Shortening of the Arctic cold air outbreak season detected by a phenomenological machine learning approach

Filip Severin von der Lippe[1], Tim Carlsen[1], Trude Storelvmo[1], and Robert Oscar David[1]

[1]University of Oslo

Correspondence: Filip Severin von der Lippe (fslippe@uio.no)

Abstract. Marine cold air outbreaks (CAOs) frequently occur in the Arctic when cold air moves over the relatively warm ocean, resulting in large turbulent fluxes, instability and cloud formation. Given the high frequency of CAOs during the Arctic winter, the associated clouds have a large impact on the region's radiative balance. Due to Arctic warming, the prevalence of CAOs and their clouds may change, impacting the Arctic radiative balance and potentially amplifying or mitigating local and global warming.

To better understand how CAO clouds respond to Arctic warming, this study has developed a phenomenological CAO cloud classification tool that utilizes machine learning methods to identify closed and open cell clouds in CAOs from MODIS satellite imagery. This new approach achieves better performance in identifying CAO clouds compared to the marine cold air outbreak index calculated using MERRA-2 reanalysis, with accuracies of 85.4 % and 78.0 %, respectively. The new approach has revealed frequent CAO cloud formation in regions of high sea surface temperatures, with occurrence maxima along the Norwegian coast and the Northern Atlantic region south of Iceland. Furthermore, the approach reveals trends in CAO cloud cover that suggest a shortening of the CAO season, characterized by an approximate 10 %, increase in cloud coverage during winter and a nearly 20 % decrease during the shoulder months over the past 25 years. These trends suggest a positive radiative feedback during winter in response to climate change, underscoring the importance of further investigating these clouds to understand the trajectory of future Arctic climate.

1 Introduction

Clouds in polar regions are often associated with marine cold air outbreaks (CAOs). These clouds form in the marine boundary layer (MBL) when cold and dry air from snow- and ice-covered regions moves over the relatively warm ocean. This produces a turbulent environment where large latent and sensible heat fluxes lead to the formation of clouds. Near the sea ice edge, these clouds form long cloud streets of densely packed closed cell stratocumulus, which transition into open cell broken cumulus when they traverse the open ocean (Brümmer, 1999; Geerts et al., 2022). These CAO clouds have a large impact on the surface radiative energy balance through their extensive coverage (Fletcher et al., 2016a). This is observed through their high albedo compared to the underlying dark ocean, reflecting incoming solar radiation back to space (cooling effect), and absorption and re-emission of outgoing terrestrial radiation (warming effect).

As the Arctic has experienced significant warming in recent decades (Serreze and Barry, 2011), it is crucial to study how CAOs may be affected. This warming is especially pronounced during the winter months where CAOs are most prevalent (Fletcher et al., 2016a; Dahlke et al., 2022), highlighting the potential of large impacts on CAOs in response to warming. Specifically, the strength of CAOs is projected to change (Landgren et al., 2019), which could affect cloud properties (Murray-Watson et al., 2023) and influence future warming. Furthermore, as the Arctic experiences polar night and day with highly seasonal variations in solar radiation, changes in CAO seasonality due to season-dependent warming may further impact the Arctic's radiative balance. This could potentially amplify or mitigate the significant warming observed in the region, underlining the importance of studying climatological shifts in the seasonality of CAOs.

To better understand the development of CAO clouds, they have been extensively studied through modeling, in situ observations and satellite studies (e.g., Hartmann et al., 1997; Abel et al., 2017; Geerts et al., 2022; Wu and Ovchinnikov, 2022). Especially, the transition from dense closed-cell stratocumulus to open-cell broken cumulus has been investigated (e.g., Abel et al., 2017; Yamaguchi et al., 2017; Tornow et al., 2021). As the denser closed cells have a higher albedo than the open cells (McCoy et al., 2017), the processes influencing the break-up to open cells become of high importance when studying the radiative impact of CAO clouds.

Several studies have suggested the onset of precipitation as a main driver for cloud break-up (e.g., Abel et al., 2017; Yamaguchi et al., 2017; Tornow et al., 2021). This precipitation, combined with increased winds, may further aid break-up into open cells by increasing the MBL moisture (Eastman et al., 2022), and favor the formation of cumuliform clouds (Stevens et al., 1998). Additionally, precipitation and evaporative cooling of the lower MBL (Abel et al., 2017) may lead to a decoupling of the stratocumulus cloud layer from the moisture-supplying surface (Bretherton and Wyant, 1997), contributing to the break-up. Furthermore, other research suggest that changes in MBL stability (McCoy et al., 2017), such as from increasing sea surface temperatures (SSTs), also play a role in this process.

While Abel et al. (2017) focused on a single CAO, introducing uncertainties regarding its universality, others have analyzed multiple CAOs by utilizing the marine cold air outbreak index M (Kolstad and Bracegirdle, 2008). This index measures the instability of the MBL and is calculated as the difference between the surface potential skin temperature and the potential temperature of a chosen pressure level, typically $850\,\text{hPa}$ (i.e. Papritz and Spengler, 2017). By utilizing reanalysis products such as Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2, Gelaro et al., 2017), which provide global climate and weather data of the past at up to hourly resolution, earlier studies have defined a CAO as a model grid point where the index M is positive, indicating instability and the possibility for clouds (e.g. Fletcher et al., 2016a; Murray-Watson and Gryspeerdt, 2024). This method provides an easy way to find the location of CAOs to use for further analysis, such as investigating cloud break-up. Despite the ease of use, reanalysis data introduces model biases especially in remote regions such as the Arctic with limited available observational data. In addition, uncertainties arise from the fact that a positive index does not necessarily result in the existence of a CAO cloud. While this can be addressed by requiring higher M indices (i.e. Murray-Watson et al., 2023), this may introduce biases from omitting clouds associated with weaker instabilities, skewing the analysis towards stronger CAO events. Furthermore, since the M index has been shown to decrease downwind (Murray-Watson et al., 2023), requiring higher M indices may result in the omission of clouds as they are advected. Consequently, as there is

no universally accepted M index for defining CAO clouds, this motivates the introduction of a phenomenological approach for defining CAOs that is based on the existence of clouds, and that is free of the biases introduced by modeling and reanalysis.

Clouds within CAOs typically cover large areas and are easily distinguishable from other clouds due to their cellular structure. This makes them easy to spot from satellite images (e.g. Fig. 1b), which in turn can be used to better understand their coverage and radiative impact. For the Arctic, this requires a polar-orbiting satellite such as Terra which since 24 February 2000 has provided multiple products of the surface, atmosphere and clouds through the onboard Moderate Resolution Imaging Spectroradiometer (MODIS) instrument. This instrument provides near-daily surface, sea ice, ocean and atmosphere data products of the entire Earth (King et al., 1992), and has been extensively used through its measurements of 36 radiance bands in the solar and thermal infrared spectral range. The 36 MODIS radiance bands are calibrated and provided as 5 minute swaths with a wide viewing angle of $55°$, giving a total coverage of 2330x2030 km. This extensive coverage makes MODIS an optimal instrument for classification of CAO clouds.

Utilizing MODIS products, a human can hand-label CAO clouds of interest, rather than using meteorological parameters to predict the possibility of a cloud. Such a labeling approach would be time-consuming and could introduce significant subjective bias from the labeler, as demonstrated in other cloud classification tasks (Stevens et al., 2020). To automate the cloud classification process, machine learning methods may be utilized. Wood and Hartmann (2006) introduced a supervised neural network (NN) utilizing the MODIS liquid water path (LWP) retrievals to classify closed and open cells in the subtropics. Despite its application in the study of CAOs (McCoy et al., 2017), this NN requires MODIS daytime retrievals, leading to a lack of data during the polar night. As Arctic CAOs are most frequent during the dark winter season from late autumn to early spring (Fletcher et al., 2016b), this model remains inapplicable for Arctic CAO studies.

Utilizing daytime MODIS calibrated radiances, Kurihana et al. (2022) developed an unsupervised machine learning approach comprising an autoencoder (Hinton and Zemel, 1993) and hierarchical agglomerative clustering for cloud classification. Although this study relied on certain radiance bands that are only available during daytime, the method can be adapted to eliminate this dependency. By modifying the approach of Kurihana et al. (2022), it becomes possible to phenomenologically classify wintertime CAO clouds using nighttime available bands in the thermal infrared. Specifically, band 31 (10.780 - 11.280 μm) in the thermal infrared can be selected as it is widely used for MODIS cloud classification algorithms given its sensitivity to clouds (Frey et al., 2008). As this approach is unsupervised, it also reduces the requirement of human labeling for training, reducing subjectivity and possible restrictions from the training dataset in producing accurate classifications. By developing a similar tool to Kurihana et al. (2022), this study will not only help in understanding drivers of cloud break-up within CAOs, but also provide an accurate database of CAO clouds for other studies, such as investigating past changes in CAO cloud cover to provide future projections of the radiative impact of CAOs.

This study aims to introduce a new phenomenological CAO cloud classification tool called CAOnet, utilizing a NN and 25 years of MODIS satellite imagery. CAOnet will provide a valuable CAO climatology database for future studies of CAO clouds, mitigating uncertainties introduced by traditional reanalysis methods. Along with a M index optimized for cloud detection, CAOnet will be employed to explore the potential radiative impact of changes in CAO cloud cover in response to projected climate change.

2 Methods

2.1 Data and model development

2.1.1 MODIS data preparation

Swaths of MODIS calibrated radiances (band 31, 10.780 - 11.280 μm) in the thermal infrared from the satellite Terra were utilized to provide a database for cloud classification during an extended winter period from September to May.

Each MODIS swath has a resolution of 1354x2030 pixels, that translates to 1 km resolution at nadir, decreasing to 4.8 km at the scan extremes. To keep the resolution as close to 1 km as possible for all pixels, the swath width was decreased from 1354 to 1024 pixels. This resulted in a resolution of approximately 2.05 km at the scan extremes, balancing uniform pixel resolution while keeping daily swath coverage in the Arctic north of 55 °N. Although additional MODIS swaths from the satellite Aqua could have been utilized to further enhance pixel uniformity, this was avoided to minimize storage and computational requirements. Finally, up to four temporal subsequent swaths were combined to extend the swath size from five to 20 minutes, resulting in a resolution of up to 1024x8120 pixels.

In order to make classifications of smaller regions containing closed and open cells, the combined satellite swaths were split into smaller image patches of 128x128 pixels each. Such a patch was large enough to cover multiple closed or open cloud cells, which made it possible for a human and the classification model to distinguish cellular structures from other cloud fields. By utilizing multiple swaths acquired between 1 March 2000 and February 28th 2025 split into patches, a database for a classification model was created.

2.1.2 Developing a phenomenological classification model

To classify the satellite images and find CAO clouds, an unsupervised machine learning approach based on the work of Kurihana et al. (2022) was developed. While Kurihana et al. (2022) used hierarchical agglomerative clustering, this study utilizes K-means clustering for faster computation and the ability to train a model for predicting unseen data. K-means is an unsupervised machine learning algorithm that classifies data into a user-defined number of clusters. It operates by optimizing cluster centroids during training, and assigning a given input to the cluster with the nearest mean. In terms of the satellite image patches, this mean equals the mean pixel value of each patch. For the case of looking at clouds, the mean pixel value may not efficiently describe cloud structure or type, giving meaningless classifications. By giving K-means more input than just the original image, it may create more meaningful clusters, especially if the input describes the features that are most important for the specific image. Such a feature could describe image contrast, cloud cell size, or brightness, as well as features that are meaningless to humans but still helps informing correct classification.

Similar to Kurihana et al. (2022), an autoencoder was used to extract dimensionally reduced information incorporating the most important information of the input patches. This is a convolutional neural network that comprises two main components (see Fig. 1a). The first component is the encoder, which performs compression on a given input image, producing a compressed feature representation. This is fed to the decoder, whose goal is to decompress the encoded features into an output that resembles

the original input image. During training, these two components work together to most accurately reproduce the original input image. The encoder achieves this through saving the most significant information in the compressed feature representation, helping the decoder to produce an accurate image reconstruction. The compressed feature representation produced by the encoder was employed in further classification tasks, aiding K-means in making meaningful classifications. Thus, the encoder and K-means clustering comprised the two main components of the classification model called CAOnet, as visualized in Fig. 1b.

A simplified structure of the autoencoder used in the subsequent analysis is shown in Fig. 1a. The encoder performs four dimensionality reductions, with each reduction halving the width and height of the input to that layer. Simultaneously, the number of features are increased from the single-feature band 31 input patch, to 16, 32, 64 and finally 128 features. While some of the final 128 features could be understandable to a human, others are likely to be meaningless. Nevertheless, this number was found to aid K-means in creating the most meaningful clusters. For each dimensionality reduction, the input passes through residual blocks, which have been shown to mitigate performance degradation in deep neural networks (He et al., 2015). Within these residual blocks, three convolutions and batch normalization are performed before the output x is processed through a leaky rectified linear unit activation function (Maas et al., 2013):

$$f(x) = max(0.3x, x).$$

This activation function introduces nonlinearities to the NN, enhancing its ability to understand complex patterns.

