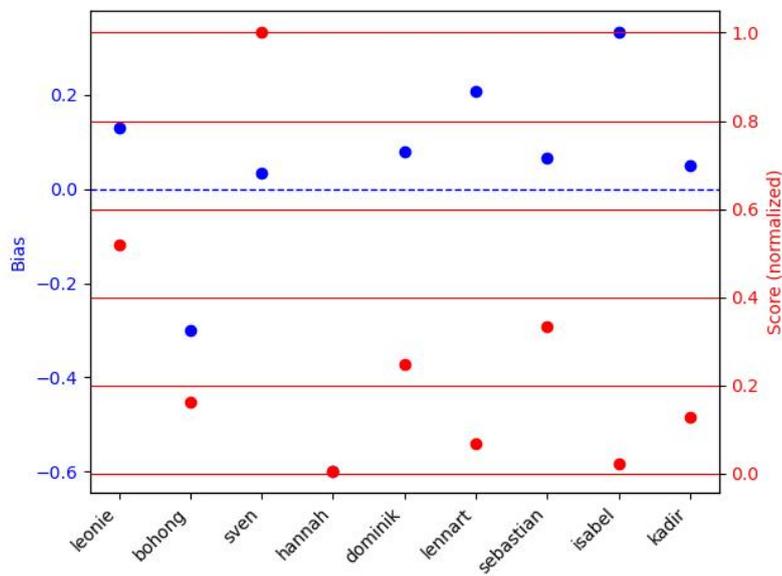


Figure A.2.: The results of the labeling party are calculated based on November 2022 as a reference month. The mean okta per image is used to calculate the bias (left y-axis and blue dots). The score is calculated by multiplying the number of labeled images by one over the bias (right y-axis and red dots). Comparing the bias for 'leonie' with the other biases shows the offset of my labels compared to meteorology students.

B. Additional Case

This case (Figure B.1) on the 2021-01-20 09:00 UTC shows topography blocking of clouds at the east coast of Svalbard. AROME-Arctic predicts almost clear sky conditions at the camera's location (Figure B.1a), while the camera itself shows fluctuating cloud coverage (Figure B.2).

Svalbard is impacted by a strong high-pressure system with its core north-east of Svalbard. The cloud area fraction in Figure B.1a shows an inhomogeneous pattern over Svalbard and a denser cell-like cloud cover over the sea, downstream of Svalbard. The wind field shows intense and diverging wind speeds away from Svalbard, and colder air over land flowing as northerlies and north-easterlies over the warmer sea. Once the air reaches the sea, convective cells begin to form, which is nicely visible in the outgoing longwave radiation (Figure B.1b).

At first, it was believed that AROME-Arctic already sees convection-like cloud features over land. However, an animation (Nellesen, 2026b) shows that these clouds are almost constant in space and time. The fact that Figure B.2 shows fluctuating cloud coverage at the camera's location might hint towards the cloud shadowing being too strong. Another potential error could be wind channeling through a nearby valley towards Isfjorden. Isfjorden might have cloud formation, which impacts the location of the camera.

Local weather situations like that are very difficult to forecast correctly. A potential solution could be increased resolution. AROME-Svalbard (Valkonen et al., 2020) the 500 m resolution model could be used to study if higher resolved topography might impact the cloud coverage distribution.

