Characterizing Mesoscale Cellular Convection in Marine Cold Air Outbreaks with a Machine Learning Approach

Christian Philipp Lackner¹, Bart Geerts¹, Timothy W Juliano², Branko Kosovic³, and Lulin Xue³

¹University of Wyoming
²National Center for Atmospheric Research
³National Center for Atmospheric Research (UCAR)

May 28, 2024

Abstract

During marine cold-air outbreaks (MCAOs), when cold polar air moves over warmer ocean, a well-recognized cloud pattern develops, with open or closed mesoscale cellular convection (MCC) at larger fetch over open water. The Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE) provided a comprehensive set of ground-based in-situ and remote sensing observations of MCAOs at a coastal location in northern Norway. MCAO periods that unambiguously exhibit open or closed MCC are determined. Individual cells observed with a profiling Ka-band radar are identified using a watershed segmentation method. Using self-organizing maps (SOMs), these cells are then objectively classified based on the variability in their vertical structure. The SOM nodes contain some information about the location of the cell transect relative to the center of the MCC. This adds classification noise, requiring numerous cell transects to isolate cell dynamical information. The SOM-based classification shows that comparatively intense convection occurs only in open MCC. This convection undergoes an apparent lifecycle. Developing cells are associated with stronger updrafts, large spectrum width, larger amounts of liquid water, lower surface precipitation rates, and lower cloud tops than mature and weakening cells. The weakening of these cells is associated with the development of precipitation-induced cold pools. The SOM classification also reveals less intense convection, with a similar lifecycle. More stratiform vertical cloud structures with weak vertical motions are common during closed MCC periods and are separated into precipitating and non-precipitating stratiform cores. Convection is observed only occasionally in the closed MCC environment.
Characterizing Mesoscale Cellular Convection in Marine Cold Air Outbreaks with a Machine Learning Approach

Christian. P. Lackner\textsuperscript{1}, Bart Geerts\textsuperscript{1}, Timothy W. Juliano\textsuperscript{2}, Branko Kosovic\textsuperscript{2} and Lulin Xue\textsuperscript{2}

\textsuperscript{1}Department of Atmospheric Science, University of Wyoming, Laramie, WY, USA. \\
\textsuperscript{2}Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO, USA.

Corresponding author: Christian P. Lackner (clackner@uwyo.edu)

Key Points:

- Cloud cells in marine cold-air outbreaks are objectively identified and classified using a profiling mm-wave radar
- The classification reveals that open-cellular clouds undergo a convective lifecycle with distinct characteristics at each lifecycle stage
- Closed-cellular clouds contain occasional convection but generally have more stratiform characteristics
Abstract

During marine cold-air outbreaks (MCAOs), when cold polar air moves over warmer ocean, a well-recognized cloud pattern develops, with open or closed mesoscale cellular convection (MCC) at larger fetch over open water. The Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE) provided a comprehensive set of ground-based in-situ and remote sensing observations of MCAOs at a coastal location in northern Norway. MCAO periods that unambiguously exhibit open or closed MCC are determined. Individual cells observed with a profiling Ka-band radar are identified using a watershed segmentation method. Using self-organizing maps (SOMs), these cells are then objectively classified based on the variability in their vertical structure. The SOM nodes contain some information about the location of the cell transect relative to the center of the MCC. This adds classification noise, requiring numerous cell transects to isolate cell dynamical information. The SOM-based classification shows that comparatively intense convection occurs only in open MCC. This convection undergoes an apparent lifecycle. Developing cells are associated with stronger updrafts, large spectrum width, larger amounts of liquid water, lower surface precipitation rates, and lower cloud tops than mature and weakening cells. The weakening of these cells is associated with the development of precipitation-induced cold pools. The SOM classification also reveals less intense convection, with a similar lifecycle. More stratiform vertical cloud structures with weak vertical motions are common during closed MCC periods and are separated into precipitating and non-precipitating stratiform cores. Convection is observed only occasionally in the closed MCC environment.

Plain Language Summary

During marine cold-air outbreaks (MCAOs) a characteristic cloud field develops over the open ocean. At large fetch, this cloud field is characterized by open and closed mesoscale cellular convection (MCC). The Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE) in northern Norway provided comprehensive observations of mixed-phase MCC using ground-based and remote sensing instruments. Distinct open or closed MCC periods are identified using a vertically pointing cloud radar. Within these periods individual cells are identified. Using a machine learning algorithm, these cells are objectively classified based on their vertical structure as observed by the radar. The classification reveals that intense convection primarily occurs in open MCC, displaying a lifecycle with developing cells characterized by strong updrafts, substantial liquid water, lower precipitation rates, and lower cloud tops compared to mature and weakening cells. Weakening cells are associated with precipitation-induced cold pools. Less intense convection with a similar lifecycle is observed during open and closed MCC. Most frequently during closed MCC, stratiform cells with weak vertical motions are observed, some of them without precipitation reaching the surface. The vertical cloud structure of the precipitating stratiform cells is very similar to weakening convective cells.

1 Introduction

The Arctic is an integral part of the global climate system. The Arctic has been warming at a rate 2-4 times faster than the rest of the globe (e.g., Serreze & Francis, 2006; Serreze et al., 2009; Rantanen et al., 2022). Models predict this trend to continue in the future (Davy & Outten, 2020). This well-documented trend is referred to as Arctic amplification. This amplification is attributed to a variety of feedback processes. Among them are Arctic cloud feedbacks (e.g., Goosse

et al., 2018; Taylor et al., 2013), for which the magnitude and sign of the feedback remains uncertain (IPCC, 2021, p. 974).

Here, we examine one specific high-latitude cloud regime, one that occurs in marine cold-air outbreaks (MCAOs). Such synoptic events produce unique cloud patterns that through surface heat exchange, vertical mixing and radiative effects may play an important role in high-latitude cloud feedbacks (Fletcher et al., 2016). When Arctic air masses move from the sea ice (or boreal continents) over the open ocean, large surface heat fluxes develop that destabilize the boundary layer (Pithan et al., 2018). This, combined with intense surface moisture fluxes, leads to shallow moist convection that deepens with fetch. Initially, as the air moves over the ocean, linear cloud patterns form, referred to as cloud streets, and at a larger fetch downstream, the cloud streets transition to open or closed mesoscale cellular convection (MCC) (e.g., Atkinson & Zhang, 1996; Brümmer, 1999; Brümmer & Pohlmann, 2000). At high latitudes, these clouds are generally mixed-phase (e.g., Geerts et al. 2022).

Understanding the differences in the characteristics of open and closed MCC is important since their differing cloud morphology and microphysical properties impact the radiative properties of clouds as well as surface fluxes (e.g., Agee, 1987). Open and closed MCC have been studied extensively in the subtropics and the mid-latitudes (e.g., Eastman et al., 2021, 2022; Jensen et al., 2021; McCoy et al., 2017, 2023; Mohrmann et al., 2021; Muhlbauer et al., 2014; Wood, 2012; Wood & Hartmann, 2006). These studies show that a variety of environmental factors, such as wind, precipitation, surface forcing, and aerosol can influence the development and evolution of open or closed MCC. McCoy et al. (2017) find that the occurrence of open and closed MCC during MCAOs is correlated with the MCAO-index $M$, a measure of low-level static stability driven by the sea surface ($M = \theta_{SST} - \theta_{500hPa}$). Mesoscale cloud morphology is strongly influenced by surface heat fluxes and low-level stability: environments with large positive $M$ values favor open MCC regimes, while under low $M$ values, a closed MCC regime prevails, with more stratiform clouds (McCoy et al., 2017). This has potential implications for the cloud feedbacks in a globally warming climate since the intensity of MCAOs is predicted to weaken in the future (Kolstad & Bracegirdle, 2008; Landgren et al., 2019).

A comprehensive dataset of ground based observations of MCAO clouds was collected as part of the Cold-Air Outbreaks in the Marine Boundary Layer Experiment in the 2019/20 cold season (COMBLE, Geerts et al., 2022). Geerts et al. (2022), Mages et al. (2023) and Lackner et al. (2023a) illustrate MCAOs with episodes of rather deep, intense open-cellular convection observed during COMBLE. These examples demonstrate a relationship between vertical velocity, supercooled liquid water path (LWP), and surface temperature anomalies: cells with strong updrafts tend to have large LWP values, while others have weak vertical motions, low LWP values, and a surface cold pool. These differences might be attributable to different cloud lifecycle stages. Geerts et al. (2022) and Lackner et al. (2023a) also illustrate closed MCC observed during COMBLE. The closed MCC periods are characterized by a lower $M$ value and weaker winds, and clouds exhibit continuous cloud cover, weaker vertical motions, and lower average cloud top heights when compared to the open MCC. The objective of this study is to comprehensively characterize the vertical structure of MCC using the full spectrum of MCAOs observed during the 6-month COMBLE campaign, to better understand the linked MCAO cloud dynamics and microphysics, i.e. the linkages between cell vertical structure, its lifecycle, and its mesoscale organization.
This study is structured as follows. Section 2 describes the COMBLE campaign, the
periods with open and closed MCC, and the methods used to identify and classify individual cells.
The results of the classification of cells and an analysis of additional measurements is presented in
Section 3. A discussion of the results providing an interpretation of the cell classifications is in
Section 4 and the study is summarized in Section 5.

2 Data Sources and Analysis Methods

2.1 The COMBLE Campaign

The data used in this study were collected as part of COMBLE, which deployed the
Atmospheric Radiation Measurement (ARM) mobile facility 1 (AMF1, Miller et al., 2016) to
Nordmela harbor (69.141 °N; 15.684 °E) on the Norwegian island Andoya from December 2019
to May 2020. A detailed summary of the COMBLE campaign can be found in Geerts et al. (2022),
and a description of the datasets can be found in Lackner et al. (2023a). The key instrument used
for analysis is the Ka-band ARM zenith radar (KAZR), providing radar observations of vertical
cloud structures. Additionally, the THERMOCLDPHASE (Van Weverberg et al., 2023) product is
utilized. All datasets are summarized in Table 1.