After passing through the encoder, an input patch of 128x128 pixels results in 128 features of 8x8 pixels. Utilizing this encoded output, the decoder performs transposed convolutions and upsampling. This results in an autoencoder with 16 trainable convolutional layers, with four of these located in the decoder. In total, the autoencoder comprised 635,953 trainable parameters, which were optimized for image reconstruction and precise compressed feature representations. This helped K-means to make meaningful cloud classifications, assigning a single label to each 128x128 image patch.

2.1.3 Training the classification model

To train CAOnet comprising an autoencoder and K-means clustering, a total of 15,200 swaths, split into 600,000 patches was used. This data covered the period from November to April for the years 2018 to 2023. First, a randomly sampled 85/15 split of the data was performed to create a training and test dataset. The autoencoder was individually trained on the whole training subset, and evaluated using the test split after one pass through the entire training dataset. The loss function optimized during training was a combination of mean squared error and Sobel loss similar to that used by Kurihana et al. (2022). Training of the autoencoder was stopped once the test loss had converged.

The encoder of the trained autoencoder was used to produce compressed feature representations that assisted K-means during training in producing meaningful clusters. In total, ten K-means models were trained, each with different numbers of predefined clusters ranging from 7 to 16. Finally, the classification of the ten K-means models were evaluated against a hand-labeled dataset to determine which of their clusters most closely aligned with CAO clouds, as further explained in the next section. Figure 1b shows an example of the satellite image classification process using a trained K-means model with 7 predefined clusters. Here, the red and pink clusters were evaluated to most closely align with CAO clouds.
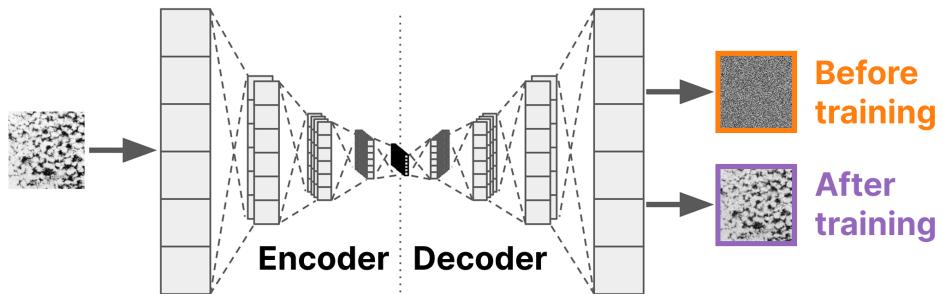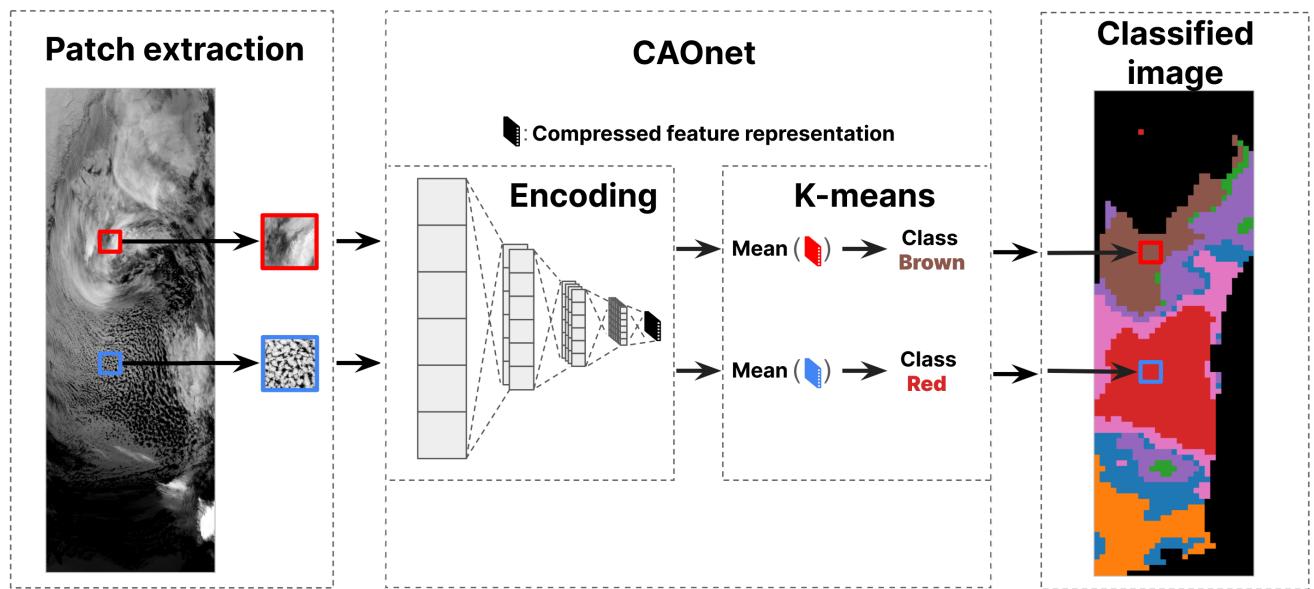
Figure 1. A visualization of a typical CAO and how it is processed by the unsupervised classification model. In panel (a) an image patch is fed through the autoencoder which before training produces noise (orange). The trained autoencoder has optimized its trainable parameters leading to an accurate reproduced image patch (purple). In panel (b), a classification example is shown, utilizing the encoder of the trained autoencoder. Two patches (marked in red and blue) are followed from extraction to encoding, K-means clustering and a final classified image. The blue patch containing typical CAO cellular cloud formation is classified as red, while the high clouds in the red patch are classified as brown. In this example, K-means clustering with seven clusters was used, with the red and pink cluster evaluated to most closely align with CAO clouds. Black regions of the classified image denote land, sea ice or regions outside of the study domain.

Table 1. Table showing the origin of the swaths making up the evaluation dataset.

Evaluation data subset	swaths per 5-year period	total swaths
CAO	25	100
No CAO	25	100
Random	25	100
Mixed	25	100
All	**100**	**500**

2.1.4 Evaluating the classification model

Even though the autoencoder and K-means clustering are unsupervised machine learning methods, their classifications require inspection and evaluation to acquire meaningful information about CAO clouds. To evaluate K-means clustering with different predefined numbers of clusters, a hand-labeled dataset was generated. To make sure there were no changes in the model's ability to classify CAO clouds as a result of potential climatological shifts, five datasets covering 5-year periods from 2000 to 2024 were created. To guarantee that each of the five subsets contained cases with and without CAOs, one K-means model with 14 clusters was randomly selected to identify satellite swaths containing CAOs. This involved manually inspecting six example images of CAOs, revealing four clusters from the example model aligning with CAO clouds (see Fig. D1).

By utilizing the randomly chosen K-means model with 14 predefined clusters, images containing CAOs (more than 30 coherent CAO patches) and those without CAOs (fewer than 2 coherent CAO patches) were identified based on predictions of the four clusters aligning with CAOs. This process resulted in CAO subsets and no-CAO subsets for each 5-year period. Additionally, random subsets were created, including randomly sampled swaths from each 5-year period. The final mixed subsets contained randomly sampled swaths for each 5-year period with equal probability corresponding to the three other subsets. In total, this resulted in 500 swaths, with 100 swaths for each 5-year period. From these 100 swaths, 25 swaths corresponded to each of the CAO, no CAO, random, and mixed subsets, as shown in Tab. 1.

The created evaluation dataset was presented to a labeler through an interactive website. The instructions were to draw regions where they believed they observed CAO-related clouds such as closed and open cells. Although closed and open cells not associated with CAOs may occur in the study region, these can be easily distinguished from CAO-related clouds during the labeling process. By recognizing cloud streets and clouds moving off the sea ice edge, it was ensured that mostly CAO related cellular structures were included in the CAO-specific evaluation dataset. The evaluation results were then compared with different K-means models, identifying the clusters that most accurately represented CAO clouds by calculating several score metrics. Consequently, when referring to closed and open cells hereafter, these are CAO-related cellular clouds.

To quantify uncertainty due to the relatively small evaluation dataset, swath-level bootstrap resampling was performed. From the original 500-swath evaluation dataset, 10,000 new 500-swath bootstrapped replicates were sampled with replacement. For each replicate, score metrics were calculated, before the standard deviation of the 10,000 bootstrap metric values was computed to provide a bootstrap estimate of the standard error.

2.1.5 Evaluation score metrics

The score metrics used to choose the most optimal clusters for CAO cloud classifications were the Matthews correlation coefficient (MCC), precision, recall and F_β scores. The MCC is a measure of association of two binary variables and defined through the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) as:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}}. \tag{1}$$

Precision is defined as the fraction of relevant predicted instances over all retrieved instances, while recall is defined as the fraction of relevant predicted instances over all relevant instances:

$$\text{precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{2}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}. \tag{3}$$

Utilizing the precision and recall, the F_β score is a measure of predictive performance in binary classification analysis and calculated as:

$$F_\beta = (1+\beta^2)\frac{\text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}}. \tag{4}$$

Here, β is a parameter which denotes the relative importance of recall compared to precision. Although the F_1 score with $\beta = 1$ is typically used, a final β of 1/1.75, valuing precision 1.75 times more than recall was chosen. This reflects a willingness to miss some true positives in order to reduce the number of false positives. It is expected that the model will be able to accurately classify clear CAO cases, and by prioritizing precision over recall, the risk of false positives biasing the data is minimized even if it results in oversight of less typical CAO cloud structures.

Finally, a combination of the F_β and MCC score can be calculated for a final evaluation score. By normalizing F_β and MCC over all models to be evaluated, the MCC and F_β scores equally contribute to the combined score (s) given as:

$$s = \frac{\text{MCC}/\text{MCC}_{\text{max}} + F_\beta/F_{\beta,\text{max}}}{2}, \tag{5}$$

where MCC_{max} and $F_{\beta,\text{max}}$ are the highest MCC and F_β scores found for all the evaluated models.

2.1.6 Producing a binary CAO classification database

After the best performing K-means model and associated CAO clusters were chosen, the final optimized CAOnet was settled on to make predictions on all MODIS swaths for autumn (Sep., Oct., Nov.), winter (Dec., Jan., Feb.) and spring (Mar., Apr., May) between March 2000 and February 2025. These predictions were re-gridded onto a 100 km resolution grid of the Arctic, creating a daily binary CAO classification database.

Potential biases may be introduced to this database through uneven MODIS coverage, confusion between clouds, land, and sea ice in coastal regions, and the uneven distribution of missing data. To address these issues, several processing methods were implemented.
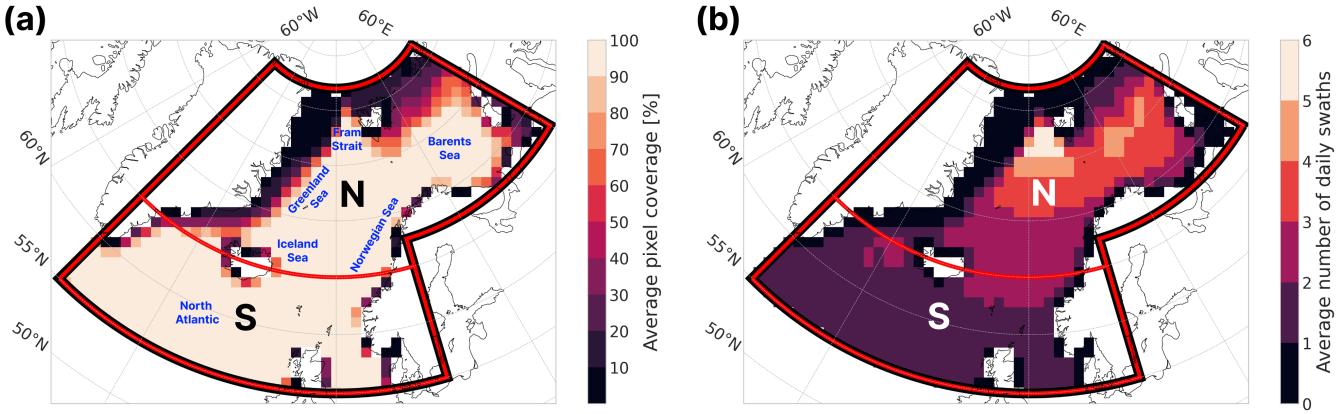
Figure 2. Average daily grid point coverage (a) and number of MODIS swaths (b). Low coverage can be found in coastal areas where patches typically contain less than 95% open ocean. S and N define a southern and northern subregion for subsequent analysis. Names of the focus regions for this study are marked in blue.

First, to prevent uneven MODIS coverage, a day was defined as extending from 01:00 UTC to 00:59 UTC the following day. This ensured that all grid points in the study areas experienced at least one overpass each day, given the orbit of the Terra satellite.

Second, to avoid bias from more overlapping MODIS swaths at higher latitudes, as shown in Fig. 2b, random sampling among overlapping swaths was performed. This guaranteed that every grid point received only one classification per day, resulting in a uniform database.