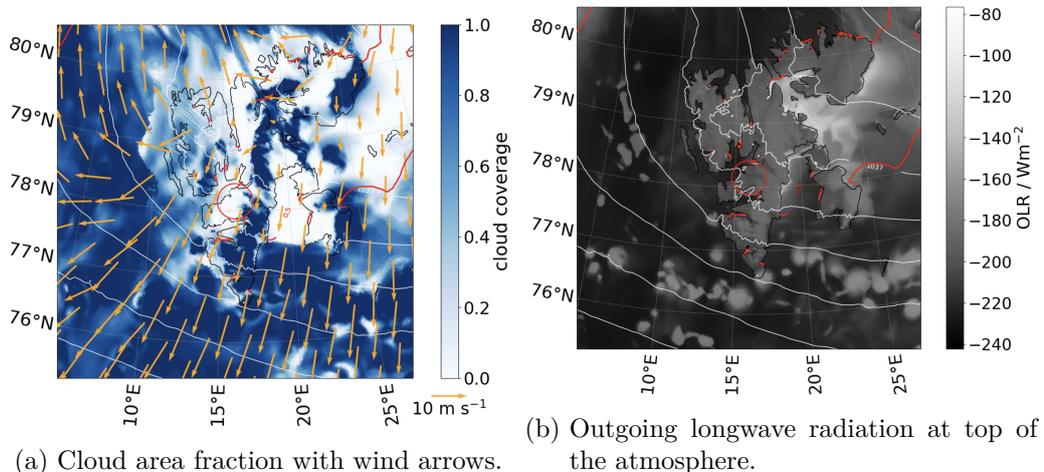
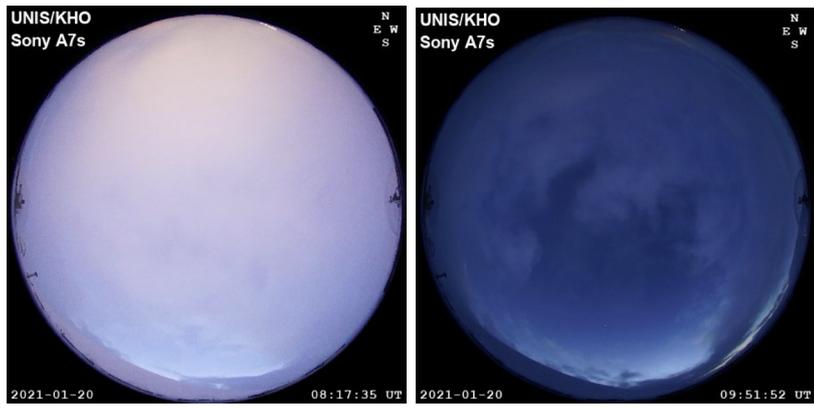


Figure B.1.: 2021-01-20 09:00 UTC: The red line indicates the area where the sea ice concentration exceeds 50% of the grid box, and the red circle indicates the camera's area of sight in AROME Arctic. The pressure lines at sea level are shown in white (upper panels) or in black (lower panels).

B. Additional Case



(a) 08:17 UTC.

(b) 09:51 UTC.

Figure B.2.: C2021-01-20 09:00 UTC: All-sky camera images at times before and after the case study. The images show fluctuation in cloud coverage.

Bibliography

- torch.optim.adam — pytorch documentation. <https://docs.pytorch.org/docs/stable/generated/torch.optim.Adam.html>, 2025a. Accessed: 2025-01-15.
- torch.optim.adamw — pytorch documentation. <https://docs.pytorch.org/docs/stable/generated/torch.optim.AdamW.html>, 2025b. Accessed: 2025-01-15.
- torch.nn.crossentropyloss — pytorch documentation. <https://docs.pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html>, 2025c. Accessed: 2025-01-15.
- C. Arbizu-Barrena, D. Pozo-Vázquez, J. A. Ruiz-Arias, and J. Tovar-Pescador. Macroscopic cloud properties in the wrf nwp model: An assessment using sky camera and ceilometer data. *Journal of Geophysical Research*, 120:10297–10312, 2015. ISSN 21562202. doi: 10.1002/2015JD023502.
- P. Bauer, A. Thorpe, and G. Brunet. The quiet revolution of numerical weather prediction, 9 2015. ISSN 14764687.
- P. Bechtold, J. W. M. Cuijpers, P. Mascart, and P. Trouilhet. Modeling of trade wind cumuli with a low-order turbulence model: Toward a unified description of cu and se clouds in meteorological models. *Journal of the Atmospheric Sciences*, 52(4):455–463, 1995. doi: 10.1175/1520-0469(1995)052<0455:MOTWCW>2.0.CO;2.
- Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 1994.
- L. Bengtsson, 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. Onville, L. Rontu, P. Samuelsson, D. S. Muñoz, A. Subias, S. Tijm, V. Toll, X. Yang, and M. Ødegaard Køltzow. The harmonie-arome model configuration in the aladin-hirlam nwp system. *Monthly Weather Review*, 145:1919–1935, 5 2017. ISSN 15200493. doi: 10.1175/MWR-D-16-0417.1.
- P. Bougeault. Cloud-ensemble relations based on the gamma probability distribution for the higher-order models of the planetary boundary layer. *Journal of the Atmospheric Sciences*, 39(12):2691–2700, 1982. doi: 10.1175/1520-0469(1982)039<2691:CERBOT>2.0.CO;2.
- S. Businger and R. J. Reed. Cyclogenesis in cold air masses. *Monthly Weather Review*, 117(6):133–156, 1989. doi: 10.1175/1520-0434(1989)004<0133:CICAM>2.0.CO;2.
- C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. doi: 10.1007/BF00994018.