The THERMOCLDPHASE product provides a vertically resolved classification of cloud
phase. This product uses KAZR and micropulse lidar data, and is based on the algorithm developed
in Shupe (2007). The micropulse lidar at the AMF1 site was only deployed starting 11 February
2020. Thus, THERMOCLDPHASE is only available for some of the periods described in Section
2.2. In addition to the COMBLE datasets, C-band scanning radar data from the Norwegian
Meteorological Institute (MET Norway) are used to analyze the horizontal structure of clouds. The
specific product that is used is a multi-radar composite of equivalent radar reflectivity factor on a
constant altitude plan position indicator (CAPPI) display at 1 km above mean sea level (AMSL).
The spatial (temporal) resolution is 1x1 km (5 min).

| Table 1. COMBLE AMF1 Data Sets Used in This Study |
| ARM Data Products | Description | Variable(s) | Units | Uncertainty |
| MET (Kyrouac et al., 2019; Ritsche & Prell, 2011) | Surface meteorological instrumentation | Atmospheric temperature Horizontal wind direction Horizontal wind speed Atmospheric pressure Precipitation rate | °C ° m s⁻¹ hPa mm hr⁻¹ | 0.2 3 0.15 0.05 |
| MAWS (Keeler et al., 2019) | Vaisala automatic weather station | Atmospheric temperature Horizontal wind direction Horizontal wind speed Atmospheric pressure | °C ° m s⁻¹ hPa | 0.25 2 max(0.1; 2 %) 0.25 |
| ARSLKAZRIKOLLIAS (Clothiaux et al., 2001; Johnson & Jensen, 2019) | KAZRARSCL: multiple outputs from first Kollias algorithm | Equivalent reflectivity factor Mean Doppler velocity Spectrum width | dBZ m s⁻¹ m s⁻¹ | 5 0.1 0.05 |
| ARSLKAZRBNDD1KOLLIAS (Johnson et al., 2019) | Cloud boundaries retrieved from KAZRARSCL | Cloud top height Cloud base height | m m | 30 30 |
### Microwave Radiometer Retrievals

<table>
<thead>
<tr>
<th>MWRRETILILCLOU</th>
<th>Microwave radiometer retrievals</th>
<th>Liquid water path</th>
<th>kg m&lt;sup&gt;-2&lt;/sup&gt;</th>
<th>0.02</th>
</tr>
</thead>
</table>

### INTERPOLATEDSONDE

<table>
<thead>
<tr>
<th>INTERPOLATEDSONDE</th>
<th>Sounding data interpolated to 1 min</th>
<th>Atmospheric temperature</th>
<th>°C</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Atmospheric pressure</td>
<td>hPa</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Horizontal wind speed</td>
<td>m s&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>0.15</td>
</tr>
</tbody>
</table>

### THERMOCLODPHASE

<table>
<thead>
<tr>
<th>THERMOCLODPHASE</th>
<th>Thermodynamic cloud phase classifications</th>
<th>Classifications: clear sky, liquid, ice, mixed phase, snow, rain, drizzle, liquid + drizzle, unknown</th>
<th>---</th>
<th>---</th>
</tr>
</thead>
</table>

### 2.2 Periods with Mesoscale Cellular Convection

Periods unambiguously exhibiting open and closed MCC are determined through visual inspection. The visual inspection involves the following steps:

1. 6-hourly KAZR time height transects during, and shortly before and after the MCAO events at the AMF1 site, as defined in Geerts et al. (2022) and Lackner et al. (2023a), are created (see Figure 1).
2. Open cellular periods are identified, when broken cloud cover, convective updrafts and numerous distinct cells of different sizes and depths are observed (Figure 1a-c).
3. Closed cellular periods are identified, when a shallow continuous cloud cover with light precipitation is observed which may not reach the surface at times (Figure 1e-f).
4. Very short periods of MCC (< 3 hours), periods with disorganized clouds, periods with polar lows (Lackner et al., 2023b), and periods with MCC connected to mid/upper-tropospheric clouds are not considered in the analysis.
5. The KAZR based identification (open, closed or neither MCC) is compared to the horizontal cloud structure in satellite imagery (see Supporting Information).

An example of the distinct vertical structure of open (closed) MCC in the KAZR data is shown in Figure 1a-c (1e-g). The closed-cell example appears to be mostly stratiform (Figure 1e). Consistent with the literature, we still refer to these cloud patches as mesoscale cellular convection (MCC), even though the convective origin may not be apparent.
Figure 1. KAZR time-height transect for an open MCC period (a-d) and a closed MCC period (e-h) showing (a,e) KAZR reflectivity factor and INTERPOLATEDSONDE temperature contours, (b,f) Doppler velocity, (c,g) spectrum width, and (d,h) objectively identified cells via the watershed segmentation method (see Section 2.3). Note that the 5 colors in (d,h) merely are recycled and do not indicate any cell grouping. The numbers on top of each cell indicate the best-matched unit, introduced in Section 2.4.
Table 2 summarizes all MCC periods used in this study. A total of 13 periods with a total of ~249 hours of open MCC were identified at the AMF1 site, compared to 4 periods with a total of ~57 hours of closed MCC. In total, this accounts for ~37% of the MCAO periods during COMBLE identified in Geerts et al. (2022) (which themselves cover 19% of the full 6-month field phase). A broad spectrum of environmental conditions prevailed during these 17 periods. The M value is above the COMBLE MCAO mean of 4.1 K for most open MCC periods (10 of 13), indicating that open MCC tends to occur during more intense MCAOs. Closed MCC on the other hand, tends to occur during weaker MCAOs, with the mean M (1.2 K) being much lower than for open MCC (6.0 K). This difference in intensity likely explains why open MCC is observed much more frequently than closed MCC at the AMF1 site: weaker MCAOs often lack the synoptic support to cover the large distance between the ice edge and the COMBLE observation site (~1000 km). Near-surface temperatures (winds) tend to be warmer (weaker) during closed MCC (Table 2). Mean cloud top heights (CTH) are typically higher during open MCC ranging from 2.2 – 3.3 km, compared to 1.9 – 2.4 km for closed MCC. However, it should be noted that CTHs are much more variable during individual open MCC periods and often exceed 4 km (see Figure 1).

Table 2. Periods of MCC used in this study, and mean environmental conditions during these MCC periods at the AMF1 site during COMBLE (Dec 2019 – May 2020).

<table>
<thead>
<tr>
<th>Start Time</th>
<th>End Time</th>
<th>Type</th>
<th>Time [min]</th>
<th>M [K]</th>
<th>T [°C]</th>
<th>U [m s⁻¹]</th>
<th>WD [°]</th>
<th>CTH [km]</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 Dec 2100 UTC</td>
<td>02 Dec 0330 UTC</td>
<td>Open</td>
<td>390</td>
<td>4.9</td>
<td>2.3</td>
<td>13.3</td>
<td>324</td>
<td>2.7</td>
</tr>
<tr>
<td>02 Dec 0700 UTC</td>
<td>02 Dec 1530 UTC</td>
<td>Open</td>
<td>510</td>
<td>4.1</td>
<td>2.2</td>
<td>12.1</td>
<td>323</td>
<td>2.5</td>
</tr>
<tr>
<td>31 Dec 0000 UTC</td>
<td>31 Dec 1915 UTC</td>
<td>Open</td>
<td>1155</td>
<td>3.7</td>
<td>2.1</td>
<td>10.5</td>
<td>286</td>
<td>2.5</td>
</tr>
<tr>
<td>04 Jan 0830 UTC</td>
<td>05 Jan 0200 UTC</td>
<td>Open</td>
<td>1050</td>
<td>6.3</td>
<td>0.3</td>
<td>10.6</td>
<td>330</td>
<td>2.6</td>
</tr>
<tr>
<td>21 Jan 2220 UTC</td>
<td>22 Jan 1830 UTC</td>
<td>Open</td>
<td>1210</td>
<td>3.1</td>
<td>2.3</td>
<td>9.2</td>
<td>328</td>
<td>2.4</td>
</tr>
<tr>
<td>02 Feb 1330 UTC</td>
<td>03 Feb 0615 UTC</td>
<td>Open</td>
<td>1005</td>
<td>5.0</td>
<td>0.8</td>
<td>10.9</td>
<td>320</td>
<td>2.2</td>
</tr>
<tr>
<td>03 Feb 1040 UTC</td>
<td>03 Feb 1845 UTC</td>
<td>Open</td>
<td>485</td>
<td>6.3</td>
<td>-1.0</td>
<td>7.3</td>
<td>308</td>
<td>2.6</td>
</tr>
<tr>
<td>04 Feb 1730 UTC</td>
<td>05 Feb 0615 UTC</td>
<td>Open</td>
<td>2775</td>
<td>7.2</td>
<td>-1.0</td>
<td>11.5</td>
<td>308</td>
<td>2.7</td>
</tr>
<tr>
<td>13 Mar 0800 UTC</td>
<td>14 Mar 0540 UTC</td>
<td>Open</td>
<td>1300</td>
<td>8.0</td>
<td>-3.0</td>
<td>9.3</td>
<td>336</td>
<td>3.3</td>
</tr>
<tr>
<td>28 Mar 0320 UTC</td>
<td>30 Mar 0945 UTC</td>
<td>Open</td>
<td>3265</td>
<td>7.1</td>
<td>-2.2</td>
<td>10.2</td>
<td>322</td>
<td>3.0</td>
</tr>
<tr>
<td>09 Apr 1420 UTC</td>
<td>10 Apr 0200 UTC</td>
<td>Open</td>
<td>700</td>
<td>4.8</td>
<td>-0.3</td>
<td>8.3</td>
<td>290</td>
<td>2.7</td>
</tr>
<tr>
<td>10 Apr 0810 UTC</td>
<td>10 Apr 1620 UTC</td>
<td>Open</td>
<td>490</td>
<td>5.0</td>
<td>0.3</td>
<td>6.8</td>
<td>304</td>
<td>2.8</td>
</tr>
<tr>
<td>10 Apr 1940 UTC</td>
<td>11 Apr 0530 UTC</td>
<td>Open</td>
<td>590</td>
<td>4.7</td>
<td>0.2</td>
<td>7.5</td>
<td>278</td>
<td>2.5</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>Open</td>
<td>14930</td>
<td>6.0</td>
<td>-0.4</td>
<td>10.1</td>
<td>314</td>
<td>2.7</td>
</tr>
<tr>
<td>17 Apr 0530 UTC</td>
<td>17 Apr 1630 UTC</td>
<td>Closed</td>
<td>660</td>
<td>1.1</td>
<td>3.3</td>
<td>8.8</td>
<td>290</td>
<td>1.9</td>
</tr>
<tr>
<td>24 Apr 1700 UTC</td>
<td>26 Apr 0010 UTC</td>
<td>Closed</td>
<td>1870</td>
<td>1.6</td>
<td>3.2</td>
<td>8.5</td>
<td>314</td>
<td>2.4</td>
</tr>
<tr>
<td>26 Apr 0300 UTC</td>
<td>26 Apr 1050 UTC</td>
<td>Closed</td>
<td>470</td>
<td>0.9</td>
<td>2.5</td>
<td>8.1</td>
<td>27</td>
<td>2.3</td>
</tr>
<tr>
<td>04 May 1050 UTC</td>
<td>04 May 1730 UTC</td>
<td>Closed</td>
<td>400</td>
<td>0.2</td>
<td>4.6</td>
<td>6.7</td>
<td>287</td>
<td>2.3</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>Closed</td>
<td>3400</td>
<td>1.2</td>
<td>3.3</td>
<td>8.3</td>
<td>317</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Note: M: MCAO-index; T: 2-m Temperature; U: 10-m wind speed; WD: 10-m wind direction; CTH: Cloud top height. The time in the summary rows (‘average’) is the total cumulative time.

2.3 Cell Identification

To examine the characteristics of open and closed MCC, individual cloud sections are identified by applying a marker-based watershed segmentation to the KAZR radar reflectivity field (Vincent & Soille, 1991). This identification is necessary for the classification of the individual
cloud sections described in Section 2.4. It should be noted that the identified cloud sections are
transects through specific parts of clouds and not necessarily representative of the three-
dimensional structure of the transected cloud. This problem is addressed by also applying the
watershed segmentation to the MET Norway C-band scanning radar data, to analyze the transect
locations. For simplicity, these cloud transects will be referred to as cells for the rest of the study.