Third, to minimize the influence of sea ice and land on cloud classifications, patches were discarded if their open ocean fraction was less than 95%. To calculate this fraction, the sea ice concentration dataset from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 2 (DiGirolamo et al., 2022) was used. This resulted in lower data coverage for grid points near the sea ice edge and land, as shown in Fig. 2a.

Finally, to prevent the influence of missing data and the variability of the sea ice edge from affecting further analysis, all grid points with more than 10% missing data over the whole study period were discarded. This led to the final grid points shown in the 90-100% bin in Fig. 2a.

2.1.7 Evaluating CAO classification based on M index

To compare the phenomenological approach to a reanalysis approach for CAO cloud classification, CAO classifications were made using the M index calculated from MERRA-2 reanalysis. To optimize its performance, various index thresholds (M_{thr}) were evaluated against the evaluation dataset:

$$M = \theta_{SST} - \theta_{850} \geq M_{\mathrm{thr}}, \tag{6}$$

where θ_{SST} represents the potential sea surface temperature and θ_{850} the potential air temperature at 850 hPa. The threshold yielding the best score (s) according to Eq. 5 was then used to create a M index binary CAO classification database for the same grid points as those used in CAOnet. Here, the average M index for any given day and grid point was used. To ensure a fair comparison between the two CAO databases, only days and regions with MODIS coverage were considered for the M index as well.

It is also important to note that in situations where overlying clouds were present, it was impossible to determine the presence of CAO clouds below. While this limitation skewed the M index scores towards scenarios without overlying clouds, it optimized the final M index threshold for conditions where CAO clouds are especially relevant for cloud radiative properties.

2.2 CAO climatology

To produce a climatology of CAO clouds, both CAOnet and the MERRA-2 M index were used. Similar to Papritz and Spengler (2017), an extended winter stretching from November through April was used for the climatological analysis. For each grid point, a relative frequency of occurrence (RFO) was calculated, representing the fraction of days with CAO coverage relative to the number of days MODIS had coverage for that grid point. For MERRA-2, two M index thresholds of 0 K and 3.75 K called M_0 and $M_{3.75}$ were used, where the latter threshold was selected based on the model evaluation as described in Sect. 2.1.7.

To further underline the importance of CAO clouds, their contribution to total cloud cover was investigated. To estimate this, a monthly climatology of the total cloud cover during their presence was calculated using MERRA-2 reanalysis data. As the CAOnet database has a daily temporal resolution, the 24-hour MERRA-2 dataset was averaged to provide daily mean cloud coverages for every CAOnet grid point.

2.3 Trend analysis

To include the study of latitudinal variations of cloud cover trends and their impact on the radiative balance, the entire domain was divided into a southern and northern region, as shown in Fig. 2. The northern region was chosen based on typical CAO trajectories, capturing CAOs earlier on in their development due to the proximity to the sea ice edge. This typically results in larger concentrations of closed cells. In contrast, the southern region was chosen for its distance from the sea ice edge, capturing CAOs later on in their development (Murray-Watson et al., 2023), where a greater prevalence of open cells are expected (Brümmer, 1999). Finally, the binary classification databases were used to make three daily CAO cloud coverage datasets: the entire region, the southern region and the northern region.

Utilizing the CAO cloud cover fraction of each region, trends were calculated using the median of pairwise slopes method, also known as the Theil-Sen trend estimator T_{TS} (Theil, 1950; Sen, 1968). This trend estimator is non-parametric, meaning it is independent of the distribution of the data, making it widely applied in climate data analysis (Gilbert, 1987; Yue et al., 2002; Collaud Coen et al., 2020). It is estimated using daily mean coverage of CAO clouds across the three regions, calculated as the

median of all possible pairwise slopes:

$$\text{T}_{\text{TS}} = \text{median}\left(\frac{y_i - y_j}{x_i - x_j}\right), \tag{7}$$

where y_i denotes the cloud cover fraction on day x_i. Additionally, a confidence interval for this trend was estimated as the interval containing α (i.e. 95%) of the pairwise slopes, for which the median is represented as the trend in Eq. 7.

To test the significance of the trend estimation, the Mann-Kendall test was used (Mann, 1945; Kendall, 1948). This method requires no specific distribution of the data, but must be applied to serially independent data (Collaud Coen et al., 2020). To account for this, prewhitening algorithms have been developed to reduce the influence of autocorrelation on the significance level of the derived trend. Such a prewhitening method is described in Yue et al. (2002), where the data is processed before performing the Mann-Kendall test. Following Yue et al. (2002), the estimated trend T_{TS} was calculated using Eq. 7, before being removed from the time series X_t and creating the detrended time series X'_t:

$$X'_t = X_t - \text{T}_{TS} \cdot t. \tag{8}$$

The lag-1 autocorrelation (r_1) was then calculated and used to produce the independent series Y'_t:

$$Y'_t = X'_t - r_1 X'_{t-1}. \tag{9}$$

The trend was then added to the independent series:

$$Y_t = Y'_t + \text{T}_{TS} \cdot t, \tag{10}$$

which was used to asses the trend significance using the Mann-Kendall test. The trend was deemed significantly different from 0 when the resulting p-value was less than $\alpha = 0.05$. Additionally, to study seasonality, the seasonal Mann-Kendall test (Hirsch et al., 1982) was used in order to acquire trends for autumn, winter and spring.

3 Results and discussions

3.1 Model evaluation

The final configuration and structure of CAOnet was determined based on the model's accuracy. It was found that CAOnet performed best when it was implemented with an autoencoder-K-means combination with 7 clusters, of which 2 are defined as CAOs. This combination achieves an accuracy and corresponding bootstrapped standard error of 85.4±0.5%, a recall of 72.4±1.3%, a true negative rate of 90.4±0.6% and precision of 74.6±1.3%, as can be seen in Fig. 3a. Meanwhile when evaluating previously used CAO criteria for the M index, it was found that a threshold of $M > 3.75\,\text{K}$ ($M_{3.75}$) performed the best, in contrast to the commonly used threshold of $M > 0\,\text{K}$ (M_0) (e.g. Fletcher et al., 2016a; Murray-Watson and Gryspeerdt, 2024). With $M_{3.75}$, an accuracy of 78±0.7%, a recall of 71.1±1.7%, true negative rate of 80.7±1.1% and precision of 59.0±1.2% is reached (see Fig. 3b), which is better than when using a threshold of M_0 reaching an accuracy of 67.4±0.9%, a recall of 93.5±0.8%, true negative rate of 57.3±1.5% and precision of 46.0±1.2% (see Fig. 3c).
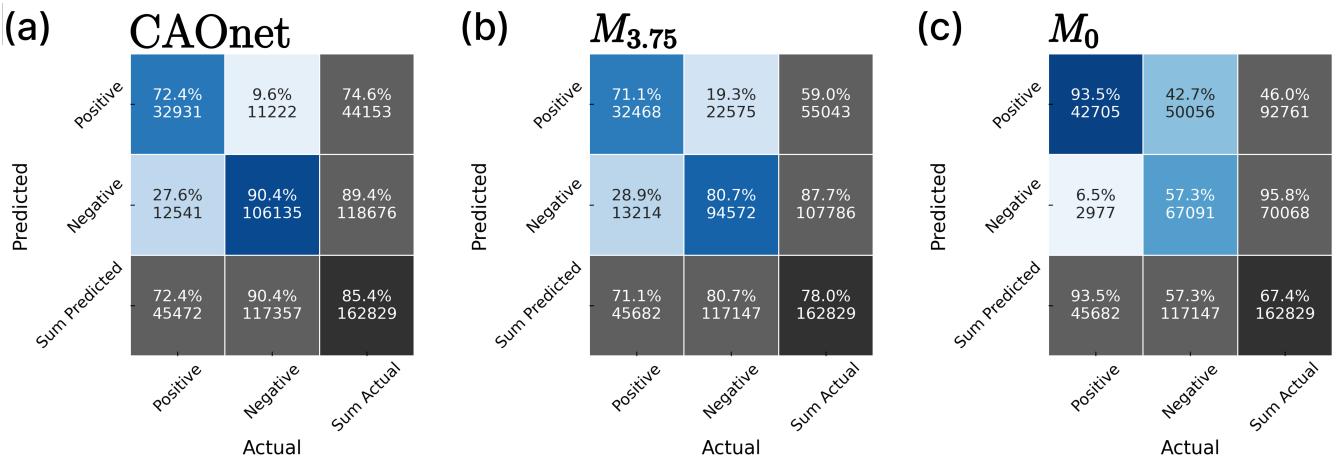
**Figure 3.** Confusion matrices for CAOnet (a), MERRA-2 with $M > 3.75\,\text{K}$ (b), and MERRA-2 with $M > 0\,\text{K}$ (c). The y-axis represents predicted classes by the models, while the x-axis represents the labeled classes. Positive corresponds to classified CAO, while negative means no CAO was classified. For each colored box, the lower number corresponds to number of classified patches, while the upper percentage corresponds to the rate of that predicted class relative to the actual class. The lower gray row shows the recall and true negative rate (upper) and number of patches labeled as that actual class (lower). The rightmost gray column shows the precision for positive and negative prediction (upper) and number of patches corresponding to that predicted class (lower). Finally the lower right corner in dark gray shows the total accuracy (upper) and total number of patches classified (lower).

The precision metrics highlight the strengths of CAOnet over the M index. For $M_{3.75}$ the precision is only 59.0%, indicating that close to half of all CAO predictions made are false positives. Furthermore, the reduction in precision to 46.0% for M_0, suggests that the majority of CAO predictions made using this threshold are false positives. This has significant implications for studies of CAO clouds employing the M index. As $M > 0$ is frequently used, these findings emphasize the importance of selecting a higher threshold and considering how any chosen threshold influences the uncertainties of such analyses.

For further evaluation, visual inspection was performed on two typical CAO cases found when screening through the hand labeled dataset. Figure 4 shows one of these cases where CAOnet and the MERRA-2 M index struggle to accurately predict CAOs, where disagreements with the evaluation dataset mostly show up as false negatives. $M_{3.75}$ in Fig. 4c appears to struggle with capturing the more open cellular structures (shown in orange), especially downwind in the lower right corner of the swath. By lowering the M index threshold to $0\,\text{K}$, these open cells are captured (see Fig. 4d), but with a slight increase in false positives (shown in red), suggesting that the threshold of $3.75\,\text{K}$ may be too high to capture the open cells in this example.

Additionally, Fig. 4 shows an example where the labeler may have been too conservative in their labeling, underlining the issue of subjective bias. Region 1 in panels a and d reveals closed cell-looking clouds labeled by the M index, but not by the human labeler. Similarly, region 2 in panel b and c shows what appears to be initial closed cell development, which has been labeled by the M index but not by CAOnet or the human. This can be explained by the uncertainties associated with the subjectivity of human labeling. Stevens et al. (2020) showed that six individuals rarely reached unanimous agreement when
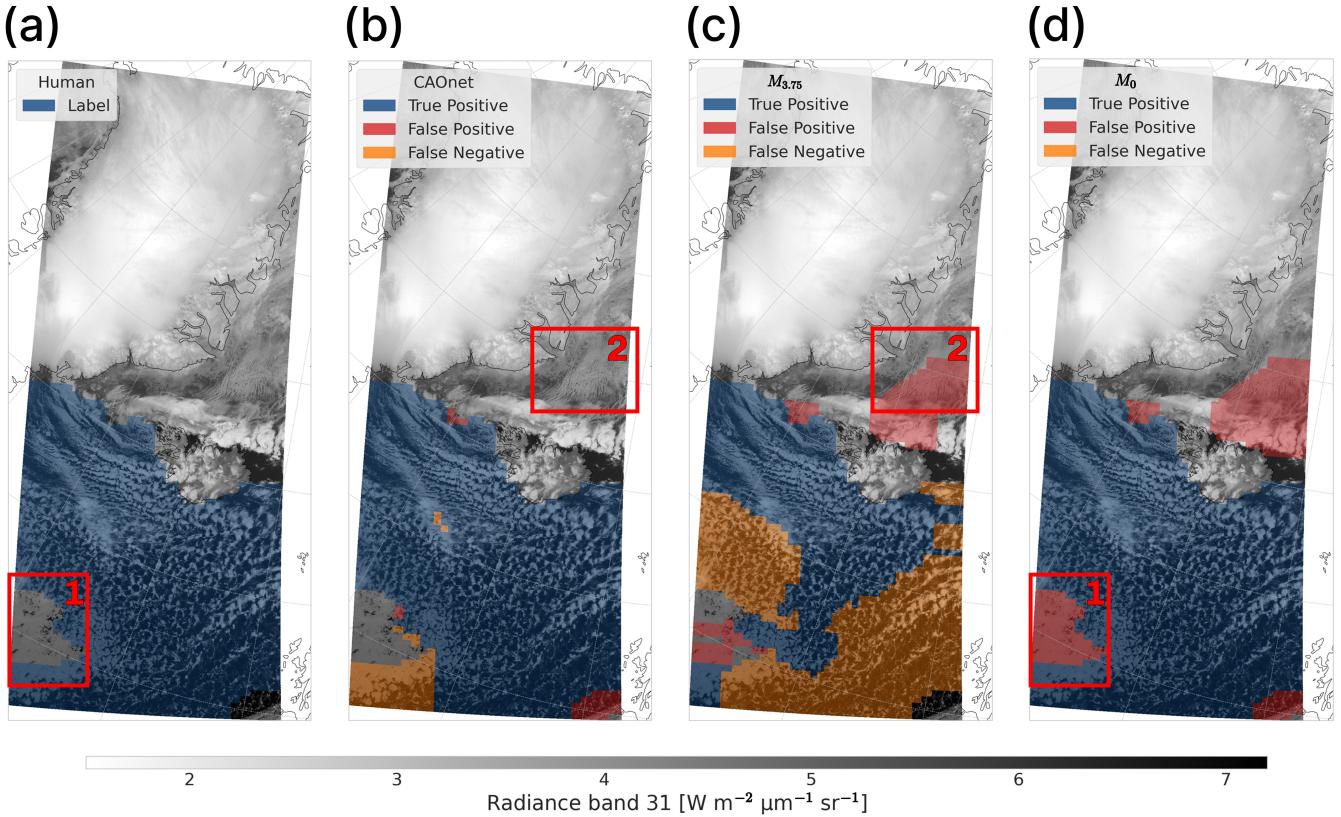
Figure 4. A CAO south of Iceland on 3 December 2013 22:50 UTC labeled by a Human (a), CAOnet (b) and MERRA-2 using a M index threshold of $3.75\,\mathrm{K}$ (c) and $0\,\mathrm{K}$ (d). Regions 1 and 2 highlight possible CAO cases not detected by CAOnet or the human labeler.