- R. Crocker and M. Mittermaier. Exploratory use of a satellite cloud mask to verify nwp models. *Meteorological Applications*, 20:197–205, 2013. ISSN 14698080. doi: 10.1002/met.1384.
- Deutsches Klimarechenzentrum (DKRZ). Levante hpc system. <https://docs.dkrz.de/doc/levante/>, 2025. Accessed: 2026-01-15.
- D. Etling and R. A. Brown. Roll vortices in the planetary boundary layer: A review. *Boundary-Layer Meteorology*, 65:215–248, 1993. doi: 10.1007/BF00705527. URL <https://doi.org/10.1007/BF00705527>.
- Y. Fouquart and B. Bonnel. Computations of solar heating of the earth’s atmosphere: A new parameterization. *Beiträge zur Physik der Atmosphäre*, 53:35–62, 1980.
- L. Frank, M. O. Jonassen, T. Remes, F. R. Schalamon, and A. Stenlund. Iwin: the isfjorden weather information network. *Earth System Science Data*, 15:4219–4234, 9 2023. ISSN 18663516. doi: 10.5194/essd-15-4219-2023.
- E. Gleeson, E. Kurzeneva, W. de Rooy, L. Rontu, D. M. Pérez, C. Clancy, K.-I. Ivarsson, B. J. Engdahl, S. Tijm, K. P. Nielsen, M. Shapkalijevski, P. Maalampi, P. Ukkonen, Y. Batrak, M. Kähnert, T. Kettler, S. M. E. van den Brekel, M. R. Adriaens, N. Theeuwes, B. Pálmason, T. Rieutord, J. Fannon, E. Whelan, S. Viana, M. Homleid, G. Bessardon, J. Onvlee, P. Samuelsson, D. Santos-Muñoz, O. N. Vignes, and R. Stappers. The cycle 46 configuration of the harmonie-arome forecast model. *Meteorology*, 3: 354–390, 11 2024. doi: 10.3390/meteorology3040018.
- D. González-Fernández, R. Román, J. C. Antuña-Sánchez, V. E. Cachorro, G. Copes, S. Herrero-Anta, C. H. del Barrio, África Barreto, R. González, R. Ramos, P. Martín, D. Mateos, C. Toledano, A. Calle, and Ángel de Frutos. A neural network to retrieve cloud cover from all-sky cameras: A case of study over antarctica. *Quarterly Journal of the Royal Meteorological Society*, 10 2024. ISSN 1477870X. doi: 10.1002/qj.4834.
- R. L. Grossman and A. K. Betts. Air–sea interaction during an extreme cold air outbreak from the eastern coast of the united states. *Monthly Weather Review*, 118(2):324–342, 1990. doi: 10.1175/1520-0493(1990)118<0324:AIDAEC>2.0.CO;2.
- J. Hartmann, C. Kottmeier, and S. Raasch. Roll vortices and boundary-layer development during a cold air outbreak. *Boundary-Layer Meteorology*, 84:45–65, 1997. doi: 10.1023/A:1000392931768. URL <https://doi.org/10.1023/A:1000392931768>.
- K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. 12 2015. URL <http://arxiv.org/abs/1512.03385>.
- R. J. Hogan and A. J. Illingworth. Deriving cloud overlap statistics from radar. *Quarterly Journal of the Royal Meteorological Society*, 126(569):2903–2909, 2000.
- R. J. Hogan, C. Jakob, and A. J. Illingworth. Comparison of ecmwf winter-season cloud fraction with radar-derived values. Technical report, 2001.