Watershed segmentation is a segmentation technique with its origins in image processing
that has also found use in analyzing atmospheric data (e.g., Martini et al., 2014; Wu &
Ovchinnikov, 2022). In the case of a marker-based watershed segmentation, all points in a dataset
that are assigned to a predetermined marker comprise an identified object. Similar to how
geographical locations are part of a water drainage basin (watershed), data points are assigned to
the marker to which they have the steepest downward gradient.

For this study, the predetermined markers are local maxima in the smoothed KAZR
reflectivity field. A Gaussian filter with a standard deviation of 1 is used for the smoothing. The
smoothing and each following step use linear units of reflectivity (Z) instead of logarithmic units
(dBZ). The local maxima are determined with two requirements: their value must exceed 2.5 dBZ
(converted to Z) and they must be at least 80 KAZR profiles away in time from a larger value (320
seconds at 4 second time resolution). The corresponding minimum width depends on the advecting
wind speed, e.g. 3.2 km for a 10 m s^{-1} layer-mean wind. Other requirements and thresholds were
tested, but this setup was found to achieve the least amount of over- and under-segmentation of
cells in a visual analysis. However, none of the tested setups were able to completely avoid
unrealistic segmentation outcomes. Clear segmentation errors are corrected by manually removing
(identifying) specific maxima associated with over (under) segmented cells. After the local
maxima are determined, the watershed segmentation is conducted on the inverted reflectivity field,
since the algorithm determines objects based on downward gradients. Cells are confined to a -30
dBZ reflectivity threshold, which exceeds the KAZR sensitivity even at a largest relevant range
(~5 km). Each object identified by the watershed segmentation is an individual cell. Lastly, small
cells (< 45 KAZR profiles = 3 minutes) and cells with shallow cloud depth (< 50 KAZR vertical
levels between the lowest and highest echo, or 1500 m) are ignored since they have too few data
points for the method used to classify cells in Section 2.4. Furthermore, cells with high radar echo
bases (> 1500 m above sea level) are ignored since these were typically not MCC but other types
of clouds. In total, 889 cells are identified during the 17 periods, 722 (167) of which are during the
open (closed) MCC periods, i.e. 81%. Examples of identified cells in a vertical transect with open
MCC and closed MCC are shown in Figure 1d, h. Note that continuous cloud structures can be
identified as multiple cells, due to the often multi-cellular structure of what might appear as single
cloud in satellite imagery. Specifically, in the case of the closed MCC periods, the cells should not
be seen as distinct clouds but rather as a continuous cloud field that is made up of separate
precipitation cores.

The corresponding horizontal structure of the vertical cell transects is identified in the MET
Norway C-band CAPPI data with a similar watershed segmentation approach: local maxima must
exceed at least 2.5 dBZ and must be separated by at least 3 grid points (3 km). Cells in C-band are
confined to a -10 dBZ threshold instead of -30 dBZ, since thresholds below -10 dBZ were found
to connect multiple cells into single large objects. Furthermore, no smoothing is applied since the
1 km resolution data is already sufficiently smooth. This horizontal identification is only conducted
in a small area (75x75 km) around the AMF1 site sufficiently large to identify the cell located over the site. Due to missing C-band data mainly during the 28-30 March 2020 period (see Table 2), a corresponding horizontal structure was identified for only 697 vertical cell transects. For an example of an identified cell, see Figure S3.

2.4 Cell Classification

The next step is to classify the identified cells based on their vertical structure of KAZR reflectivity, Doppler velocity (i.e., hydrometeor vertical motion), and spectrum width. Since the melting level is usually at or very close to sea level, the radar echoes are dominated by snow. Spectrum broadening in snow is generally largely due to atmospheric turbulence (Doviak & Zrnic, 1993). For instance, the red-colored cell around 11:20 UTC (Figure 1a-d) is marked by strong convective updrafts and high spectrum width, due to high buoyantly-generated turbulent kinetic energy. The four closed MCC periods all occurred in late April – early May with melting levels a few hundreds of meters above the surface, which impacts the profiles of all three radar moments.

For the purpose of classifying the 889 cells, we ignore the horizontal cloud structure (i.e., the time dimension of KAZR profiles, shown in Figure 1), since it is highly dependent on deep-layer wind shear in the direction of cell motion. Instead, the vertical distribution of the three radar moments, i.e. the contoured frequency-by-altitude diagrams (CFADs) of reflectivity, Doppler velocity, and spectrum width, provide a more standardized representation of the individual cells.

An unsupervised machine learning algorithm referred to as the self-organizing map (SOM) is utilized for this CFAD-based classification (Kohonen, 1982). The main objective of the SOM is to represent archetypical patterns in datasets using an interconnected node structure, with the user defining the structure arrangement. The basis of the SOM algorithm is the following equation, which updates the SOM nodes \( m_{i,j} \) in each training step \( t \):

\[
 m_{i,j} (t + 1) = m_{i,j} (t) + a(t) \times n(t) \times [x(t) - m_{i,j} (t)] 
\]

(1)

The subscripts \( i \) and \( j \) denote the location of the node in the grid and \( x(t) \) is the input vector. The learning rate \( a(t) \) and neighborhood function \( n(t) \) determine how much a node is updated in a training step. The learning rate \( a(t) \) at each training step is the following function:

\[
a(t) = \frac{a(t = 0)}{1 + \frac{5t}{t_n}}
\]

(2)

The total number of training steps is denoted by \( t_n \), here selected as 10,000. The neighborhood function is a Gaussian distribution (provided by the minisom package for python) with a standard deviation \( \sigma(t) \), which follows the same decrease at each training step as Equation 2. At the first training step, \( a(t = 0) = 0.2 \) and \( \sigma(t = 0) = 3 \). The center of \( n(t) \) is always located on the node that has the smallest Euclidean distance to the input vector \( x(t) \). This node is referred to as the best match unit (BMU). Thus, values of \( n(t) \) decrease the further a node is away from the BMU, leading to the arrangement of the SOM nodes by similarity. The decrease of \( \sigma(t) \) with each training step continually reduces the value of \( n(t) \) at nodes other than the BMU, where \( n(t) \) always equals 1. The selected values for the number of training steps (Figure S4), the learning rate
and neighborhood function minimize the quantization error (QE, Kohonen, 2001) and topographic error (TE, Kiviluoto, 1996), compared to other tested values. Furthermore, the number of cells each node is the BMU for, does not have major outliers (Figure 2) indicating good performance of the SOM and the topography of the SOM is well organized (Figure S5).

The input vector \( \mathbf{x}(t) \) contains the data that is analyzed by the SOM algorithm, in this case the CFADs of all three KAZR moments, for a specific cell. The CFADs’ vertical structures are normalized by CTH (a cell’s highest echo top), to focus the algorithm’s outcome on the distribution of the radar variables within cloud, and not the variability of CTHs (which otherwise would dominate the SOM nodes). These CFADs (with 32x32 bins each) provide a standardized and statistically comparable representation of each cell. Furthermore, the CFADs are limited to specific ranges of the radar variables to better highlight the differences between nodes (see Figures 2 – 4). This excludes a very small fraction of available radar pixels (< 0.6 %), but does not impact the results. Note that while cell width and CTH do not define the SOM nodes, they will be examined as we characterize each node.

Each training step of the SOM consists of calculating Equation 1 for all 889 cells at each node. In each training step, cells are selected in a random order. For this study, a grid of 3x3 nodes with a rectangular topology is selected. Larger node setups with grids of 4x4 and 5x5 nodes were tested as well. While these setups have a lower QE and a similar TE compared to the 3x3 grid, the additional information contained in the additional nodes is small and outweighed by the physical interpretability of the relatively few nodes of the 3x3 grid (see Supporting Information). The final SOM resulting from the training process is referred to as the master SOM, which classifies the cells into nine categories. The characteristics of each node of the master SOM are described, first in terms of its composite radar moment vertical structure (e.g., reflectivity, shown in Figure 2) (Section 3.1), then in terms of cloud morphological parameters (Section 3.2), and finally in terms of cloud microphysical and dynamical parameters (Section 3.3). Through this analysis, insights into the dynamical and microphysical characteristics of each node are gained and in Section 4 possible physical meanings are discussed for each node. Note that the SOM technique treats each cell independently. It ignores the temporal succession of cells or cell nodes as illustrated in Figure 1d, h, and thus any interaction that may occur between cells.
3  Results

3.1  Cloud Vertical Structure

In this section, the individual nodes of the master SOM are used to describe the vertical cloud structure of MCC observed during the identified periods (Table 2). Figures 2 – 4 display the reflectivity, Doppler velocity (hydrometeor vertical motions) and spectrum width parts of the master SOM. For a better comparison of the mean reflectivity profiles in Figure 2, they are displayed together in Figure S8.
Figure 3. As Figure 2 but for Doppler velocity (hydrometeor vertical motion). The white vertical dashed line indicates 0 m s\(^{-1}\) Doppler velocity. The percentages on top indicate the percentage of positive Doppler velocities in each CFAD.

The results of the SOM highlight distinct differences between different types of cells. Starting with the reflectivity part of the master SOM (Figure 2), the mean reflectivity profile has the largest values in the left column (Figure 2a,d,g) and the lowest values in the right column (Figure 2c,f,i). Node [2,1] has the largest mean reflectivity maximum (~18 dBZ) and node [1,3] the smallest (~4 dBZ). Despite having similarly large mean reflectivity values, the most frequent reflectivity values in node [1,1] (Figure 2a) are lower compared to the two other nodes in the same column. The two lower nodes in the left column and also node [3,2] (Figure 2h) have a structure indicative of precipitation with relatively large reflectivity values (> 10 dBZ) being frequent in the lower half of the cloud. However, for node [3,2] reflectivity values are overall lower, and the most frequent reflectivity values are between 0 – 10 dBZ in the upper half of the cells. The right column is characterized by frequent low reflectivity values (< 0 dBZ) in the upper half of the cells and fewer echoes near the surface, indicating the presence of non-precipitating cell parts. Especially, for nodes [1,3] and [2,3] the frequency of reflectivity values above 0 dBZ is very low close to the surface. Compared to the right column, nodes [1,2] and [2,2] (Figure 2b,e) have larger mean...
reflectivity values and the highest frequencies are closer to the middle of the cells. Nevertheless, these two nodes have weak or no surface precipitation as indicated by the low frequency of higher reflectivity values close to the surface.

Continuing with the Doppler velocity part of the master SOM, a distinct decrease in the strength of the hydrometeor vertical motions is evident from the top left to the bottom right (Figure 3). The comparatively strong vertical motions in nodes [1,1] and [2,1] (Figure 3a,d) are indicative of strong convection. By definition, convection is characterized by ascending hydrometeors during part of the cloud lifecycle or in a cloud region (Houze, 2014). While the overall fraction of upward hydrometeor motions is larger in node [2,1], stronger updrafts (> 1 m s\(^{-1}\)) are more frequent in node [1,1]. Hydrometeor vertical motions are somewhat weaker in nodes [1,2], [2,2] and [3,1] (Figure 3b,e,g), still indicative of convection, but overall the convection is weaker compared to the two nodes with strong convection. Even weaker are the hydrometeor vertical motions in node [3,2] and the right column (Figure 3c,f,h,i). If convection is at all present within these cells it is either very weak or limited to a small part of the cell. Notable is node [1,3] which has a strong increase in mean Doppler velocities near cloud top, possibly indicative of cloud top processes.