labeling mesoscale shallow clouds in the trade winds. This illustrates how subjectivity can lead to an evaluation dataset that may

310   not accurately reflect ground truth. Consequently, the classifications from a model such as CAOnet and MERRA-2 using the M index is likely to provide more stable and accurate classifications than that produced by a single labeler. However, the evaluation dataset may still indicate which model is better at classifying easily distinguishable clouds, such as clear cases of closed and open cells that are easily distinguishable by a human. As a result, the higher accuracy of CAOnet over $M_{3.75}$ combined with a significantly higher true negative rate (90.4% vs 80.7%) and precision (74.6% vs 59.0%), suggests that CAOnet has captured

315   more typical CAO clouds without the cost of more false positives that could bias its predictions.

Figure 5 shows a second classification example. Here, CAOnet (panel b) agrees well with the hand labeled data, while struggling to capture some of the initial closed cell development (marked as region 3). This is a general tendency of CAOnet seen in multiple classification examples throughout all seasons and years. It is a result of overlap between non-CAO cloud types and the cluster aligning with initial dense closed cell development. Discarding that cluster results in better overall accuracy, but

320   at the cost of missing CAO classifications close to the sea ice edge. This limitation must be considered in all further analysis,

as it may significantly impact the results. Especially, a shift in the prevalence of initial closed cell development off the sea ice edge as a result of climate change could result in lower or higher CAO detection over time, affecting the upcoming trend analysis. However, since the extent of initial cloud development is small compared to the total extent of a typical CAO, the correct classification of these clouds may be insignificant regarding the overall radiative impact of CAOs.

In Fig. 5c, $M_{3.75}$ also fails to detect some initial closed cell cloud development off the sea ice edge. While decreasing the M index threshold down to $0\,\text{K}$ solves some of the missing detection (see Fig. 5d), it also greatly increases false positives. One reason for these missed detections may be related to a typical shallow MBL of only a few hundred meters close to the sea ice edge (Fletcher et al., 2016a). As a result, the initial cloud formation may be a result of instabilities not stretching all the way to $850\,\text{hPa}$ in MERRA-2. Consequently, a higher pressure level (lower altitude) would have to be used for more efficient cloud detection. However, as the large turbulent fluxes act on the MBL, the potential temperature at a lower altitude may no longer describe the presence of instability and CAO clouds. This suggests that relying on the potential temperature from a single pressure level like 700, 800 or 850 as used in previous studies (e.g. Kolstad and Bracegirdle, 2008; Fletcher et al., 2016a; Papritz and Spengler, 2017), may not be optimal for CAO cloud classifications using MERRA-2 or other reanalysis products over the entire region where CAOs are found in the North Atlantic.

Additionally, the commonly used pressure levels for calculating the M index account only for the lower troposphere, missing out on a potential upper troposphere cloud layer. This limitation can lead to M index CAO classifications even when only high clouds are visible from space. Figure 5c, region 4 shows such an example where a high cloud extending from southern Norway along the CAO towards the sea ice edge to the northwest has been misclassified as a CAO. Although there may be CAO clouds present below, their radiative influence is greatly reduced by the high cloud above. Moreover, the high cloud region is quite large, giving a considerable contribution to the CAO database produced by the M index for this specific date. Even though an M index threshold most closely aligning with the label data has been selected, such discrepancies often occur, indicating that careful consideration is required when aiming to use the M index to study the radiative impact of CAO clouds.

3.2 Climatology

Having established the classification performance of CAOnet, $M_{3.75}$ and M_0 on individual images, they were applied to produce climatologies of CAO clouds for the months November to April from 2000 to 2025. In Fig. 6a, CAOnet shows RFO maxima close to 30% west of Lofoten along the Norwegian coast, and south of Iceland. These regions are located far from the sea ice edge and closely align with warmer SSTs (see Fig. A1b), where weaker CAOs are expected to be found (Papritz and Spengler, 2017). Similarly, a low RFO is observed in the Iceland and Greenland seas, where SSTs are expected to be relatively low. While this suggests a CAO cloud SST dependency, these findings may be a consequence of the higher CAOnet detection rate of open cells and their formation farther from the sea ice where higher SSTs are found. The findings do however show an interesting SST-aligning pattern, which suggests higher SSTs as a potential necessity for CAOnet detection and thereby open cell formation.

The Fram Strait is a region known for CAOs due to frequent instability in the MBL (Papritz and Spengler, 2017; Dahlke et al., 2022). However, this region shows relatively low RFO for CAOnet in Fig. 6a. This discrepancy may be attributed to
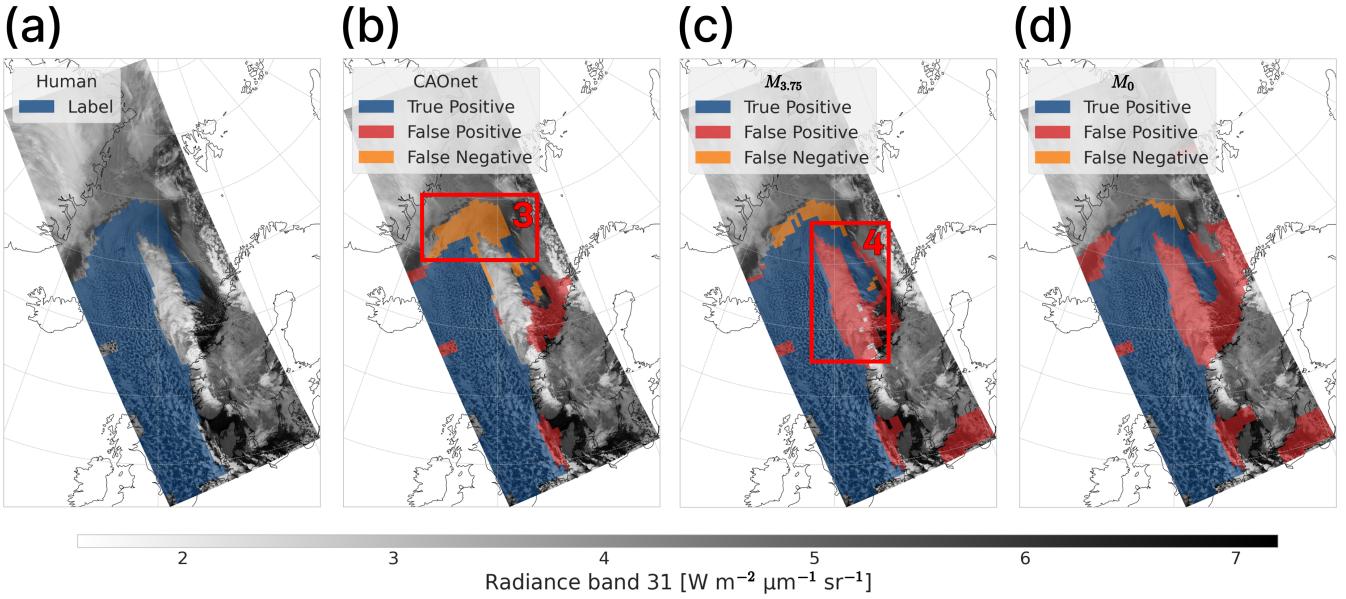
Figure 5. A CAO west of Norway 21 November 2008 21:00 UTC labeled by a Human (a), CAOnet (b) and MERRA-2 using a M index threshold of $3.75\,\mathrm{K}$ (c) and $0\,\mathrm{K}$ (d). Region 3 shows a typical example of CAOnet struggling in capturing initial closed cell development, while region 4 shows an example of the M index classifying a high cloud as a CAO.

earlier studies employing the M index which suggests instability without necessarily indicating cloud formation. It could also result from high clouds obscuring the CAO clouds below or from CAOnet failing to capture the initial development of closed cell clouds, as shown in Fig. 5. In contrast, $M_{3.75}$ and M_0 aligns with expectations, showing maxima in the RFO over the Fram Strait (Fig. 6 panel b and c). This maximum also extends towards the Norwegian coast, giving higher RFO in this region than CAOnet.

The Norwegian coast is a region where closed cells are expected to start breaking up into open cells. As illustrated in Fig. 4 and 5, CAOnet shows high sensitivity to open cell clouds, suggesting that CAOnet likely represents their RFO of approximately 30%. Consequently, $M_{3.75}$ overestimates this occurrence with an RFO of more than 35% . Additionally, the low $M_{3.75}$ CAO precision of just 59% (see Fig. 3b), implies a large number of false CAO classifications. These inaccuracies for $M_{3.75}$ may mainly result from high clouds obscuring the CAO clouds below (i.e. region 2 in Fig. 5c), giving non-aligning MERRA-2 classifications. While this is not necessarily wrong, such cases are less relevant when studying the radiative impact of CAO clouds.

Another region of disagreement between CAOnet and the M index is south of Iceland, where CAOnet (Fig. 6a) shows up to 20% higher RFO than $M_{3.75}$ (Fig. 6b). Given the CAOnet precision of 74.6% (Fig. 3a), it is unlikely that this is the result of large amounts of false positives. Instead, it may be explained by lacking CAO classifications by the M index, which is

370   shown as an example in Fig. 4. This Figure illustrates that the region south of Iceland can be dominated by open cells, which MERRA-2 may fail to detect, unless the M index threshold is lowered (see Fig. 6c).

    By decreasing the M index threshold to 0, the CAO occurrence south of Iceland increases by approximately 30%, but this comes at a significant cost to overall precision. With a precision of only 46.0% (see Fig. 3c), most of the M index CAO predictions using M_0 are false positives. This shows how lowering the M index threshold can be beneficial in certain cases such
375   as in Fig. 4, but with the consequences of lowering the total accuracy. This highlights the strength of the phenomenological approach using CAOnet as a CAO cloud predictor, provided that its limitations regarding missing classifications of initial closed cell clouds are acknowledged.

    To improve classifications by MERRA-2 in regions where open cells are expected, it may be beneficial to use another pressure level when calculating the M index. Near the sea ice edge, CAOs and the MBL depth are expected to be a few
380   hundred meters, increasing to up to 2 km downstream (Fletcher et al., 2016a). As a result, higher pressure levels may lead to missing classification close to the sea ice edge, making the 850 hPa potential temperature more suitable for detecting general instability and the potential for convective clouds. However, as open cells develop, the MBL may further deepen through mixing, suggesting that a lower pressure level might be better suited for detection of these open cells. Implementing a varying pressure level or M index threshold based on region and CAO lifecycle could address these issues, but would require extensive
385   research to determine accurate parameters.

    In total, the climatology from CAOnet (Fig. 6a) identifies regions where CAO clouds play a significant role for the radiative balance. This offers an alternative perspective to previous studies, focusing on MBL instability associated with CAOs. While instability remains an important factor for exchange of heat fluxes (Papritz and Spengler, 2017), CAOnet demonstrates its strengths in applications directly aimed at studying CAO clouds. Other than potentially missing important areas like the Fram
390   Strait and close to the sea ice edge, the identified areas of high RFO point to the coastal Norwegian Sea and the northern Atlantic south of Iceland as key regions for investigating the radiative impact of CAO clouds.