- R. J. Hogan, E. J. O'Connor, and A. J. Illingworth. Verification of cloud-fraction forecasts. *Quarterly Journal of the Royal Meteorological Society*, 135:1494–1511, 7 2009. ISSN 00359009. doi: 10.1002/qj.481.
- A. J. Illingworth, R. J. Hogan, E. J. O'Connor, D. Bouniol, M. E. Brooks, J. Delanoë, D. P. Donovan, J. D. Eastment, N. Gaussiat, J. W. Goddard, M. Haeffelin, H. K. Baltinik, O. A. Krasnov, J. Pelon, J. M. Piriou, A. Protat, H. W. Russchenberg, A. Seifert, A. M. Tompkins, G. J. van Zadelhoff, F. Vinit, U. Willen, D. R. Wilson, and C. L. Wrench. Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bulletin of the American Meteorological Society*, 88: 883–898, 6 2007. ISSN 00030007. doi: 10.1175/BAMS-88-6-883.
- IPCC. *Climate Change 2021: The Physical Science Basis*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.
- KHO. Aurora and kjell henriksen observatory. <https://aurora.unis.no>, 2026. Accessed: 2026-01-22.
- B. Y. Kim, J. W. Cha, and K. H. Chang. Twenty-four-hour cloud cover calculation using a ground-based imager with machine learning. *Atmospheric Measurement Techniques*, 14:6695–6710, 10 2021. ISSN 18678548. doi: 10.5194/amt-14-6695-2021.
- M. Kähnert, H. Sodemann, W. C. D. Rooy, and T. M. Valkonen. On the utility of individual tendency output: Revealing interactions between parameterized processes during a marine cold air outbreak. 2021. doi: 10.1175/WAF-D-21. URL <https://doi.org/10.1175/WAF-D-21->.
- P. Le Moigne. Surfex scientific documentation. Cnrm technical report, CNRM / Météo-France, 2009.
- Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.
- C. M. Lee, M. Degrandpre, J. Guthrie, V. Hill, R. Kwok, J. Morison, C. J. Cox, H. Singh, T. P. Stanton, and J. Wilkinson. Emerging technologies and approaches for in situ, autonomous observing in the arctic. *SPECIAL ISSUE ON THE NEW ARCTIC OCEAN*, 35:210–221, 2022. doi: 10.2307/27182719.
- G. Lenderink and A. A. M. Holtslag. An updated length-scale formulation for turbulent mixing in clear and cloudy boundary layers. *Quarterly Journal of the Royal Meteorological Society*, 130(604):3405–3427, 2004. doi: 10.1256/qj.03.117.
- T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-October, pages 2999–3007. Institute of Electrical and Electronics Engineers Inc., 12 2017. ISBN 9781538610329. doi: 10.1109/ICCV.2017.324.