**Figure 4.** As Figure 2 but for spectrum width.
Lastly, the spectrum width part of the master SOM (Figure 4) shows a very similar pattern compared to the Doppler velocity: the largest spectrum width values are found in the top left and the lowest in the bottom right. Again, this is a result of the strength and presence of convective motions, which produce larger spectrum width values through the generation of turbulence. Comparable to Doppler velocity, node [1,3] has an increase in spectrum width near cloud top.

3.2 Cloud Morphological Characteristics

**Figure 5.** Fraction of cells, which are from open (red) and closed (blue) MCC periods at a specific SOM node is displayed by the bars. The sum of the two bars in each node equals 100%. The numbers listed in each bar indicate the percentages of the total number of open (left) or closed (right) MCC cells that are in each node: the sum of the nine numbers equals 100%.
To examine the characteristics of the SOM nodes further, other observations collected during the identified cells are collectively analyzed for each node. In this section, cell morphological characteristics are investigated. First, it is analyzed what fraction of cells in each node is from either open or closed MCC periods (Figure 5). For instance, 108 of the 109 cells in node [2,1] are from open MCC periods, accounting for 99.1% of the cells in this node and 15.0% of all 722 cells identified during open MCC periods. The left column of the SOM contains almost exclusively cells from open MCC periods (Figure 5a,d,g). Over 44% of all cells during open MCC periods are within these three nodes. The other six nodes show a mix of cells from open and closed MCC periods. Notable are nodes [1,3] and [3,2] (Figure 5c,h). Together they contain more than 62% of all cells from closed MCC periods. Node [1,3] is dominated by closed MCC periods as it has the largest (smallest) fraction of closed (open) cells of all nodes, while node [3,2] has an almost equal split between open and closed cells.

Next, the horizontal width of the cells is examined. The horizontal width of the cells is determined by multiplying the temporal width of the cells at normalized height levels with the closest available INTERPOLATEDSONDE wind speed (Figure 6), i.e., it is the along-wind width only. Due to the standardized CFADs, the classification contains only height-relative information about the cloud width, e.g., upper vs lower levels. The absolute cloud width matters less, it depends in part on a choice in the watershed segmentation method. What matters more is the vertical structure of the cell width. There are substantial differences between the nodes in terms of the vertical structure of their width. Nodes [1,1] and [1,2] (Figure 6a,b) are the smallest in width on average. Additionally, node [1,1] shows a gradual decrease in width in the upper half of the cells, not seen for any other node. The nodes with the largest average widths are nodes [2,1] and [3,1] (Figure 6d,g). The right column and node [3,1] (Figure 6c,f,g,i) show a clear tendency to be wider in the upper half of the cells than closer to the surface, while for node [2,1] and the middle column (Figure 6b,d,e,h) width changes much less with height (except very close to cloud top).

Figure 7 displays CTH, cloud top temperature (CTT), and cloud base height (CBH) distributions for each node. A distribution of CTHs, CBHs and CTTs is available for each cell. Here, we use the 90th percentile CTHs (10th percentile CBHs and CTTs) of each cell to avoid cells being represented by outliers of these variables. The CTH of a cell is determined based on the cell mask from the watershed segmentation (see Figure 1d,h). The CTT is the INTERPOLATEDSONDE temperature at this height at the matching time. CBH estimates are from the ARSCLKAZRBNDIKOLLAS product, which is derived from ceilometer and lidar data. Since cells can overlap in time, cloud base heights are assigned to the cell that is the closest to the surface in any given KAZR profile. The distributions of CTH and CTT show a large spread within individual nodes (Figure 7), which can be explained by the variability of the mean CTHs of the MCC periods (see Table 2). However, the distributions also reveal differences between the nodes. Mean CTHs (CTT) reach from 1.9 km (-14°C) in node [1,3] (Figure 7c) to 3.6 km (-32°C) in node [2,1] (Figure 7g). These two nodes have other notable features compared to the rest of the nodes. Node [1,3] has very narrow distributions of CTH and CTT, while node [2,1] is the node that has more cells than any other node (~25%) reaching homogeneous freezing temperatures at cloud top [<-38°C (Ickes et al., 2015)]. The rest of the nodes mostly have mean CTHs (CTTs) between 2 – 3 km (-20 – -30°C). Mean CBH is between 0.8 – 1.1 km for most nodes, except nodes [2,1] and [3,1] for which mean CBH is ~0.5 km (Figure 7d,g). Thus, cells in these two nodes have...
on average the largest cloud depths (CTH – CBH). The CBH distribution of node [3,3] is notable for having the largest spread of all nodes (Figure 7i).

Figure 6. Width of the identified cells at each node of the master SOM. The solid line indicates the mean and the shaded area the interquartile range.
The last morphological characteristic that is analyzed is the horizontal size of the cells (area) and the location at which the KAZR transects the cells. This analysis is based on linking the vertical cell transects with horizontal cell maps (C-band radar echoes) as they pass over the AMF1 site (see Section 2.3). To analyze whether specific nodes are correlated with transects through specific parts of cells (cell cores vs. cell edges), a minimum relative distance $r_{\text{min}}$ from the cell center (geographical center of cell; see Supporting Information) is calculated for each cell. $r_{\text{min}}$ is
the closest distance of the KAZR to the cell center relative to the cell size. The calculation of $r_{min}$ is explained in the Supporting Information and is used instead of an absolute distance to account for the different sizes of cells. Note that this analysis applies to a height of 1 km since it is based on a CAPPI product at 1 km AMSL (see Section 2.3). The calculation of $r_{min}$ also assumes that cells are circular for simplicity. This simplification explains why $r_{min}$ can exceed 100%, although this is not very common (Figure 8). If cells were all circular, then the average $r_{min}$ would be 50% and the distribution uniform, given that transect locations are stochastic. The average $r_{min}$ is <50% for most nodes, simply because very marginal cell transects are too small to be identified as individual cells.

Mean cell sizes are \( \sim 80 \text{ – } 110 \text{ km}^2 \) depending on the node with smaller (larger) mean cell sizes found in the top (bottom) row. However, there is a large spread of cell sizes with the smallest cells being only a few km\(^2\) and the largest exceeding 200 km\(^2\). Based on satellite imagery Wu & Ovchinnikov (2022) report mean cell sizes \( \sim 300 \text{ km}^2 \) in the vicinity of the AMF1 site for the two March 2020 open MCC periods during COMBLE. This discrepancy is likely due to our radar-based approach that may distinguish multiple cells that may appear as a single cloud in satellite imagery (see Figure 1). Moreover, the cell sizes apply to a height of 1 km instead of cloud top.

Most nodes have a large spread of where cells are transected (Figure 8). Cell cores ($r_{min} < 25\%$) and edges ($r_{min} > 75\%$) are transected in each node. Notable are nodes [2,1], [2,2], [3,1] and [3,2], which are more related with transects of cell cores, as over 50% of cell transects are through the cell core (Figure 8d,e,g,h). Barely any transects ($< 5\%$) within these nodes only transect cell edges. On the other hand, nodes [2,3] and [3,3] are the two nodes that appear to more frequently ($\sim 25\%$) have transects that only sample cell edges (Figure 8f,i). To summarize there are three groups of nodes based on where cells are transected: the top row contains transects of all parts of cells, the left two nodes in the bottom two rows mostly contain cell transects of cell cores, and most transects only going through cell edges are found in the right nodes of the bottom two rows. These groups can also be justified by comparing the $p$-values (based on a t-test) of the $r_{min}$ distributions. Distributions that have large $p$-values ($> 0.2$) when compared to each other are in the same group.
Figure 8. Distributions of cell size and minimum relative distance of the KAZR from the cell center based on horizontal C-band scanning radar data at each node of the master SOM. The displayed cell sizes are at the time of the minimum relative distance. For a description of how the minimum relative distance is calculated, see Supporting Information. Box plots are as in Figure 7.

3.3 Cloud Microphysical and Dynamical Characteristics

In this section, microphysical and dynamical characteristics of the SOM nodes are analyzed, specifically four variables: microwave radiometer liquid water path (LWP) (Figure 9), a cloud phase classification from the THERMOCLDPHASE product (Figure 10), surface precipitation rate, and surface virtual potential temperature ($\theta_v$) anomaly (a measure of cold pool strength) (Figure 11). Since multiple cells can be identified in a single vertical column, it is not
always possible to unambiguously assign a LWP measurement (a vertically integrated quantity) to
a single cell. Therefore, LWP is only assigned to a cell if that cell makes up more than 80% of
cloud in a vertical profile and is the closest cell to the surface. Data points from the 2D
THERMOCLOD PHASE product are assigned to the cell they cover. Precipitation measurements and
the surface $\theta_v$ anomaly are assigned to the cell closest to the surface. The minimum (including
negative values) difference between the instantaneous and the six-hour mean surface $\theta_v$ (centered
on the instantaneous value) during each cell is used as a measure for cold pool strength.

The SOM nodes show strong contrasts in terms of presence of liquid water (Figure 9). The
largest LWP values are found in node [1,1] with a mean value (90th percentile) of 0.33 kg m$^{-2}$ (0.86
kg m$^{-2}$) (Figure 9a). Other nodes with substantial but lower LWP values are nodes [1,2], [1,3],
[2,1], and [3,2]. The lowest mean (90th percentile) values of LWP are in node [3,3] with 0.05 kg
m$^{-2}$ (0.12 kg m$^{-2}$) (Figure 9i) and nodes [2,3] and [3,1] have only slightly larger values. Noticeable
for almost all nodes is that the mode of the distribution is at low LWP values at or below 0.05 kg
m$^{-2}$, indicating that the presence of liquid is confined to specific parts of the cells. The exceptions
are nodes [1,3] and [3,2]. While the latter has a secondary mode at larger values (Figure 9h), node
[1,3] completely lacks the mode at low values (Figure 9c). Instead, the mode is similar to the mean
value, indicating a constant presence of liquid within cells associated with this node. These two
nodes are the same nodes that have the largest fraction of cells from closed MCC periods (Figure
5).

**Figure 9.** Distributions of liquid water path (LWP) of the identified cells at each node of the master
SOM. Box plots are as in Figure 7.
Figure 10. Fraction of retrieved cloud phase of the identified cells at each node of the master SOM. Note that R+D refers to rain and drizzle combining the rain, drizzle, and liquid + drizzle classification of the THERMOCLDPHASE product (see Table 1). Note that this product is only available for cells identified after 11 February 2020 (441 of 889).

The THERMOCLDPHASE product corroborates these findings (Figure 10). The largest fraction of mixed phase and liquid in any node is found in node [1,3] (Figure 10c), the node associated with a constant presence of liquid in cells. Furthermore, the node with the largest mean LWP, node [1,1], has the second largest fraction of combined mixed and pure liquid phase (Figure 10a). Generally, the other nodes follow this pattern of a correlation between the frequency of larger LWP path values and the fraction of mixed phase. Two notable deviations from this pattern are nodes [2,1] and [3,2]. Node [2,1] has a relatively low fraction of mixed phase (~15%, Figure 10d),
but much larger LWP values as compared to nodes with a similar mixed phase fraction. The low fraction of mixed phase is likely related to much larger ice water contents within this node than other nodes, as indicated by the large mean reflectivity values (Figure 2d). For node [3,2], the mixed phase fraction is very low (∼7 %, Figure 10h), however, there is a comparatively large fraction of rain and drizzle (∼ 6 %, Figure 10h) which is the reason for the secondary mode with larger LWP values (Figure 9h). The rain is associated with cells from closed MCC periods in this node.