    Supplementing the RFO climatology, the average CAO cloud coverage for each month is investigated. As visualized in Figure 8a, the average CAO coverage peaks during the winter months for each model. This aligns with RFO maxima reported in earlier Arctic CAO studies (e.g. Kolstad et al., 2009; Dahlke et al., 2022), suggesting that both CAOnet and the M index
395   effectively capture the expected CAO seasonality. When further comparing the monthly coverage, the results from CAOnet and $M_{3.75}$ mostly agree. This is expected, as both the M index threshold and CAOnet have been optimized for the best fit with the evaluation data. Nevertheless, discrepancies are still observed. While both models show similar coverage in winter and spring, they disagree in autumn, with $M_{3.75}$ showing lower CAO coverage. This discrepancy may indicate that $M_{3.75}$ has a too high threshold to detect clouds associated with weaker CAOs, which may occur when Arctic air masses are warmer during
400   autumn. M_0 yields more comparable results to CAOnet for September, but the adjustment leads to a doubling of coverage in the following months. The alternative offered by CAOnet suggests that varying the M index thresholds by season may be necessary to optimize an M-index-based CAO cloud detection.

    Despite the disagreements between the models, the clear seasonal pattern obtained suggests a corresponding pattern in terms of the cloud radiative impact. This is further illustrated in Fig. 7, showing that CAO clouds contribute up to 20% of the region's
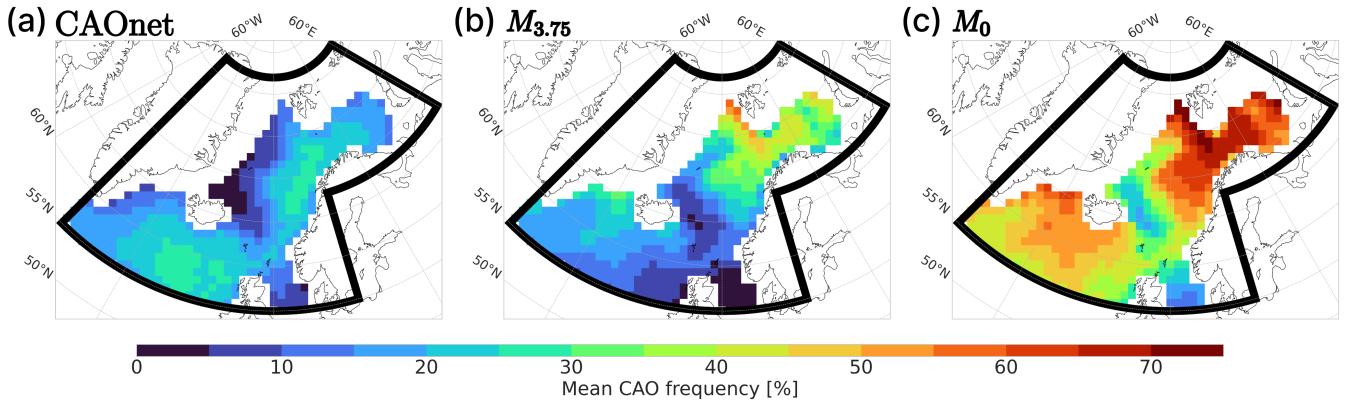
Figure 6. CAO climatologies for CAOnet (a) and the M index calculated using MERRA-2 reanalysis for the months November-April using data from March 2000 to February 2025. In panel b and c, M index thresholds of 3.75 and $0\,\mathrm{K}$ were used respectively. Regions with less than $90\,\%$ data availability because of missing MODIS swaths or sea ice have been discarded.
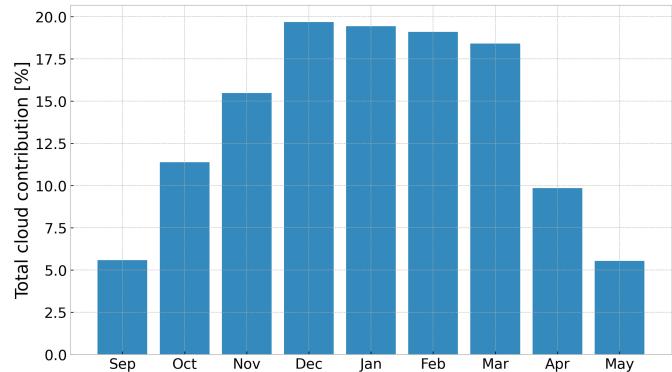
**Figure 7.** Monthly CAO contribution to the total cloud cover in the study region following CAOnet.

405   total cloud cover. With maximum coverage and cloud contribution occurring during winter and early spring, it is expected that potential trends during these months will be of the greatest significance for the radiative balance. Consequently, December to March are likely to play an important role in CAO influence on the Arctic climate, especially if trends in coverage are present. To further explore this, these months will be emphasized in the following sections, with a focus on CAOnet, while utilizing supporting results from MERRA-2 with $M_{3.75}$.
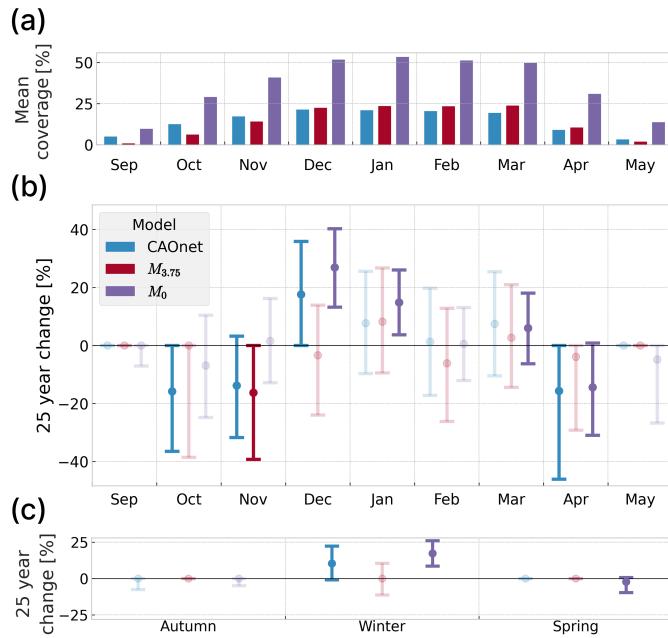
Figure 8. Trends in CAO coverage for September to May, 2000 to 2025 calculated using CAOnet (blue) and MERRA-2 with a M index threshold of 3.75 K (red) and 0 K (purple). Panel (a) shows the mean coverage of CAOs for the whole 25-year period. Panel (b) shows the relative 25-year change in CAO coverage for each month while panel (c) shows the relative 25-year change in CAO coverage for each season. Error bars indicate the 95 % confidence interval of the Theil-Sen slope, with the Theil-Sen estimate shown as dots. Significant trends are shown as solid colors, indicating a p-value less than 0.05 estimated using the non-parametric Mann-Kendall test. Insignificant trends are shown as transparent colors.

3.3 Trends

3.3.1 Arctic CAO cloud cover trends

To investigate how CAO clouds are responding to a warming Arctic, trends in their coverage are analyzed. Figure 8b shows trends in CAO cloud coverage for September-May as a 25-year relative change for CAOnet and the M index calculated using MERRA-2. CAOnet suggests a seasonal trend in which the shoulder months of the CAO season (October, November and April) show a significant relative decrease in cloud coverage of nearly 20 %, in contrast to the winter season and December that show increases of approximately 10 % and almost 20 % respectively. While the winter increase is not significant on a month-by-month basis, except for December, the nearly 10 % total winter season trend in Fig. 8c is significant. As the winter also contains the highest CAO coverage (see Fig. 8a), this relative trend has a great impact on the total area covered by CAO clouds.

To further support these findings, trends obtained using the M index can be examined. In Fig. 8b, $M_{3.75}$ generally agrees with the sign of the CAOnet trends, except for December, when CAOnet suggests a significant increase and the M index a

18

decrease. However, the confidence interval of this trend is well inside the confidence interval of CAOnet, suggesting a non-significant disagreement. When the M index threshold is reduced to $0\,\mathrm{K}$, the trend sign shifts to a strong and significant increase. However, as explored in Sect. 3.1, this threshold does not typically align well with CAO clouds. Nevertheless, the significance of the trend suggests an increase in areas with MBL instability for December across the Arctic.

Overall, the seasonality of the MBL instability and cloud cover trends, which is further supported by CAO cloud occurrence trends in Fig. C1, suggests clear seasonal drivers that may impact both the exchange of heat fluxes and the radiative balance of the Arctic.

3.3.2 Explaining the seasonality of Arctic trends

The observed trends motivate future studies to investigate the climatological parameters driving them. Especially, the seasonal nature of these trends underscores the potential importance of other season-dependent parameters such as temperatures and sea ice extent. Through CAOnet, future studies have access to a phenomenological classifier that provides an extensive database for correlation analysis with potential CAO-influencing climatological parameters. To aid future research in their analysis, we propose several factors for further investigation in order to better understand CAOs in a future Arctic.

First, Arctic air temperature warming tends to be greater (Chylek et al., 2009; Johannessen et al., 2016; Rantanen et al., 2022) and more surface-confined during winter than in the shoulder seasons (Graversen et al., 2008; Alexeev et al., 2012). This may serve as a wintertime destabilizing factor for the MBL, in contrast to the more vertically distributed warming during the shoulder seasons. However, increasing air temperatures close to the ocean surface may also limit moist convection and cloud formation, showing uncertainties in the impact of atmospheric warming profiles. By investigating the correlation between the warming profiles and CAO occurrence, the drivers behind the increasing prevalence of CAO clouds during winter and their decreasing prevalence during the shoulder seasons may be better understood.

Second, as indicated by CAOnet (Fig. 6a), CAO clouds are more frequently observed in regions of higher SSTs (Fig.A1b), revealing a potential SST dependence. However, as suggested by the CAO-SST correlation (Fig. A1a) and the relatively stable SSTs over the study period (Fig. A2), SSTs are unlikely the driver of the observed CAO trends. However, in the Tropics, Sandu and Stevens (2011) showed that closed-to-open cell transition occurred as a result of increasing SSTs. Consequently, as SSTs rise due to climate change, conditions may become more favorable for cumulus and open cells. This highlights the need for future research to investigate how changing SSTs affect the distribution of closed and open cell clouds, which is important for understanding the future radiative impact of CAOs.

Third, the impact of projected Arctic sea ice loss (DeRepentigny et al., 2016) could be studied using CAOnet. Observations indicate larger wintertime sea ice loss (Garcia-Soto et al., 2021), motivating future studies to conduct correlation analysis to assess how this loss has influenced the observed increase in wintertime CAO cloud cover. Investigating how total cloud cover and cloud properties evolve in relation to the location of the sea ice edge may help to quantify its effects on CAOs.

3.3.3 Radiative impacts of Arctic trends

Low-level clouds contribute the most to the Arctic surface radiative energy balance (Shupe and Intrieri, 2004), making trends in the low-level CAO clouds that contribute to up to 20 % of the total cloud cover in the study domain (Fig. 7) of high importance for Arctic warming. Their net radiative effect does however largely depend on solar radiation, which is negligible during winter for large portions of the study region. This results in a dominant longwave radiative warming from the increasing coverage of the CAO clouds during winter. With CAO presence accounting for up to 20 % of the region's clouds, they are likely responsible for a large longwave effect that may have contributed to an enhanced and region-dependent wintertime Arctic amplification (Rantanen et al., 2022). In contrast, during the shoulder months, solar radiation contributes significantly to the radiative balance, which may lead to a warming effect associated with the decreasing prevalence of CAO clouds.

However, changes in CAO cloud cover potentially align with shifts in other cloud types, complicating their radiative impact. This motivates future work to uncover the exact radiative impact of the observed CAO trends by using spaceborne radiative flux products such as from the clouds and earth radiant energy system instrument (CERES, Wielicki et al., 1996).

3.3.4 Trends and coverage in the north and south

To better understand changes in the radiative impact of CAO clouds, it is important to study both seasonal and latitudinal trends. As the North Atlantic ocean stretching to the Arctic maintains fairly stable temperatures throughout the year (Fig. A2), the outgoing longwave radiation from the ocean surface remains close to constant. However, the incoming solar radiation may vary greatly, particularly in the winter season across the domain between 55 and 80 °N. This latitudinal dependency motivates a separate trend analysis for the southern and northern part of the domain (see Fig. 2).

Figure 9 shows the trends and CAO cloud coverage for the northern (panels a-c) and southern regions (d-f). Similar to the whole domain in Fig. 8, the highest CAO coverage occurs in both regions during winter. However, there is a shift in maximum coverage towards late winter and spring in the northern region (a) for both the M index and CAOnet, while in the southern region CAOnet shows a shift towards early winter and autumn (d). In the North, the M index generally suggests more CAO coverage than CAOnet, whereas CAOnet suggests higher coverage than the M index in the South. This aligns with the climatology presented in Fig. 6, where the M index shows higher occurrences than CAOnet in the north and lower in the south.