- Y. Liu, J. R. Key, R. A. Frey, S. A. Ackerman, and W. P. Menzel. Nighttime polar cloud detection with modis. *Remote Sensing of Environment*, 92:181–194, 8 2004. ISSN 00344257. doi: 10.1016/j.rse.2004.06.004.
- I. Loshchilov and F. Hutter. Decoupled weight decay regularization. 1 2019. URL <http://arxiv.org/abs/1711.05101>.
- L. Mack, M. Kähnert, Q. Rauschenbach, L. Frank, F. H. Hasenburg, J. M. Huss, M. O. Jonassen, M. Malpas, Y. Batrak, T. Remes, N. Pirk, and C. K. Thomas. Stable boundary layers in an arctic fjord-valley system: Evaluation of temperature profiles observed from fiber-optic distributed sensing and comparison to numerical weather prediction systems at different resolutions. *Journal of Geophysical Research: Atmospheres*, 130, 1 2025. ISSN 21698996. doi: 10.1029/2024JD042825.
- I. A. Makhotina, D. G. Chechin, and A. P. Makshtas. Cloud radiative forcing over sea ice in the arctic during the polar night according to north pole-37, -39, and -40 drifting stations. *Izvestiya - Atmospheric and Ocean Physics*, 57:451–460, 9 2021. ISSN 00014338. doi: 10.1134/S0001433821050091.
- A. Marocco. KHO-AutomaticCloudinessEstimation: An algorithm to estimate cloud cover from rgb nighttime images. <https://github.com/alessandromarocco/KHO-AutomaticCloudinessEstimation>, 2025. GitHub repository, accessed 2025-02-26.
- E. J. Mlawer, S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough. Radiative transfer for inhomogeneous atmospheres: Rrtm, a validated correlated-k model for the longwave. *Journal of Geophysical Research Atmospheres*, 102:16663–16682, 7 1997. ISSN 01480227. doi: 10.1029/97jd00237.
- M. Mommert. Cloud identification from all-sky camera data with machine learning. *The Astronomical Journal*, 159:178, 4 2020. ISSN 0004-6256. doi: 10.3847/1538-3881/ab744f.
- M. Moser, C. Voigt, T. Jurkat-Witschas, V. Hahn, G. Mioche, O. Jourdan, R. Dupuy, C. Gourbeyre, A. Schwarzenboeck, J. Lucke, Y. Boose, M. Mech, S. Borrmann, A. Ehrlich, A. Herber, C. Lüpkes, and M. Wendisch. Microphysical and thermodynamic phase analyses of arctic low-level clouds measured above the sea ice and the open ocean in spring and summer. *Atmospheric Chemistry and Physics*, 23:7257–7280, 7 2023. ISSN 16807324. doi: 10.5194/acp-23-7257-2023.
- M. Müller, Y. Batrak, J. Kristiansen, M. A. Kølitzow, G. Noer, and A. Korosov. Characteristics of a convective-scale weather forecasting system for the european arctic. *Monthly Weather Review*, 145:4771–4787, 12 2017. ISSN 15200493. doi: 10.1175/MWR-D-17-0194.1.
- NASA Earth Observing System Data and Information System (EOSDIS). Nasa worldview. <https://worldview.earthdata.nasa.gov/>, 2024. Accessed: 2025-12-13.
- R. A. Neggers, M. Köhler, and A. C. Beljaars. A dual mass flux framework for boundary layer convection. part i: Transport. *Journal of the Atmospheric Sciences*, 66:1465–1487, 2009. ISSN 00224928. doi: 10.1175/2008JAS2635.1.

- L. Nellesen. Case 1 north south cross section, 2026a. URL <https://doi.org/10.5281/zenodo.18430285>.
- L. Nellesen. Cloud coverage of appendix case, 2026b. URL <https://doi.org/10.5281/zenodo.18430860>.
- Norwegian Meteorological Institute. Arome-arctic archive (thredds dataset). <https://thredds.met.no/thredds/catalog/aromearcticarchive/catalog.html>, 2024. Accessed: 2025-01-13.
- Norwegian Meteorological Institute. The arome-arctic weather model, 2025. URL <https://www.met.no/en/projects/The-weather-model-AROME-Arctic>. Accessed: 2026-01-08.
- L. Papritz and S. Pfahl. Importance of latent heating in mesocyclones for the decay of cold air outbreaks: A numerical process study from the pacific sector of the southern ocean. *Monthly Weather Review*, 144:315–336, 2016. ISSN 15200493. doi: 10.1175/MWR-D-15-0268.1.
- L. Papritz and H. Sodemann. Characterizing the local and intense water cycle during a cold air outbreak in the nordic seas. 2018. doi: 10.1175/MWR-D-18.
- J.-P. Pinty and P. Jabouille. A mixed-phase cloud parameterization for use in mesoscale non-hydrostatic model: simulations of a squall line and of orographic precipitations. In *Conf. on cloud physics*, pages 217–220. Amer. Meteor. Soc. Everett, WA, 1998.
- W. C. D. Rooy, P. Siebesma, P. Baas, G. Lenderink, S. R. D. Roode, H. D. Vries, E. V. Meijgaard, J. F. Meirink, S. Tijm, and B. V. Veen. Model development in practice: A comprehensive update to the boundary layer schemes in harmonie-arome cycle 40. *Geoscientific Model Development*, 15:1513–1543, 2 2022. ISSN 19919603. doi: 10.5194/gmd-15-1513-2022.
- E. M. Samuelson and R. G. Graversen. Weather situation during observed ship-icing events off the coast of northern norway and the svalbard archipelago. *Weather and Climate Extremes*, 24, 6 2019. ISSN 22120947. doi: 10.1016/j.wace.2019.100200.
- Y. Seity, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson. The arome-france convective-scale operational model. *Monthly Weather Review*, 139:976–991, 3 2011. ISSN 00270644. doi: 10.1175/2010MWR3425.1.
- A. P. Siebesma, P. M. Soares, and J. Teixeira. A combined eddy-diffusivity mass-flux approach for the convective boundary layer. *Journal of the Atmospheric Sciences*, 64: 1230–1248, 4 2007. ISSN 00224928. doi: 10.1175/JAS3888.1.
- P. M. M. Soares, P. M. A. Miranda, A. P. Siebesma, and J. Teixeira. An eddy-diffusivity/mass-flux parametrization for dry and shallow cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, 130(604):3365–3383, 2004. doi: 10.1256/qj.03.223.
- Sony A7s. Sony a7s all-sky color camera. <https://aurora.unis.no/data/SonyA7s.html>, 2026. Accessed: 2026-01-06.