Figure 11. Distribution of precipitation rate and virtual potential temperature anomaly of the identified cells at each node of the master SOM. Box plots are as in Figure 7.
Lastly, surface precipitation rates and cold pools are analyzed. Precipitation rates and (negative) $\theta_v$ anomalies increase steadily in each column from top to bottom (Figure 11). In the top row, precipitation rates are negligible and the mean surface $\theta_v$ anomaly is close to 0 K, and the distribution has few cells with larger anomalies. On average the strongest precipitation rates (0.7 mm hr$^{-1}$) and surface $\theta_v$ anomalies (-0.8 K) are found in node [3,1] (Figure 11g). The distributions of precipitation rates for all nodes have a mode at very low values (< 0.1 mm hr$^{-1}$), indicating that precipitation associated with the analyzed cells is usually caused by brief showers. Overall, the precipitation rates and the occurrence of cold anomalies correlate well. This indicates that cold pools are likely precipitation induced.

4 Discussion

4.1 Interpretation of the Cell Classification

Our analysis shows that a SOM-based cell classification can determine distinct characteristics of open and closed MCC clouds observed during MCAOs. For instance, the vertical distributions of the radar variables differ significantly between nodes. Furthermore, observations made independent from the KAZR such as LWP, surface precipitation, and surface $\theta_v$ anomalies show strong differences between the nodes. These distinct nodes reveal information about dynamical/microphysical characteristics of cell transects (the “signal”), but also about varying locations of the transects in evolving three-dimensional cloud structures. In this section, an interpretation of the individual nodes is provided concerning what possible types of cells they might describe. Cell transect location can be considered “noise” in the SOM classification, however neighboring cells (possibly in multi-cellular cloud systems) may interact, e.g., a decaying cell may trigger a new cell. Such interactions are not examined here: each cell is examined in isolation.

The left column of the SOM has two major things in common: these three nodes have the largest mean reflectivity profile values (Figure 2), and they are almost exclusively from open MCC periods (Figure 5). Thus, they describe the most intense cells observed during open MCC periods. There are major differences between these three nodes. Node [1,1] has strong convection as indicated by the vertical distributions of Doppler velocity and spectrum width (Figures 3a, 4a), but has on average shallower cloud depths, smaller cell widths and sizes than the bottom two nodes (Figures 6 – 8). Based on these characteristics, node [1,1] is interpreted as representing developing intense cells, that have strong convection, developing in depth and size. For node [2,1] convection is similarly strong (Figures 3d, 4d), but mean reflectivity values and cloud depth are the largest (Figures 2, 7). Moreover, cell widths and sizes are larger (Figures 6, 8). Thus, the interpretation of node [2,1] is that it contains mature intense cells. Lastly, compared to the mature intense cells, node [3,1] has substantially weaker convection (Figures 3g, 4g), cells are slightly shallower in depth and have an increased width in the upper half of cells. These characteristics indicate that node [3,1] generally contains weakening intense cells. To summarize, these three nodes describe a convective lifecycle of intense cells with a developing, mature and weakening stage.

The microphysical and dynamical characteristics of these nodes highlight important processes associated with this lifecycle. The presence of liquid decreases substantially as the lifecycle progresses (Figure 9, 10), while cold pools and surface precipitation rates intensify (Figure 11; also shown by the reflectivity distributions in Figure 2). Thus, the glaciation of cells at
the mature lifecycle stage, the ensuing precipitation formation, and the development of cold pools appear to play a key role in the cell lifecycle. At the developing and mature stages, CTTs can be sufficiently cold for homogenous freezing of supercooled liquid droplets for some cells, but in most cases CTTs are higher, requiring heterogenous freezing (Figure 7a, b).

There is one other difference between these nodes that needs to be mentioned. Nodes [2,1] and [3,1] tend to sample mostly cell cores, while node [1,1] does not show a clear trend in terms of cell center proximity. It is unlikely that transect location variations significantly impact the lifecycle stage interpretation: the transects of cell edges in node [1,1] are likely caused by some intense updrafts developing on the edges of pre-existing cells. If these updrafts are not sufficiently isolated, their horizontal reflectivity structure might not be separable from the cell edge on which they develop. Furthermore, a newly developing updraft might make up a whole cell such that it can be observed at the edge of that cell. Yet, the strong precipitation cores of the mature and weakening intense cells are unlikely to be located on cell edges. This raises the question concerning which nodes contain the edge transects of these lifecycle stages. This is discussed in the interpretation of nodes [2,3] and [3,3].

Node [1,3] is dominated by cells from closed MCC periods. This node is characterized by rather stratiform cloud profiles with frequent low reflectivity values near cloud top and weak vertical motions (Figures 2 – 3). Near cloud top, an increase in upward hydrometeor motions and spectrum width is found (Figures 3 – 4). This cloud top turbulence may be attributed to generating cells (e.g., Hobbs & Locatelli, 1978; Rosenow et al., 2014) driven by cloud top radiative cooling (Keeler et al., 2016). Such cells, with a reflectivity and Doppler velocity structure similar to that of generating cells at the top of frontal clouds (e.g., Plummer et al., 2015; Zaremba et al., 2022), become apparent in a zoom-in of these cells, e.g., around 10:00 UTC in Fig. 1e (not shown). These traits are indicative of stratiform clouds, with hydrometeors initiated at cloud top and slowly growing as they fall through the cloud. Based on these stratiform characteristics and the lack of surface precipitation, this node is interpreted as non-precipitating stratiform cells. Furthermore, these cells have the shallowest cloud depth and the lowest spread in CTHs (Figure 7), and it is the only node with a presence of liquid in all parts of cells (Figure 9). At the same time, there are no cold pools (Figure 11). Since these cells are rather horizontally homogeneous, the transect location is of little consequence (Figure 8). Very few cells from open MCC periods are in this node because the relatively high $M$ values, strong surface heating, and intense convection during such periods prevent the development of long-lasting stratiform clouds. The non-precipitating stratiform cells are likely the “base state” in a closed MCC environment. They dominate in the closed MCC example in Figure 1h. Yet, this cloud regime exists in an $M > 0$ environment and contains occasional convective updrafts (Figure 1f), classified in different nodes. Hence, we still refer to this as mesoscale convective updrafts (Figure 1f), classified in different nodes. Hence, we still refer to this as mesoscale cellular convection (MCC) (consistent with McCoy et al., 2017 and others), even though stratiform processes may dominate clouds and precipitation.

In terms of the vertical structures of the radar variables, cloud morphological and microphysical characteristics, the comparison between nodes [1,2] and [2,2] is very similar to the comparison between the developing and mature intense cells. The major difference is that reflectivity values are lower, the convection is less intense (Figures 2 – 4), and the cells are shallower (Figure 7). Thus, these two nodes are interpreted as developing ([1,2]) and mature ([2,2])
moderately convective cells. Unlike the intense convection, the moderate convection can occur in the continuous cloud deck of closed MCC periods (Figure 5).

Node [3,2] is comparable to [3,1], but with lower reflectivity values and even weaker convection (Figures 2 – 4). The cells are also slightly shallower (Figure 7). Thus, this node can be interpreted as containing weakening convective cells as well, which can be of an intense or moderate convective origin. In addition to these convective cells, it is notable that node [3,2] has the second most cells from closed MCC periods (Figure 5). These cells appear to have different development mechanisms than the convective cells in the same node. For instance, some cells in Figure 1h are classified in this node but are rather stratiform, while two adjacent cells observed around 13 March 2020 1040 UTC are also classified in this node and appear to be the remnants of active convection (Figure 1d). This overlap is remarkable: cells with very different development mechanisms (convective vs. stratiform) result in very similar vertical cloud structures at some point during their development. Compared to the non-precipitating stratiform cells (node [1,3]), the stratiform cells in this node produce surface precipitation (rain) (Figures 10, 11) and are dominated by in-cloud ice as opposed to liquid (Figures 9, 10). This could indicate that the formation of ice is important for the development of significant precipitation in closed MCC. To summarize, node [3,2] contains a mix of precipitating stratiform cells and weakening convective cells.

This leaves nodes [2,3] and [3,3] to be interpreted. Cell transects included in these two nodes relatively frequently occur near cell edges (Figure 8). This is interpreted as these two nodes being the only nodes that are not predominantly associated with a lifecycle stage and cell intensity, but also the relative location of the transects. These two nodes include transects near the cell center; these are interpreted as dissipating convective cells at the very end of the convective lifecycle. This interpretation is based on the rather stratiform vertical cloud structures, cells being mostly from open MCC periods, much larger cell widths near cloud top than closer to the surface, and often elevated cloud bases (Figures 2 – 7). These two nodes also contain transects near the cell edge. Such transects may be associated with any intensity or lifecycle stage. Especially when cells reach their maximum size during the mature and weakening stages, cell edges likely lack precipitation and have weak vertical motions making their vertical structure comparable to that of dissipating cells. This leads to the SOM classifying cell edges from all lifecycle stages and intensities in these two nodes instead of their corresponding node, and answers the question concerning the lack of cell edge transects in specific nodes. In other words, these cloud-edge transects add noise, revealing no information about intensity or lifecycle stage.

Lastly, it should be mentioned that the two nodes are not entirely the same: node [2,3] has slightly stronger convective motions. This might be due to the development of new updrafts in dissipating cells or edge transects through a convective updraft. The reason for the development of new updrafts in node [2,3] might be related to the lack of surface precipitation and associated cold pools (Figure 11). To summarize, nodes [2,3] and [3,3] contain a specific lifecycle stage (dissipating convective cells), but also edge transects from all lifecycle stages.

4.2 Study Limitations and Future Work
A few things need to be considered when interpreting the results of the cloud classification. While the SOM yields discrete classifications, it should be remembered that these clouds occur on a continuous spectrum in terms of their development stage, the intensity of convection, or stratiform precipitation growth. The SOM with more nodes provided in the Supporting Information show this continuous spectrum in more detail, but, as previously argued, the main information is captured by the 3x3 node SOM. Also, an individual cell defined objectively (watershed segmentation method) may consist of multiple smaller convective turrets at different life stages, e.g., the blue cell around 12:20 UTC in Figure 1a-d. The analysis provides a comparison of the horizontal and vertical structures of cells, but it is not possible to examine the temporal evolution of the three-dimensional structure of individual cells. Rather, the statistical analysis of a large sample of cells allows differentiation in terms of intensity, lifecycle stage, and transect location.

The results provide strong evidence that a convective lifecycle dominates in open MCC environments. This sets the stage for further work that examines cell interactions and mergers in MCAO convection.

The approach to objectively classify clouds helps to conceptualize the defining characteristics of convective cells during MCAOs, and to elucidate the driving dynamical-microphysical processes. While buoyant ascent (possibly forced by outflow boundary convergence) clearly is the driving mechanism during relatively intense MCAOs with strong surface heating, stratiform cloud and precipitation processes play a role as well, especially under low $M$ value conditions. This may result in a closed MCC organization, which was less common during COMBLE, and observed only late in the field campaign (April-May).