In general, a higher occurrence of CAOs is expected farther north as a result of CAO initiation near the sea-ice and snow-covered surfaces, which are more prevalent in the north. However, as these clouds move over the ocean toward the south, they develop into open cells that may expand to cover large areas. Additionally, CAOs originating outside of the study area, specifically west of Greenland, can extend into the southern region as visualized in Fig. 5. This results in a higher cloud coverage in the South than the North, despite the expectation of higher CAO occurrence in the North. Consequently, the high cloud coverage indicated by CAOnet in the South (Fig. 9d), along with CAOnet demonstrating high sensitivity to open cell clouds in this region (Fig. 5b), enhances confidence in the results produced by CAOnet. This motivates the importance of the southerly trends (Fig. 9 panels e and f), especially when interpreting these findings as mostly reflecting changes in open cell coverage.
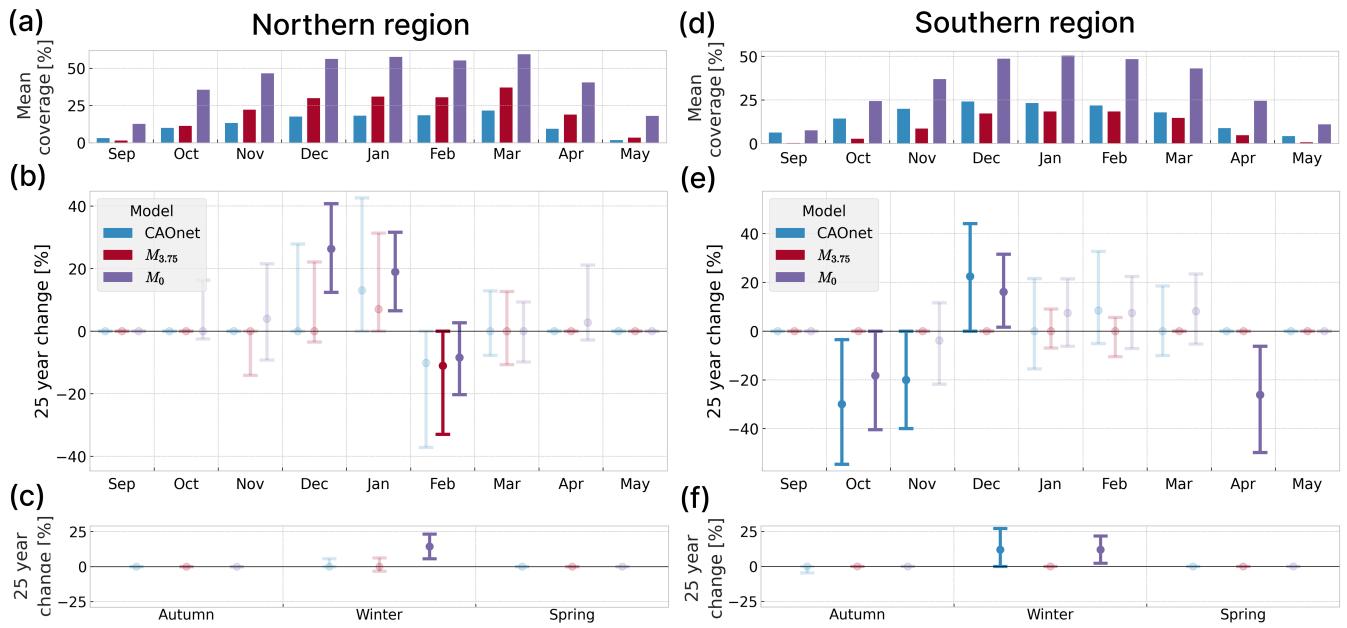
Figure 9. Similar to Fig. 8, but for the northern region north of 65 °N (a-c), and the southern region between 55 and 65 °N (d-f) as shown in Fig. 2. Significant trends are shown as solid colors, indicating a p-value less than 0.05 estimated using the non-parametric Mann-Kendall test. Insignificant trends are shown as transparent colors.

While $M_{3.75}$ sees a significant decrease in February, no general trend in CAO coverage is observed for CAOnet in the northern region (panels b and c). In contrast, M_0 reveals significant increasing trends for December and January and decreasing for February, as well as an increasing trend for the whole winter season combined. This suggests an increase in areas with MBL instability over the last 25 years, with corresponding insignificant trends in cloud cover as indicated by CAOnet. This can be explained by the M index trend contributions largely resulting from increasing areas of very weak instability, insufficient to form clouds.

In the southern region (Fig. 9 panels e and f), a seasonal pattern is observed, marked by an increase in cloud cover during winter and decrease in the shoulder months October and November. This pattern aligns with the seasonality of the whole domain (compare Fig. 8b), suggesting that the trends in the southern region are driving those observed over the entire domain. As the southern region was chosen for its distance from the sea ice edge, capturing more developed CAO clouds and open cells, the trends likely reflect changes in the prevalence of open cell clouds. Notably, significant increases in cloud cover are found in December (Fig. 9e) and the winter season combined (Fig. 9f), while a significant decrease is found in the shoulder months October and November (Fig. 9e). M_0 exhibits similar significant trends for October and December, while also suggesting a significant cloud cover decrease decrease for April. In contrast, $M_{3.75}$ suggests no overall trend, deviating from the two others and indicating that the M index threshold may be too high to capture the cloud structures typically found in the southern region.

21

The overall trends attributed to the southern region and its open cell coverage suggest potential changes in atmospheric circulation and the efficiency of cloud dissipation over the past 25 years. While shifts in circulation patterns may result in more or fewer CAO clouds reaching the South, these trends could also be caused by delayed or enhanced dissipation of open cell clouds. Regardless of the underlying cause, the increasing prevalence of open cells during winter and decreasing prevalence during the shoulder seasons will change the radiative impact of CAOs. Although these factors only provide plausible explanations for the observed trends, they emphasize the importance of understanding both circulation changes and cloud dissipation within CAOs, as these processes may themselves be influenced by climate change.

Consequently, further studies could utilize a phenomenological CAO classifier like CAOnet to directly assess changes in radiative effects of regions experiencing CAO trends. Additionally, by further developing this method to assess closed and open cells separately, closed and open cell trend attribution may be determined accurately. This would yield more precise insights into the future impacts of Arctic CAOs on both the local and global climate.

4   Conclusion

Clouds associated with CAOs are important for the Arctic radiative energy budget, particularly in the regions surrounding the Norwegian sea, Barents sea and Northern Atlantic due to their extensive coverage. In the rapidly warming Arctic, it becomes important to understand how these clouds respond to climate change, as changes in their prevalence and properties may either amplify or dampen local and global warming. To explore shifts in CAO cloud coverage, a phenomenological CAO cloud classification method named CAOnet has been developed. This approach, based on MODIS data and machine learning, has demonstrated promising results compared to the traditional marine cold air outbreak index (M) in detecting CAO clouds. Additionally, an M index threshold of 3.75 K has been identified for optimal detection of CAO clouds, providing future studies a basis for the instability required for CAO cloud formation.

By employing CAOnet, a CAO climatology focusing on clouds rather than MBL instability has been produced. In contrast to the M index calculated using MERRA-2 reanalysis, this climatology has provided an alternative perspective on CAOs, highlighting the Norwegian coast and the North Atlantic region south of Iceland as key areas for CAO clouds and their associated radiative impact. While these regions align with relatively high SSTs, as well as frequent open cells to which CAOnet shows a high sensitivity, low CAO-SST correlation is found. Although this suggests SSTs to be of little importance for CAO trends, future studies may find clearer correlations as more data becomes available.

Utilizing data from the past 25 years, CAO clouds have been found to contribute to up to 20 % of the total cloud cover in the study region. This underscores the importance of the observed trends, revealing a shortening of the CAO season as indicated by a 10 % increase in wintertime CAO cloud cover and nearly a 20 % decrease during the shoulder months of October, November and April. These shifts are in large part linked to changes in open cell cloud cover as well as changes in the MBL (as indicated by the M index). By utilizing CAOnet, future studies have an easy-to-use database to investigate the potential drivers of these trends, such as correlation analysis with atmospheric warming profiles and the position of the sea ice edge.

Due to the lack of incoming solar radiation during winter, the radiative impact of the observed wintertime cloud cover trends are likely dominated by the terrestrial radiative warming effect. This may have contributed to the anomalously strong and region-dependent wintertime Arctic amplification. Conversely, during the shoulder months when decreasing cloud cover is observed, the increased solar radiative effect introduces uncertainties regarding the overall radiative effect of the cloud cover trends.

While these trends do not directly indicate future changes in the Arctic's radiative energy balance, they clearly indicate that CAO clouds are influenced by climate change. This emphasizes the importance of accurately characterizing the radiative impact of these clouds and understanding their role in the local and global climate system. This motivates future work to utilize CAOnet together with spaceborne flux products such as from CERES, to accurately uncover the radiative impact of the observed CAO trends.

Code and data availability. Data, including CAOnet and MERRA-2 CAO masks, along with the code, are available from Zenodo at https://doi.org/10.5281/zenodo.16680335 (von der Lippe et al., 2025). MODIS level 1B calibrated radiances can be accessed at https://ladsweb.modaps.eosdis.nasa.gov/search/order/1/MOD021KM--61 (MODIS Characterization Support Team (MCST), 2017). NIMBUS sea ice concentration data is available at https://nsidc.org/data/nsidc-0051/versions/2 (DiGirolamo et al., 2022). The MERRA-2 reanalysis products are accessible at https://disc.gsfc.nasa.gov/datasets/M2T1NXSLV_5.12.4/summary (Global Modeling and Assimilation Office (GMAO), 2015)

Appendix A: Sea surface temperatures
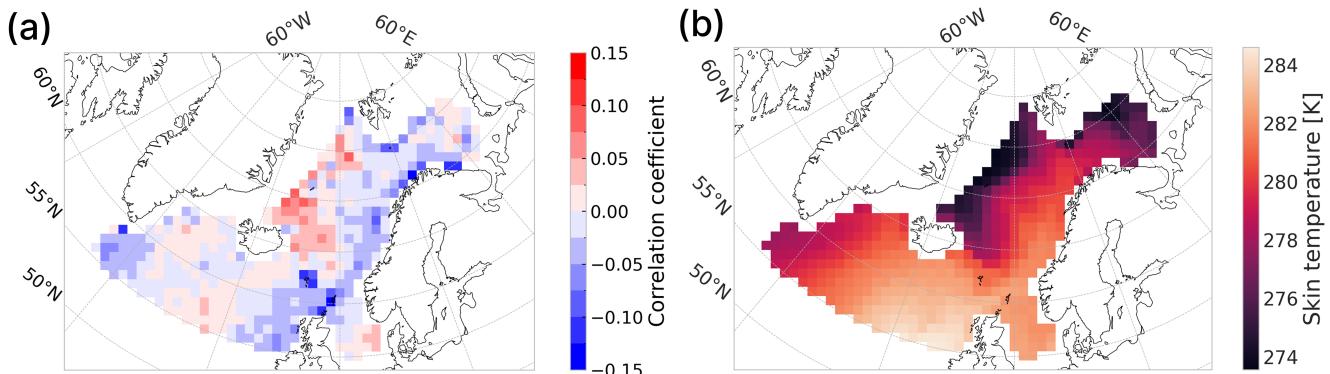
Figure A1. Panel a shows the SST-CAO correlation, with positive values indicating higher CAO occurrence for higher temperatures, while panel B shows the average skin temperature over the whole period.
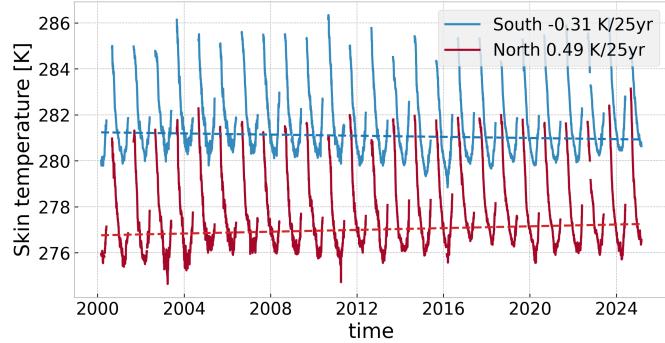
Figure A2. Variations in the average temperature for the southern and northern region for the months September through May, including the significant Theil-Sen trend estimate.