- H. Sverdrup. The norwegian north polar expedition with the maud. *Vol. II, Meteorology, Geophysical Institute Bergen*, 331, 1933.
- M. Tjernström, C. Leck, P. O. G. Persson, M. L. Jensen, S. P. Oncley, and A. Targino. The summertime arctic atmosphere: Meteorological measurements during the arctic ocean experiment 2001. *Bulletin of the American Meteorological Society*, 85(9):1305–1322, 2004. doi: 10.1175/BAMS-85-9-1305.
- M. Tjernström, C. E. Birch, I. M. Brooks, M. D. Shupe, P. O. Persson, J. Sedlar, T. Mauritsen, C. Leck, J. Paatero, M. Szczodrak, and C. R. Wheeler. Meteorological conditions in the central arctic summer during the arctic summer cloud ocean study (ascos). *Atmospheric Chemistry and Physics*, 12:6863–6889, 2012. ISSN 16807316. doi: 10.5194/acp-12-6863-2012.
- T. Valkonen, P. Stoll, Y. Batrak, M. Køltzow, T. M. Schneider, E. E. Stigter, O. B. Aashamar, E. Støylen, and M. O. Jonassen. Evaluation of a sub-kilometre nwp system in an arctic fjord-valley system in winter. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 72:1–21, 1 2020. ISSN 16000870. doi: 10.1080/16000870.2020.1838181.
- J. Zhang, P. Liu, F. Zhang, and Q. Song. Cloudnet: Ground-based cloud classification with deep convolutional neural network. *Geophysical Research Letters*, 45:8665–8672, 8 2018. ISSN 19448007. doi: 10.1029/2018GL077787.
- X. Zhong, F. Du, Y. Hu, X. Hou, Z. Zhu, X. Zheng, K. Huang, Z. Ren, and Y. Hou. Automatic classification of all-sky nighttime cloud images based on machine learning. *Electronics (Switzerland)*, 13, 4 2024. ISSN 20799292. doi: 10.3390/electronics13081503.

Acknowledgment

I want to thank my primary supervisor, Prof. Dr. Stefan Bühler, for the opportunity to write my thesis. I would also like to thank my second supervisor, Dr. Marvin Kähnert, for his relentless support. Despite him being in Oslo, the communication with him was outstanding. Emails were usually answered within less than 15 minutes (sometimes only 15 seconds), which made working with him very efficient and enjoyable.

Furthermore, I would like to thank Prof. Noora Partamies. She is the reason this project existed in the first place, as she provided the images from the all-sky camera.

Not to mention my boyfriend, Jon, for proofreading and emotional support. Special thanks to my fellow students and friends who made studying a lot more enjoyable. Thanks for the last five and a half years!

Lastly, I would like to thank my parents, Bernd and Birgit Nellesen, for supporting me financially and mentally during the last years. Without your support, this would not have been possible!

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit im Studiengang Atmospheric Science selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe. Sofern im Zuge der Erstellung der vorliegenden Abschlussarbeit auf generativer künstlicher Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten Wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Hamburg, den _____ Unterschrift: _____