Our analysis is limited to the comparatively short data record of COMBLE. While a certain range of MCAO intensities is captured, it is not necessarily representative of the long-term climatological conditions over the high-latitude oceans, and certainly not elsewhere. The approach presented here, using the COMBLE dataset, could be used to analyze MCC clouds at observational sites with a much longer data record such as ARM's North Slope of Alaska and Eastern North Atlantic (ENA) sites. On average, $M$ values are lower in the area of the ENA site during MCAOs (or post-frontal periods) (Fletcher et al., 2016), and the amount of time during which open and closed MCC occurs is much more equal (Jensen et al., 2021). A comparison with the ENA site allows examination of the importance of the ice phase in MCAO convection dynamics: the ENA site is considerably warmer, resulting in less common mixed-phase MCC.

In future work, we will analyze the MCAOs listed in Table 2 using a large-eddy simulation (LES), starting with the 13 March 2020 case (Kosović et al., 2023), following a similar approach as presented in this study. A radar simulator such as the Cloud Resolving Model Radar Simulator (Oue et al., 2020) will be used to create synthetic radar data from the LES that is comparable to KAZR observations. A comparison of the observational SOM classification to an LES-output based SOM classification may reveal whether the model reproduces MCC with the characteristics observed in this study. Another benefit of conducting this analysis with LES is that the three-dimensional spatial structure and the temporal evolution of the variables presented in this study can be analyzed, giving further insight into the impact of locations of cell transects on the cell classification. The LES further allows analysis of a large number of cells given that the simulation covers a large domain, stretching from north of the ice edge to south of the COMBLE observation site (~1,200 km distance).
5 Conclusions

In this study, the characteristics of open and closed mesoscale cellular convection (MCC) during marine cold-air outbreaks (MCAO) are investigated. The data were collected as part of the COMBLE field campaign at a coastal site in northern Norway. Unambiguous periods of open and closed MCC are determined within MCAOs, based on radar and satellite data. Periods of open and closed MCC are associated with different MCAO intensities and display distinct cloud top height differences. Watershed segmentation and self-organizing map (SOM) algorithms are used to objectively identify and classify individual cells observed during these periods. The identification and classification of cells are based on the composite frequency-by-altitude structure of the cells as observed by a high-resolution vertically-pointing radar. Furthermore, the horizontal structure associated with the vertical cell transects is identified with a similar watershed segmentation in scanning radar data. The profiling radar’s time-height transects across advecting cells may occur anywhere relative to MCC, which often has a complex, multi-core footprint as documented by the scanning radar. Invariably, the SOM nodes contain some distinguishing information about transect location, which adds noise to a process-focused classification. A sample size of 889 cell transects is large enough for a physically meaningful cell classification to emerge from this stochastic noise.

The main findings are as follows:

- The SOM algorithm yields nine nodes with distinct vertical distributions of the three radar moments, i.e., reflectivity, Doppler velocity, and spectrum width. Furthermore, there are distinct differences between the cloud morphological characteristics of the individual nodes. Some nodes group vertical transects of convective cells in distinct lifecycles stages and at different intensities, while others describe rather stratiform clouds. Such interpretation of the nodes is corroborated by independent microphysical and dynamical observations.

- A group of three nodes is interpreted as intense convection. These cells only occur during open MCC periods and are separated into developing, mature, and weakening intense cells. Developing cells are associated with strong updrafts, high spectrum width, shallower cloud depth, pockets of large liquid water path, and minimal precipitation. As the convection matures and then weakens, updrafts decrease in strength, cloud tops are higher and colder, the amount of liquid decreases indicating that cells glaciate, surface precipitation intensifies and cold pools develop below the cells. Two additional nodes contain cell transects in the decaying stage of the convective lifecycle. New updrafts can develop below these dissipating cells.

- Cells with more moderate convection compared to the intense cells are observed during open MCC and occasionally during closed MCC. This moderate convection has a similar lifecycle.

- During closed MCC periods, cells with stratiform characteristics are observed most frequently. These stratiform cells are separated into non-precipitating and precipitating cells. Non-precipitating stratiform cells are relatively shallow with cloud tops mostly >-15°C, have weak vertical drafts, somewhat enhanced spectrum width near cloud top, and mostly mixed-phase
cloud. The precipitating stratiform cells are dominated by ice, and are structurally very similar to weakening convective cells.

Acknowledgments: The COMBLE campaign was enabled by an ARM Mobile Facility deployment proposal to the Office of Science of the U.S. Dept. of Energy (DOE). This research was supported by DOE Atmospheric System Research (ASR) Grants DE-SC0018927 and DE-SC0021151. We appreciate the assistance from several ARM instrument and data mentors, especially Scott Giangrande. The National Center for Atmospheric Research is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977.

Open Research: COMBLE campaign data at the Andenes site (ANX) used in this study are available through the references found in Table 1. C-band equivalent radar reflectivity factor is obtained from the Norwegian Meteorological Institute (2020). MODIS satellite data used in the Supporting Information is from the MYD021KM (MODIS Characterization Support Team (MCST), 2017a) and MYD03 products (MODIS Characterization Support Team (MCST), 2017b). The watershed segmentation was implemented in python code using the sci-kit image package (Van Der Walt et al., 2014) and the self-organizing map algorithm was implemented in python code with the minisom package (Vettigli, 2018).

References


Figure 1.
Figure 3.
Figure 5.
Figure 6.
Figure 7.
Figure 10.
Figure 11.
Characterizing Mesoscale Cellular Convection in Marine Cold Air Outbreaks
with a Machine Learning Approach

Christian. P. Lackner\textsuperscript{1}, Bart Geerts\textsuperscript{1}, Timothy W. Juliano\textsuperscript{2}, Branko Kosovic\textsuperscript{2} and Lulin Xue\textsuperscript{2}

\textsuperscript{1}Department of Atmospheric Science, University of Wyoming, Laramie, WY, USA.
\textsuperscript{2}Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO, USA.
Corresponding author: Christian P. Lackner (clackner@uwyo.edu)

Key Points:
- Cloud cells in marine cold-air outbreaks are objectively identified and classified using a profiling mm-wave radar
- The classification reveals that open-cellular clouds undergo a convective lifecycle with distinct characteristics at each lifecycle stage
- Closed-cellular clouds contain occasional convection but generally have more stratiform characteristics
Abstract

During marine cold-air outbreaks (MCAOs), when cold polar air moves over warmer ocean, a well-recognized cloud pattern develops, with open or closed mesoscale cellular convection (MCC) at larger fetch over open water. The Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE) provided a comprehensive set of ground-based in-situ and remote sensing observations of MCAOs at a coastal location in northern Norway. MCAO periods that unambiguously exhibit open or closed MCC are determined. Individual cells observed with a profiling Ka-band radar are identified using a watershed segmentation method. Using self-organizing maps (SOMs), these cells are then objectively classified based on the variability in their vertical structure. The SOM nodes contain some information about the location of the cell transect relative to the center of the MCC. This adds classification noise, requiring numerous cell transects to isolate cell dynamical information. The SOM-based classification shows that comparatively intense convection occurs only in open MCC. This convection undergoes an apparent lifecycle. Developing cells are associated with stronger updrafts, large spectrum width, larger amounts of liquid water, lower surface precipitation rates, and lower cloud tops than mature and weakening cells. The weakening of these cells is associated with the development of precipitation-induced cold pools. The SOM classification also reveals less intense convection, with a similar lifecycle. More stratiform vertical cloud structures with weak vertical motions are common during closed MCC periods and are separated into precipitating and non-precipitating stratiform cores. Convection is observed only occasionally in the closed MCC environment.

Plain Language Summary

During marine cold-air outbreaks (MCAOs) a characteristic cloud field develops over the open ocean. At large fetch, this cloud field is characterized by open and closed mesoscale cellular convection (MCC). The Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE) in northern Norway provided comprehensive observations of mixed-phase MCC using ground-based and remote sensing instruments. Distinct open or closed MCC periods are identified using a vertically pointing cloud radar. Within these periods individual cells are identified. Using a machine learning algorithm, these cells are objectively classified based on their vertical structure as observed by the radar. The classification reveals that intense convection primarily occurs in open MCC, displaying a lifecycle with developing cells characterized by strong updrafts, substantial liquid water, lower precipitation rates, and lower cloud tops compared to mature and weakening cells. Weakening cells are associated with precipitation-induced cold pools. Less intense convection with a similar lifecycle is observed during open and closed MCC. Most frequently during closed MCC, stratiform cells with weak vertical motions are observed, some of them without precipitation reaching the surface. The vertical cloud structure of the precipitating stratiform cells is very similar to weakening convective cells.

1 Introduction

The Arctic is an integral part of the global climate system. The Arctic has been warming at a rate 2-4 times faster than the rest of the globe (e.g., Serreze & Francis, 2006; Serreze et al., 2009; Rantanen et al., 2022). Models predict this trend to continue in the future (Davy & Outten, 2020). This well-documented trend is referred to as Arctic amplification. This amplification is attributed to a variety of feedback processes. Among them are Arctic cloud feedbacks (e.g., Goosse
et al., 2018; Taylor et al., 2013), for which the magnitude and sign of the feedback remains uncertain (IPCC, 2021, p. 974).

Here, we examine one specific high-latitude cloud regime, one that occurs in marine cold-air outbreaks (MCAOs). Such synoptic events produce unique cloud patterns that through surface heat exchange, vertical mixing and radiative effects may play an important role in high-latitude cloud feedbacks (Fletcher et al., 2016). When Arctic air masses move from the sea ice (or boreal continents) over the open ocean, large surface heat fluxes develop that destabilize the boundary layer (Pithan et al., 2018). This, combined with intense surface moisture fluxes, leads to shallow moist convection that deepens with fetch. Initially, as the air moves over the ocean, linear cloud patterns form, referred to as cloud streets, and at a larger fetch downstream, the cloud streets transition to open or closed mesoscale cellular convection (MCC) (e.g., Atkinson & Zhang, 1996; Brümmer, 1999; Brümmer & Pohlmann, 2000). At high latitudes, these clouds are generally mixed-phase (e.g., Geerts et al. 2022).

Understanding the differences in the characteristics of open and closed MCC is important since their differing cloud morphology and microphysical properties impact the radiative properties of clouds as well as surface fluxes (e.g., Agee, 1987). Open and closed MCC have been studied extensively in the subtropics and the mid-latitudes (e.g., Eastman et al., 2021, 2022; Jensen et al., 2021; McCoy et al., 2017, 2023; Mohrmann et al., 2021; Muhlbauer et al., 2014; Wood, 2012; Wood & Hartmann, 2006). These studies show that a variety of environmental factors, such as wind, precipitation, surface forcing, and aerosol can influence the development and evolution of open or closed MCC. McCoy et al. (2017) find that the occurrence of open and closed MCC during MCAOs is correlated with the MCAO-index $M$, a measure of low-level static stability driven by the sea surface ($M = \theta_{SST} - \theta_{500hPa}$). Mesoscale cloud morphology is strongly influenced by surface heat fluxes and low-level stability: environments with large positive $M$ values favor open MCC regimes, while under low $M$ values, a closed MCC regime prevails, with more stratiform clouds (McCoy et al., 2017). This has potential implications for the cloud feedbacks in a globally warming climate since the intensity of MCAOs is predicted to weaken in the future (Kolstad & Bracegirdle, 2008; Landgren et al., 2019).