## Appendix B: Optimizing M index threshold
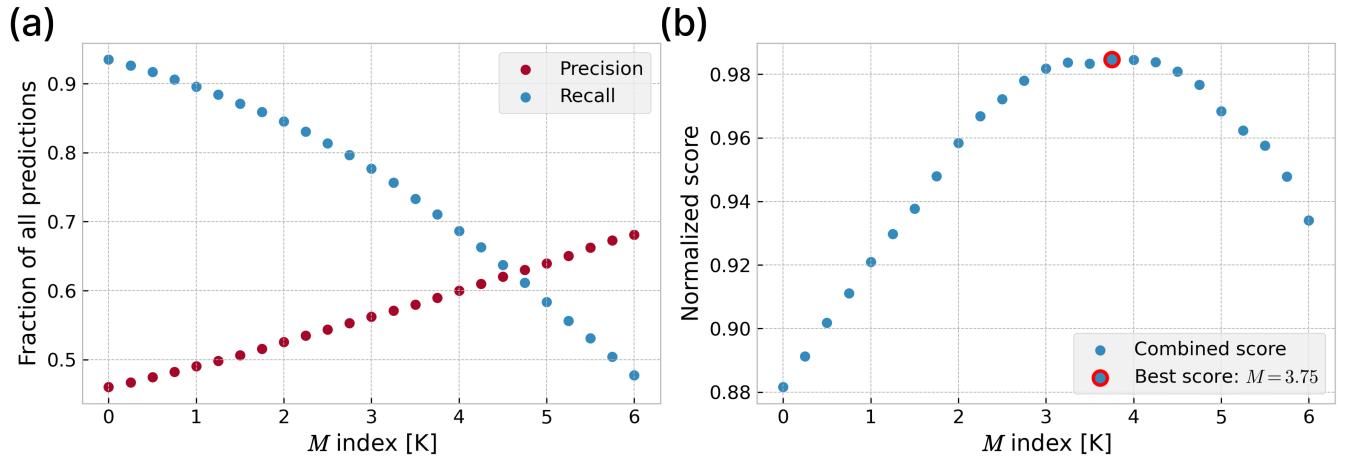
Figure B1. CAO evaluation scores calculated using different M index thresholds. In panel a, a higher precision comes at the cost of recall. This leads to the the most optimal threshold of 3.75 K for the normalized score in panel b.

Appendix C: Occurrence trends

In addition to the CAO cloud coverage trends presented in Fig. 8 and 9, monthly CAO cloud occurrence trends were calculated on a grid point by grid point basis, as shown in Fig. C1. This was accomplished using the median of pairwise slopes (Theil-Sen) method and Mann-Kendall test, as described in Sect. 2.3. However, as each grid point has values of either 1 or 0 depending on whether a CAO is present or not, and most grid points having zeros most of the time, most pairwise slopes will also be 0. Consequently, the final median and trend will also be 0. To overcome this limitation, monthly means for each grid point were calculated, resulting in monthly occurrence fractions that were used for the final trend estimate.

Furthermore, a p-value was calculated using the Mann-Kendall test. A trend was deemed significant if it satisfied a false discovery rate of 0.2, following the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995). This method is recommended for addressing multiple hypothesis testing in atmospheric sciences (Wilks, 2016), allowing control over the expected fraction of false positives among the significant trends to be 20%.

The resulting significant trends indicate a decreasing occurrence of CAO clouds in the southern region for both October and May over the past 25 years. This aligns with the cloud coverage trends for October shown in Fig. 8, and further supports the observed trend seasonality, suggesting a shortening of the CAO season.
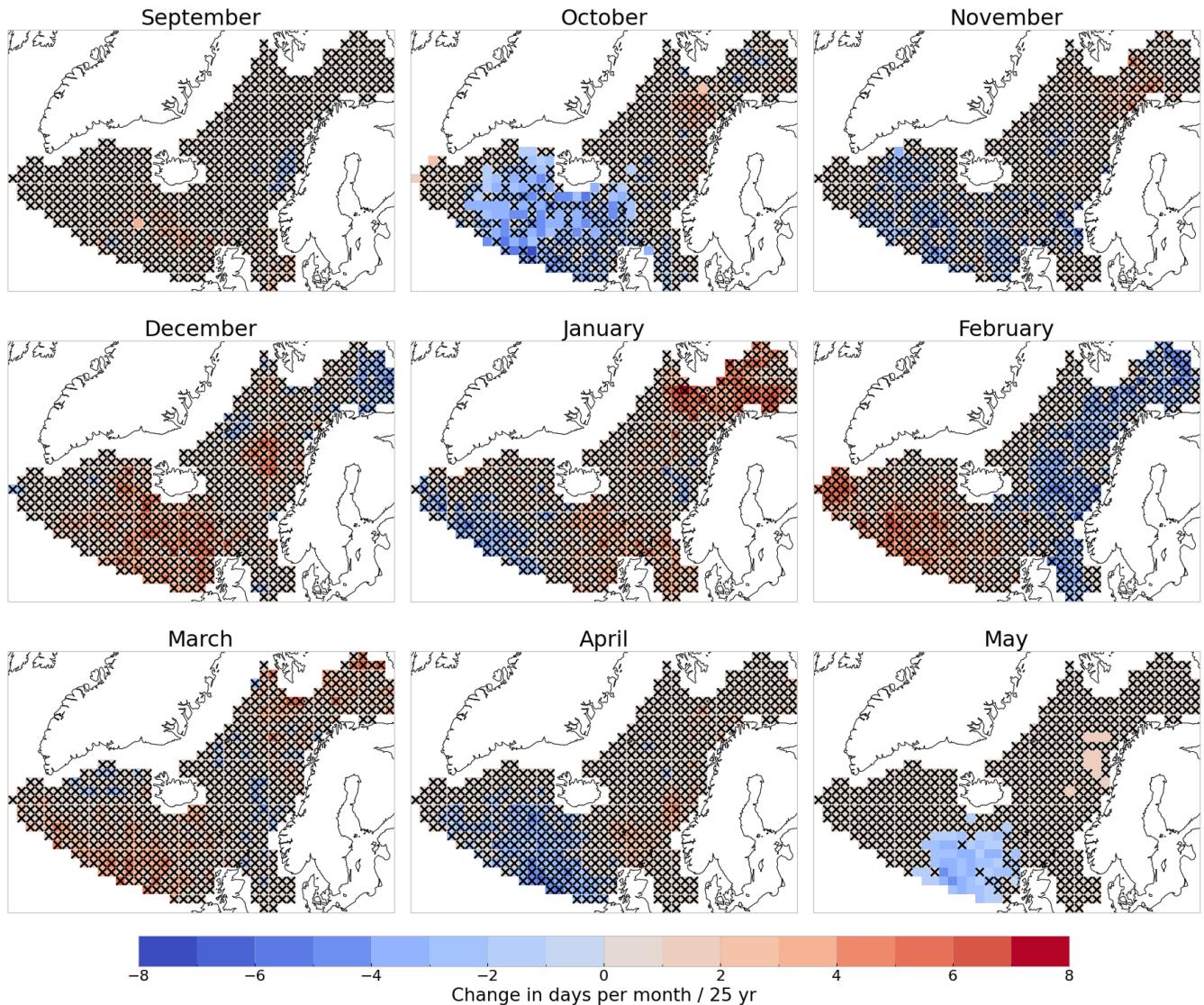
Figure C1. Trends for each month calculated using the Theil-Sen estimator. Stippled points are insignificant by not satisfying a false discovery rate of 0.2.

Appendix D: Manual inspection of CAO images
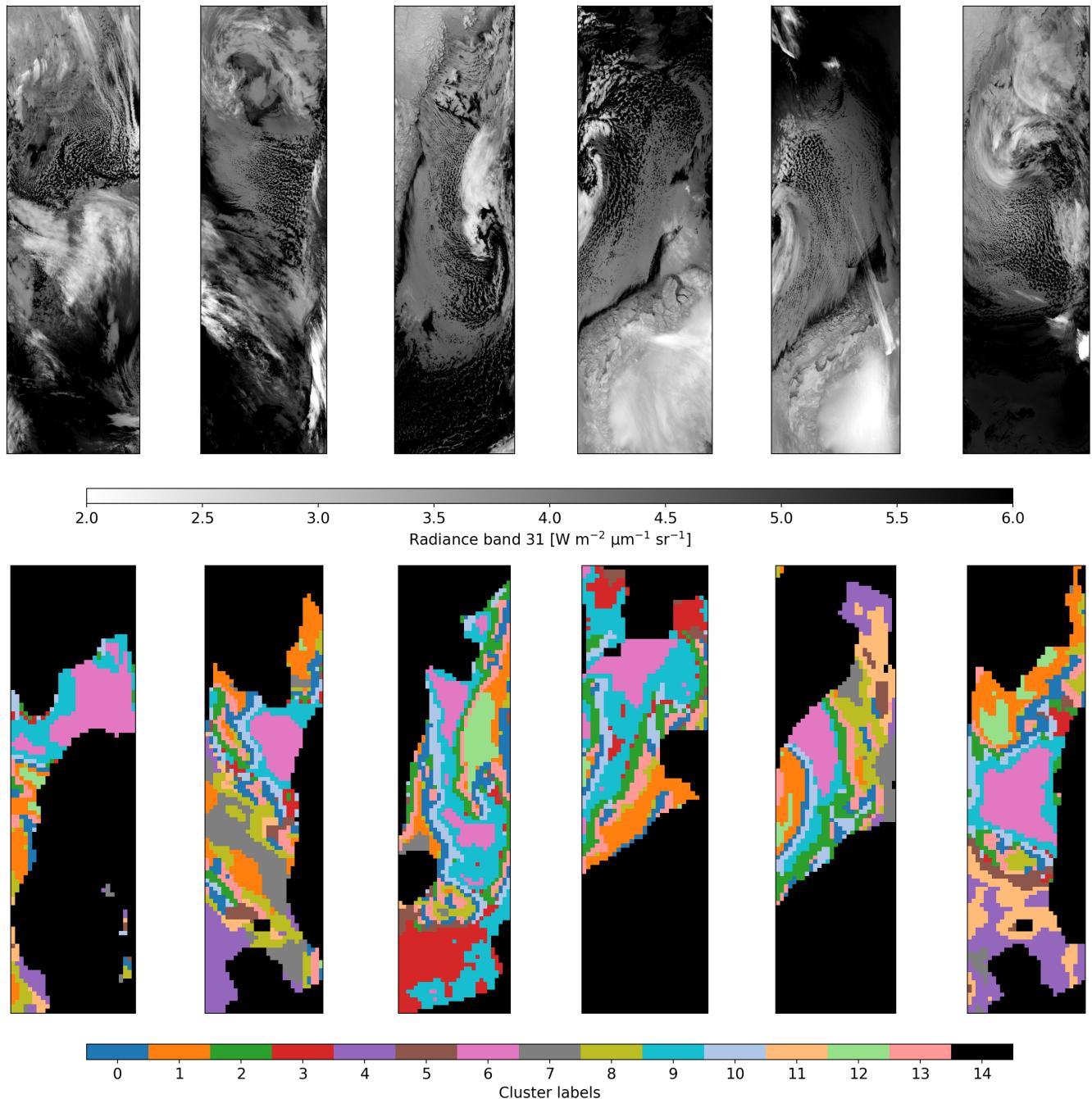
Figure D1. Example images of CAOs used to identify CAO-aligning clusters to facilitate the creation of a labeling dataset. Out of the total 14 K-means clusters, manual inspection identified clusters 2, 6, 9 and 10 to align with typical CAO clouds.

References

Abel, S. J., Boutle, I. A., Waite, K., Fox, S., Brown, P. R. A., Cotton, R., Lloyd, G., Choularton, T. W., and Bower, K. N.: The Role of Precipitation in Controlling the Transition from Stratocumulus to Cumulus Clouds in a Northern Hemisphere Cold-Air Outbreak, Journal of the Atmospheric Sciences, 74, 2293 – 2314, https://doi.org/10.1175/JAS-D-16-0362.1, place: Boston MA, USA Publisher: American Meteorological Society, 2017.

Alexeev, V. A., Esau, I., Polyakov, I. V., Byam, S. J., and Sorokina, S.: Vertical structure of recent Arctic warming from observed data and reanalysis products, Climatic Change, 111, 215–239, https://doi.org/10.1007/s10584-011-0192-8, publisher: Springer, 2012.

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society. Series B (Methodological), 57, 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x, publisher: [Royal Statistical Society, Oxford University Press], 1995.

Bretherton, C. S. and Wyant, M. C.: Moisture transport, lower-tropospheric stability, and decoupling of cloud-topped boundary layers, Journal of the atmospheric sciences, 54, 148–167, https://doi.org/10.1175/1520-0469(1997)054<0148:MTLTSA>2.0.CO;2, publisher: American Meteorological Society, 1997.

Brümmer, B.: Roll and Cell Convection in Wintertime Arctic Cold-Air Outbreaks, Journal of the Atmospheric Sciences, 56, https://doi.org/10.1175/1520-0469(1999)056<2613:racciw>2.0.co;2, section: Journal of the Atmospheric Sciences, 1999.

Chylek, P., Folland, C. K., Lesins, G., Dubey, M. K., and Wang, M.: Arctic air temperature change amplification and the Atlantic Multidecadal Oscillation, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL038777, publisher: Wiley Online Library, 2009.

Collaud Coen, M., Andrews, E., Bigi, A., Martucci, G., Romanens, G., Vogt, F. P., and Vuilleumier, L.: Effects of the prewhitening method, the time granularity, and the time segmentation on the Mann–Kendall trend detection and the associated Sen's slope, Atmospheric measurement techniques, 13, 6945–6964, https://doi.org/10.5194/amt-13-6945-2020, publisher: Copernicus Publications Göttingen, Germany, 2020.