A comprehensive dataset of ground based observations of MCAO clouds was collected as part of the Cold-Air Outbreaks in the Marine Boundary Layer Experiment in the 2019/20 cold season (COMBLE, Geerts et al., 2022). Geerts et al. (2022), Mages et al. (2023) and Lackner et al. (2023a) illustrate MCAOs with episodes of rather deep, intense open-cellular convection observed during COMBLE. These examples demonstrate a relationship between vertical velocity, supercooled liquid water path (LWP), and surface temperature anomalies: cells with strong updrafts tend to have large LWP values, while others have weak vertical motions, low LWP values, and a surface cold pool. These differences might be attributable to different cloud lifecycle stages. Geerts et al. (2022) and Lackner et al. (2023a) also illustrate closed MCC observed during COMBLE. The closed MCC periods are characterized by a lower $M$ value and weaker winds, and clouds exhibit continuous cloud cover, weaker vertical motions, and lower average cloud top heights when compared to the open MCC. The objective of this study is to comprehensively characterize the vertical structure of MCC using the full spectrum of MCAOs observed during the 6-month COMBLE campaign, to better understand the linked MCAO cloud dynamics and microphysics, i.e. the linkages between cell vertical structure, its lifecycle, and its mesoscale organization.
This study is structured as follows. Section 2 describes the COMBLE campaign, the periods with open and closed MCC, and the methods used to identify and classify individual cells. The results of the classification of cells and an analysis of additional measurements is presented in Section 3. A discussion of the results providing an interpretation of the cell classifications is in Section 4 and the study is summarized in Section 5.

2 Data Sources and Analysis Methods

2.1 The COMBLE Campaign

The data used in this study were collected as part of COMBLE, which deployed the Atmospheric Radiation Measurement (ARM) mobile facility 1 (AMF1, Miller et al., 2016) to Nordmela harbor (69.141 °N; 15.684 °E) on the Norwegian island Andoya from December 2019 to May 2020. A detailed summary of the COMBLE campaign can be found in Geerts et al. (2022), and a description of the datasets can be found in Lackner et al. (2023a). The key instrument used for analysis is the Ka-band ARM zenith radar (KAZR), providing radar observations of vertical cloud structures. Additionally, the THERMOCLDPHASE (Van Weverberg et al., 2023) product is utilized. All datasets are summarized in Table 1.

The THERMOCLDPHASE product provides a vertically resolved classification of cloud phase. This product uses KAZR and micropulse lidar data, and is based on the algorithm developed in Shupe (2007). The micropulse lidar at the AMF1 site was only deployed starting 11 February 2020. Thus, THERMOCLDPHASE is only available for some of the periods described in Section 2.2. In addition to the COMBLE datasets, C-band scanning radar data from the Norwegian Meteorological Institute (MET Norway) are used to analyze the horizontal structure of clouds. The specific product that is used is a multi-radar composite of equivalent radar reflectivity factor on a constant altitude plan position indicator (CAPPI) display at 1 km above mean sea level (AMSL). The spatial (temporal) resolution is 1x1 km (5 min).

### Table 1. COMBLE AMF1 Data Sets Used in This Study

<table>
<thead>
<tr>
<th>ARM Data Products</th>
<th>Description</th>
<th>Variable(s)</th>
<th>Units</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MET</strong></td>
<td>Surface meteorological instrumentation</td>
<td>Atmospheric temperature, Horizontal wind direction, Horizontal wind speed, Atmospheric pressure, Precipitation rate</td>
<td>°C, °, m s⁻¹, hPa, mm hr⁻¹</td>
<td>0.2, 3, 2%, 0.15, 0.05</td>
</tr>
<tr>
<td><strong>MAWS</strong></td>
<td>Vaisala automatic weather station</td>
<td>Atmospheric temperature, Horizontal wind direction, Horizontal wind speed, Atmospheric pressure</td>
<td>°C, °, m s⁻¹, hPa</td>
<td>0.25, 2, max(0.1; 2%), 0.25</td>
</tr>
<tr>
<td><strong>ARSCLKAZRIKOLLIAS</strong></td>
<td>multiple outputs from first Kollias algorithm</td>
<td>Equivalent reflectivity factor, Mean Doppler velocity, Spectrum width</td>
<td>dBZ, m s⁻¹, m s⁻¹</td>
<td>5, 0.1, 0.05</td>
</tr>
<tr>
<td><strong>ARSCLKAZRBNDIKOLLIAS</strong></td>
<td>Cloud boundaries retrieved from KAZRARSCL</td>
<td>Cloud top height, Cloud base height</td>
<td>m, m</td>
<td>30, 30</td>
</tr>
</tbody>
</table>
2.2 Periods with Mesoscale Cellular Convection

Periods unambiguously exhibiting open and closed MCC are determined through visual inspection. The visual inspection involves the following steps:

1. 6-hourly KAZR time height transects during, and shortly before and after the MCAO events at the AMF1 site, as defined in Geerts et al. (2022) and Lackner et al. (2023a), are created (see Figure 1).

2. Open cellular periods are identified, when broken cloud cover, convective updrafts and numerous distinct cells of different sizes and depths are observed (Figure 1a-c).

3. Closed cellular periods are identified, when a shallow continuous cloud cover with light precipitation is observed which may not reach the surface at times (Figure 1e-f).

4. Very short periods of MCC (< 3 hours), periods with disorganized clouds, periods with polar lows (Lackner et al., 2023b), and periods with MCC connected to mid/upper-tropospheric clouds are not considered in the analysis.

5. The KAZR based identification (open, closed or neither MCC) is compared to the horizontal cloud structure in satellite imagery (see Supporting Information).

An example of the distinct vertical structure of open (closed) MCC in the KAZR data is shown in Figure 1a-c (1e-g). The closed-cell example appears to be mostly stratiform (Figure 1e). Consistent with the literature, we still refer to these cloud patches as mesoscale cellular convection (MCC), even though the convective origin may not be apparent.
Figure 1. KAZR time-height transect for an open MCC period (a-d) and a closed MCC period (e-h) showing (a,e) KAZR reflectivity factor and \textit{INTERPOLATEDSONDE} temperature contours, (b,f) Doppler velocity, (c,g) spectrum width, and (d,h) objectively identified cells via the watershed segmentation method (see Section 2.3). Note that the 5 colors in (d,h) merely are recycled and do not indicate any cell grouping. The numbers on top of each cell indicate the best-matched unit, introduced in Section 2.4.
Table 2 summarizes all MCC periods used in this study. A total of 13 periods with a total of ~249 hours of open MCC were identified at the AMF1 site, compared to 4 periods with a total of ~57 hours of closed MCC. In total, this accounts for ~37% of the MCAO periods during COMBLE identified in Geerts et al. (2022) (which themselves cover 19% of the full 6-month field phase). A broad spectrum of environmental conditions prevailed during these 17 periods. The $M$ value is above the COMBLE MCAO mean of 4.1 K for most open MCC periods (10 of 13), indicating that open MCC tends to occur during more intense MCAOs. Closed MCC on the other hand, tends to occur during weaker MCAOs, with the mean $M$ (1.2 K) being much lower than for open MCC (6.0 K). This difference in intensity likely explains why open MCC is observed much more frequently than closed MCC at the AMF1 site: weaker MCAOs often lack the synoptic support to cover the large distance between the ice edge and the COMBLE observation site (~1000 km). Near-surface temperatures (winds) tend to be warmer (weaker) during closed MCC (Table 2). Mean cloud top heights (CTH) are typically higher during open MCC ranging from 2.2 – 3.3 km, compared to 1.9 – 2.4 km for closed MCC. However, it should be noted that CTHs are much more variable during individual open MCC periods and often exceed 4 km (see Figure 1).

<table>
<thead>
<tr>
<th>Start Time</th>
<th>End Time</th>
<th>Type</th>
<th>Time</th>
<th>$M$</th>
<th>$T$</th>
<th>$U$</th>
<th>WD</th>
<th>CTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 Dec 2100 UTC</td>
<td>02 Dec 0330 UTC</td>
<td>Open</td>
<td>390</td>
<td>4.9</td>
<td>2.3</td>
<td>13.3</td>
<td>324</td>
<td>2.7</td>
</tr>
<tr>
<td>02 Dec 0700 UTC</td>
<td>02 Dec 1530 UTC</td>
<td>Open</td>
<td>510</td>
<td>4.1</td>
<td>2.2</td>
<td>12.1</td>
<td>323</td>
<td>2.5</td>
</tr>
<tr>
<td>31 Dec 0000 UTC</td>
<td>31 Dec 1915 UTC</td>
<td>Open</td>
<td>1155</td>
<td>3.7</td>
<td>2.1</td>
<td>10.5</td>
<td>286</td>
<td>2.5</td>
</tr>
<tr>
<td>04 Jan 0830 UTC</td>
<td>05 Jan 0200 UTC</td>
<td>Open</td>
<td>1050</td>
<td>6.3</td>
<td>0.3</td>
<td>10.6</td>
<td>330</td>
<td>2.6</td>
</tr>
<tr>
<td>21 Jan 2220 UTC</td>
<td>22 Jan 1830 UTC</td>
<td>Open</td>
<td>1210</td>
<td>3.1</td>
<td>2.3</td>
<td>9.2</td>
<td>328</td>
<td>2.4</td>
</tr>
<tr>
<td>02 Feb 1330 UTC</td>
<td>03 Feb 0615 UTC</td>
<td>Open</td>
<td>1005</td>
<td>5.0</td>
<td>0.8</td>
<td>10.9</td>
<td>320</td>
<td>2.2</td>
</tr>
<tr>
<td>03 Feb 1040 UTC</td>
<td>03 Feb 1845 UTC</td>
<td>Open</td>
<td>485</td>
<td>6.3</td>
<td>-1.0</td>
<td>7.3</td>
<td>308</td>
<td>2.6</td>
</tr>
<tr>
<td>04 Feb 1730 UTC</td>
<td>06 Feb 1545 UTC</td>
<td>Open</td>
<td>2775</td>
<td>7.2</td>
<td>-1.0</td>
<td>11.5</td>
<td>308</td>
<td>2.7</td>
</tr>
<tr>
<td>13 Mar 0800 UTC</td>
<td>14 Mar 0540 UTC</td>
<td>Open</td>
<td>1300</td>
<td>8.0</td>
<td>-3.0</td>
<td>9.3</td>
<td>336</td>
<td>3.3</td>
</tr>
<tr>
<td>28 Mar 0320 UTC</td>
<td>30 Mar 0945 UTC</td>
<td>Open</td>
<td>3265</td>
<td>7.1</td>
<td>-2.2</td>
<td>10.2</td>
<td>322</td>
<td>3.0</td>
</tr>
<tr>
<td>09 Apr 1420 UTC</td>
<td>10 Apr 0200 UTC</td>
<td>Open</td>
<td>700</td>
<td>4.8</td>
<td>-0.3</td>
<td>8.3</td>
<td>290</td>
<td>2.7</td>
</tr>
<tr>
<td>10 Apr 0810 UTC</td>
<td>10 Apr 1620 UTC</td>
<td>Open</td>
<td>490</td>
<td>5.0</td>
<td>0.3</td>
<td>6.8</td>
<td>304</td>
<td>2.8</td>
</tr>
<tr>
<td>10 Apr 1940 UTC</td>
<td>11 Apr 0530 UTC</td>
<td>Open</td>
<td>590</td>
<td>4.7</td>
<td>0.2</td>
<td>7.5</td>
<td>278</td>
<td>2.5</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td>14930</td>
<td>6.0</td>
<td>-0.4</td>
<td>10.1</td>
<td>314</td>
<td>2.7</td>
</tr>
<tr>
<td>17 Apr 0530 UTC</td>
<td>17 Apr 1630 UTC</td>
<td>Closed</td>
<td>660</td>
<td>1.1</td>
<td>3.3</td>
<td>8.8</td>
<td>290</td>
<td>1.9</td>
</tr>
<tr>
<td>24 Apr 1700 UTC</td>
<td>26 Apr 0010 UTC</td>
<td>Closed</td>
<td>1870</td>
<td>1.6</td>
<td>3.2</td>
<td>8.5</td>
<td>314</td>
<td>2.4</td>
</tr>
<tr>
<td>26 Apr 0300 UTC</td>
<td>26 Apr 1050 UTC</td>
<td>Closed</td>
<td>470</td>
<td>0.9</td>
<td>2.5</td>
<td>8.1</td>
<td>27</td>
<td>2.3</td>
</tr>
<tr>
<td>04 May 1050 UTC</td>
<td>04 May 1730 UTC</td>
<td>Closed</td>
<td>400</td>
<td>0.2</td>
<td>4.6</td>
<td>6.7</td>
<td>287</td>
<td>2.3</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td>3400</td>
<td>1.2</td>
<td>3.3</td>
<td>8.3</td>
<td>317</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Note: $M$: MCAO-index; $T$: 2-m Temperature; $U$: 10-m wind speed; WD: 10-m wind direction; CTH: Cloud top height. The time in the summary rows (‘average’) is the total cumulative time.