Dahlke, S., Solbès, A., and Maturilli, M.: Cold air outbreaks in Fram Strait: Climatology, trends, and observations during an extreme season in 2020, Journal of Geophysical Research: Atmospheres, 127, e2021JD035 741, https://doi.org/10.1029/2021JD035741, publisher: Wiley Online Library, 2022.

DeRepentigny, P., Tremblay, L. B., Newton, R., and Pfirman, S.: Patterns of sea ice retreat in the transition to a seasonally ice-free Arctic, Journal of Climate, 29, 6993–7008, https://doi.org/10.1175/JCLI-D-15-0733.1, 2016.

DiGirolamo, N., Parkinson, C., Cavalieri, D., Gloersen, P., and Zwally, H.: Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 2, https://doi.org/10.5067/MPYG15WAA4WX, 2022.

Eastman, R., McCoy, I. L., and Wood, R.: Wind, Rain, and the Closed to Open Cell Transition in Subtropical Marine Stratocumulus, Journal of Geophysical Research: Atmospheres, 127, e2022JD036 795, https://doi.org/10.1029/2022JD036795, publisher: John Wiley & Sons, Ltd, 2022.

Fletcher, J., Mason, S., and Jakob, C.: The climatology, meteorology, and boundary layer structure of marine cold air outbreaks in both hemispheres, Journal of Climate, 29, 1999–2014, https://doi.org/10.1175/jcli-d-15-0268.1, 2016a.

Fletcher, J. K., Mason, S., and Jakob, C.: A climatology of clouds in marine cold air outbreaks in both hemispheres, Journal of Climate, 29, 6677–6692, https://doi.org/10.1175/JCLI-D-15-0783.1, publisher: American Meteorological Society, 2016b.

610 Frey, R. A., Ackerman, S. A., Liu, Y., Strabala, K. I., Zhang, H., Key, J. R., and Wang, X.: Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for collection 5, Journal of atmospheric and oceanic technology, 25, 1057–1072, https://doi.org/10.1175/2008JTECHA1052.1, 2008.

Garcia-Soto, C., Cheng, L., Caesar, L., Schmidtko, S., Jewett, E. B., Cheripka, A., Rigor, I., Caballero, A., Chiba, S., Báez, J. C., Zielinski, T., and Abraham, J. P.: An Overview of Ocean Climate Change Indicators: Sea Surface Temperature, Ocean Heat Content, Ocean pH, Dis-
615 solved Oxygen Concentration, Arctic Sea Ice Extent, Thickness and Volume, Sea Level and Strength of the AMOC (Atlantic Meridional Overturning Circulation), Frontiers in Marine Science, 8, https://doi.org/10.3389/fmars.2021.642372, publisher: Frontiers, 2021.

Geerts, B., Giangrande, S. E., McFarquhar, G. M., Xue, L., Abel, S. J., Comstock, J. M., Crewell, S., DeMott, P. J., Ebell, K., Field, P., Hill, T. C. J., Hunzinger, A., Jensen, M. P., Johnson, K. L., Juliano, T. W., Kollias, P., Kosovic, B., Lackner, C., Luke, E., Lüpkes, C., Matthews, A. A., Neggers, R., Ovchinnikov, M., Powers, H., Shupe, M. D., Spengler, T., Swanson, B. E., Tjernström, M., Theisen, A. K., Wales,
620 N. A., Wang, Y., Wendisch, M., and Wu, P.: The COMBLE Campaign: A Study of Marine Boundary Layer Clouds in Arctic Cold-Air Outbreaks, https://doi.org/10.1175/BAMS-D-21-0044.1, section: Bulletin of the American Meteorological Society, 2022.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A. M. d., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and
625 Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419 – 5454, https://doi.org/10.1175/JCLI-D-16-0758.1, place: Boston MA, USA Publisher: American Meteorological Society, 2017.

Gilbert, R. O.: Statistical methods for environmental pollution monitoring, John Wiley & Sons, 1987.

Global Modeling and Assimilation Office (GMAO): MERRA-2 tavg1_2d_slv_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Single-Level Diagnostics V5.12.4,, https://doi.org/10.5067/VJAFPLI1CSIV, 2015.

630 Graversen, R. G., Mauritsen, T., Tjernström, M., Källén, E., and Svensson, G.: Vertical structure of recent Arctic warming, Nature, 451, 53–56, https://doi.org/10.1038/nature06502, publisher: Nature Publishing Group UK London, 2008.

Hartmann, J., Kottmeier, C., and Raasch, S.: Roll vortices and boundary-layer development during a cold air outbreak, Boundary-Layer Meteorology, 84, 45–65, https://doi.org/10.1023/A:1000392931768, publisher: Springer, 1997.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, CoRR, abs/1512.03385, http://arxiv.org/abs/1512.
635 03385, arXiv: 1512.03385, 2015.

Hinton, G. E. and Zemel, R.: Autoencoders, minimum description length and Helmholtz free energy, Advances in neural information processing systems, 6, 1993.

Hirsch, R. M., Slack, J. R., and Smith, R. A.: Techniques of trend analysis for monthly water quality data, Water resources research, 18, 107–121, https://doi.org/10.1029/WR018i001p00107, publisher: Wiley Online Library, 1982.

640 Johannessen, O. M., Kuzmina, S. I., Bobylev, L. P., and Miles, M. W.: Surface air temperature variability and trends in the Arctic: new amplification assessment and regionalisation, Tellus A: Dynamic Meteorology and Oceanography, 68, 28 234, https://doi.org/10.3402/tellusa.v68.28234, publisher: Taylor & Francis, 2016.

Kendall, M. G.: Rank Correlation Methods, C. Griffin, https://books.google.no/books?id=hiBMAAAAMAAJ, 1948.

King, M. D., Kaufman, Y. J., Menzel, W. P., and Tanre, D.: Remote sensing of cloud, aerosol, and water vapor properties
645 from the moderate resolution imaging spectrometer(MODIS), IEEE transactions on geoscience and remote sensing, 30, 2–27, https://doi.org/10.1109/36.124212, publisher: Citeseer, 1992.

Kolstad, E. W. and Bracegirdle, T. J.: Marine cold-air outbreaks in the future: an assessment of IPCC AR4 model results for the Northern Hemisphere, Climate dynamics, 30, 871–885, https://doi.org/10.1007/s00382-007-0331-0, publisher: Springer, 2008.

Kolstad, E. W., Bracegirdle, T. J., and Seierstad, I. A.: Marine cold-air outbreaks in the North Atlantic: Temporal distribution and associations with large-scale atmospheric circulation, Climate dynamics, 33, 187–197, https://doi.org/10.1007/s00382-008-0431-5, publisher: Springer, 2009.

Kurihana, T., Foster, I., Willett, R., Jenkins, S., Koenig, K., Werman, R., Lourenco, R. B., Neo, C., and Moyer, E.: Cloud classification with unsupervised deep learning, arXiv preprint arXiv:2209.15585, https://doi.org/10.48550/arxiv.2209.15585, 2022.

Landgren, O. A., Seierstad, I. A., and Iversen, T.: Projected future changes in marine cold-air outbreaks associated with polar lows in the northern North-Atlantic Ocean, Climate Dynamics, 53, 2573–2585, https://doi.org/10.1007/s00382-019-04642-2, publisher: Springer, 2019.

Maas, A. L., Hannun, A. Y., and Ng, A. Y.: Rectifier Nonlinearities Improve Neural Network Acoustic Models, International Conference on Machine Learning, 30, 2013.

Mann, H. B.: Nonparametric tests against trend, Econometrica: Journal of the econometric society, pp. 245–259, https://doi.org/10.2307/1907187, publisher: JSTOR, 1945.

McCoy, I. L., Wood, R., and Fletcher, J. K.: Identifying meteorological controls on open and closed mesoscale cellular convection associated with marine cold air outbreaks, Journal of Geophysical Research: Atmospheres, 122, 11–678, https://doi.org/10.1002/2017JD027031, publisher: Wiley Online Library, 2017.

MODIS Characterization Support Team (MCST): MODIS 1km Calibrated Radiances Product, https://doi.org/http://dx.doi.org/10.5067/MODIS/MOD021KM.061, 2017.

Murray-Watson, R. J. and Gryspeerdt, E.: Air mass history linked to the development of Arctic mixed-phase clouds, Atmospheric Chemistry and Physics, 24, 11 115–11 132, https://doi.org/10.5194/acp-24-11115-2024, publisher: Copernicus GmbH, 2024.

Murray-Watson, R. J., Gryspeerdt, E., and Goren, T.: Investigating the development of clouds within marine cold-air outbreaks, Atmospheric Chemistry and Physics, 23, 9365–9383, https://doi.org/10.5194/acp-23-9365-2023, publisher: Copernicus Publications Göttingen, Germany, 2023.

Papritz, L. and Spengler, T.: A Lagrangian climatology of wintertime cold air outbreaks in the Irminger and Nordic Seas and their role in shaping air–sea heat fluxes, Journal of Climate, 30, 2717–2737, https://doi.org/10.1175/JCLI-D-16-0605.1, publisher: American Meteorological Society, 2017.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., and Laaksonen, A.: The Arctic has warmed nearly four times faster than the globe since 1979, Communications earth & environment, 3, 168, https://doi.org/10.1038/s43247-022-00498-3, publisher: Nature Publishing Group UK London, 2022.

Sandu, I. and Stevens, B.: On the factors modulating the stratocumulus to cumulus transitions, Journal of the Atmospheric Sciences, 68, 1865–1881, https://doi.org/10.1175/2011JAS3614.1, publisher: American Meteorological Society, 2011.

Sen, P. K.: Estimates of the regression coefficient based on Kendall's tau, Journal of the American statistical association, 63, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934, publisher: Taylor & Francis, 1968.

Serreze, M. C. and Barry, R. G.: Processes and impacts of Arctic amplification: A research synthesis, Global and planetary change, 77, 85–96, https://doi.org/10.1016/j.gloplacha.2011.03.004, publisher: Elsevier, 2011.

Shupe, M. D. and Intrieri, J. M.: Cloud radiative forcing of the Arctic surface: The influence of cloud properties, surface albedo, and solar zenith angle, Journal of climate, 17, 616–628, https://doi.org/10.1175/1520-0442(2004)017<0616:CRFOTA>2.0.CO;2, 2004.

685   Stevens, B., Cotton, W. R., Feingold, G., and Moeng, C.-H.: Large-eddy simulations of strongly precipitating, shallow, stratocumulus-topped boundary layers, Journal of the atmospheric sciences, 55, 3616–3638, https://doi.org/10.1175/1520-0469(1998)055<3616:LESOSP>2.0.CO;2, 1998.

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., L'Ecuyer, T. S., Naumann, A. K., Schulz, H., Siebesma, P. A., and others: Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds, Quarterly Journal of the Royal Meteorological

690   Society, 146, 141–152, https://doi.org/10.1002/qj.3662, publisher: Wiley Online Library, 2020.

Theil, H.: A rank-invariant method of linear and polynomial regression analysis, Indagationes mathematicae, 12, 173, 1950.

Tornow, F., Ackerman, A. S., and Fridlind, A. M.: Preconditioning of overcast-to-broken cloud transitions by riming in marine cold air outbreaks, Atmospheric Chemistry and Physics, 21, 12 049–12 067, https://doi.org/10.5194/acp-21-12049-2021, publisher: Copernicus Publications Göttingen, Germany, 2021.

695   von der Lippe, F. S., Carlsen, T., Storelvmo, T., and David, R. O.: Shortening of the Arctic cold air outbreak season detected by a phenomeno-logical machine learning approach, EGUsphere, pp. 1–29, https://doi.org/10.5194/egusphere-2025-3711, publisher: Copernicus GmbH, 2025.

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., and Cooper, J. E.: Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment, https://doi.org/10.1175/1520-0477(1996)077<0853:catere>2.0.co;2, section: Bulletin

700   of the American Meteorological Society, 1996.

Wilks, D. S.: "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinter-preted, and What to Do about It, https://doi.org/10.1175/BAMS-D-15-00267.1, section: Bulletin of the American Meteorological Society, 2016.

Wood, R. and Hartmann, D. L.: Spatial variability of liquid water path in marine low cloud: The importance of mesoscale cellular convection,

705   Journal of Climate, 19, 1748–1764, https://doi.org/10.1175/JCLI3702.1, 2006.

Wu, P. and Ovchinnikov, M.: Cloud morphology evolution in Arctic cold-air outbreak: Two cases during COMBLE period, Journal of Geophysical Research: Atmospheres, 127, e2021JD035 966, https://doi.org/10.1029/2021JD035966, publisher: Wiley Online Library, 2022.

Yamaguchi, T., Feingold, G., and Kazil, J.: Stratocumulus to cumulus transition by drizzle, Journal of Advances in Modeling Earth Systems,

710   9, 2333–2349, https://doi.org/10.1002/2017MS001104, publisher: Wiley Online Library, 2017.

Yue, S., Pilon, P., Phinney, B., and Cavadias, G.: The influence of autocorrelation on the ability to detect trend in hydrological series, Hydrological processes, 16, 1807–1829, https://doi.org/10.1002/hyp.1095, publisher: Wiley Online Library, 2002.