2.3 Cell Identification

To examine the characteristics of open and closed MCC, individual cloud sections are identified by applying a marker-based watershed segmentation to the KAZR radar reflectivity field (Vincent & Soille, 1991). This identification is necessary for the classification of the individual
cloud sections described in Section 2.4. It should be noted that the identified cloud sections are transects through specific parts of clouds and not necessarily representative of the three-dimensional structure of the transected cloud. This problem is addressed by also applying the watershed segmentation to the MET Norway C-band scanning radar data, to analyze the transect locations. For simplicity, these cloud transects will be referred to as cells for the rest of the study.

Watershed segmentation is a segmentation technique with its origins in image processing that has also found use in analyzing atmospheric data (e.g., Martini et al., 2014; Wu & Ovchinnikov, 2022). In the case of a marker-based watershed segmentation, all points in a dataset that are assigned to a predetermined marker comprise an identified object. Similar to how geographical locations are part of a water drainage basin (watershed), data points are assigned to the marker to which they have the steepest downward gradient.

For this study, the predetermined markers are local maxima in the smoothed KAZR reflectivity field. A Gaussian filter with a standard deviation of 1 is used for the smoothing. The smoothing and each following step use linear units of reflectivity (Z) instead of logarithmic units (dBZ). The local maxima are determined with two requirements: their value must exceed 2.5 dBZ (converted to Z) and they must be at least 80 KAZR profiles away in time from a larger value (320 seconds at 4 second time resolution). The corresponding minimum width depends on the advecting wind speed, e.g. 3.2 km for a 10 m s\(^{-1}\) layer-mean wind. Other requirements and thresholds were tested, but this setup was found to achieve the least amount of over- and under-segmentation of cells in a visual analysis. However, none of the tested setups were able to completely avoid unrealistic segmentation outcomes. Clear segmentation errors are corrected by manually removing (identifying) specific maxima associated with over (under) segmented cells. After the local maxima are determined, the watershed segmentation is conducted on the inverted reflectivity field, since the algorithm determines objects based on downward gradients. Cells are confined to a -30 dBZ reflectivity threshold, which exceeds the KAZR sensitivity even at a largest relevant range (~5 km). Each object identified by the watershed segmentation is an individual cell. Lastly, small cells (< 45 KAZR profiles = 3 minutes) and cells with shallow cloud depth (< 50 KAZR vertical levels between the lowest and highest echo, or 1500 m) are ignored since they have too few data points for the method used to classify cells in Section 2.4. Furthermore, cells with high radar echo bases (> 1500 m above sea level) are ignored since these were typically not MCC but other types of clouds. In total, 889 cells are identified during the 17 periods, 722 (167) of which are during the open (closed) MCC periods, i.e. 81%. Examples of identified cells in a vertical transect with open MCC and closed MCC are shown in Figure 1d, h. Note that continuous cloud structures can be identified as multiple cells, due to the often multi-cellular structure of what might appear as single cloud in satellite imagery. Specifically, in the case of the closed MCC periods, the cells should not be seen as distinct clouds but rather as a continuous cloud field that is made up of separate precipitation cores.

The corresponding horizontal structure of the vertical cell transects is identified in the MET Norway C-band CAPPI data with a similar watershed segmentation approach: local maxima must exceed at least 2.5 dBZ and must be separated by at least 3 grid points (3 km). Cells in C-band are confined to a -10 dBZ threshold instead of -30 dBZ, since thresholds below -10 dBZ were found to connect multiple cells into single large objects. Furthermore, no smoothing is applied since the 1 km resolution data is already sufficiently smooth. This horizontal identification is only conducted
in a small area (75x75 km) around the AMF1 site sufficiently large to identify the cell located over
the site. Due to missing C-band data mainly during the 28-30 March 2020 period (see Table 2), a
corresponding horizontal structure was identified for only 697 vertical cell transects. For an
example of an identified cell, see Figure S3.

2.4 Cell Classification

The next step is to classify the identified cells based on their vertical structure of KAZR
reflectivity, Doppler velocity (i.e., hydrometeor vertical motion), and spectrum width. Since the
melting level is usually at or very close to sea level, the radar echoes are dominated by snow.
Spectrum broadening in snow is generally largely due to atmospheric turbulence (Doviak & Zrnic,
1993). For instance, the red-colored cell around 11:20 UTC (Figure 1a-d) is marked by strong
convective updrafts and high spectrum width, due to high buoyantly-generated turbulent kinetic
energy. The four closed MCC periods all occurred in late April – early May with melting levels a
few hundreds of meters above the surface, which impacts the profiles of all three radar moments.

For the purpose of classifying the 889 cells, we ignore the horizontal cloud structure (i.e.,
the time dimension of KAZR profiles, shown in Figure 1), since it is highly dependent on deep-
layer wind shear in the direction of cell motion. Instead, the vertical distribution of the three radar
moments, i.e. the contoured frequency-by-altitude diagrams (CFADs) of reflectivity, Doppler
velocity, and spectrum width, provide a more standardized representation of the individual cells.

An unsupervised machine learning algorithm referred to as the self-organizing map (SOM)
is utilized for this CFAD-based classification (Kohonen, 1982). The main objective of the SOM is
to represent archetypical patterns in datasets using an interconnected node structure, with the user
defining the structure arrangement. The basis of the SOM algorithm is the following equation,
which updates the SOM nodes \( m_{i,j} \) in each training step \( t \):

\[
m_{i,j} (t + 1) = m_{i,j} (t) + a(t) \times n(t) \times [x(t) - m_{i,j} (t)]
\]

The subscripts \( i \) and \( j \) denote the location of the node in the grid and \( x(t) \) is the input
vector. The learning rate \( a(t) \) and neighborhood function \( n(t) \) determine how much a node is
updated in a training step. The learning rate \( a(t) \) at each training step is the following function:

\[
a(t) = \frac{a(t = 0)}{1 + \frac{5t}{t_n}}
\]

The total number of training steps is denoted by \( t_n \), here selected as 10,000. The
neighborhood function is a Gaussian distribution (provided by the minisom package for python)
with a standard deviation \( \sigma(t) \), which follows the same decrease at each training step as Equation
2. At the first training step, \( a(t = 0) = 0.2 \) and \( \sigma(t = 0) = 3 \). The center of \( n(t) \) is always located
on the node that has the smallest Euclidean distance to the input vector \( x(t) \). This node is referred
to as the best match unit (BMU). Thus, values of \( n(t) \) decrease the further a node is away from
the BMU, leading to the arrangement of the SOM nodes by similarity. The decrease of \( \sigma(t) \) with
each training step continually reduces the value of \( n(t) \) at nodes other than the BMU, where \( n(t) \)
always equals 1. The selected values for the number of training steps (Figure S4), the learning rate
and neighborhood function minimize the quantization error (QE, Kohonen, 2001) and topographic error (TE, Kiviluoto, 1996), compared to other tested values. Furthermore, the number of cells each node is the BMU for, does not have major outliers (Figure 2) indicating good performance of the SOM and the topography of the SOM is well organized (Figure S5).

The input vector \( \mathbf{x}(t) \) contains the data that is analyzed by the SOM algorithm, in this case the CFADs of all three KAZR moments, for a specific cell. The CFADs’ vertical structures are normalized by CTH (a cell’s highest echo top), to focus the algorithm’s outcome on the distribution of the radar variables within cloud, and not the variability of CTHs (which otherwise would dominate the SOM nodes). These CFADs (with 32x32 bins each) provide a standardized and statistically comparable representation of each cell. Furthermore, the CFADs are limited to specific ranges of the radar variables to better highlight the differences between nodes (see Figures 2 – 4). This excludes a very small fraction of available radar pixels (< 0.6 %), but does not impact the results. Note that while cell width and CTH do not define the SOM nodes, they will be examined as we characterize each node.

Each training step of the SOM consists of calculating Equation 1 for all 889 cells at each node. In each training step, cells are selected in a random order. For this study, a grid of 3x3 nodes with a rectangular topology is selected. Larger node setups with grids of 4x4 and 5x5 nodes were tested as well. While these setups have a lower QE and a similar TE compared to the 3x3 grid, the additional information contained in the additional nodes is small and outweighed by the physical interpretability of the relatively few nodes of the 3x3 grid (see Supporting Information). The final SOM resulting from the training process is referred to as the master SOM, which classifies the cells into nine categories. The characteristics of each node of the master SOM are described, first in terms of its composite radar moment vertical structure (e.g., reflectivity, shown in Figure 2) (Section 3.1), then in terms of cloud morphological parameters (Section 3.2), and finally in terms of cloud microphysical and dynamical parameters (Section 3.3). Through this analysis, insights into the dynamical and microphysical characteristics of each node are gained and in Section 4 possible physical meanings are discussed for each node. Note that the SOM technique treats each cell independently. It ignores the temporal succession of cells or cell nodes as illustrated in Figure 1d, h, and thus any interaction that may occur between cells.
Figure 2. Reflectivity part of the master SOM. Each node (labeled as [row #, column #]) shows the normalized CFADs of reflectivity resulting from the training of the SOM. The numbers on the top right of each node indicate for how many cells (out of a total of 889) each node is the BMU. The red dashed line indicates the 10th percentile, the solid line the mean, and the red dotted line the 90th percentile.

3 Results

3.1 Cloud Vertical Structure

In this section, the individual nodes of the master SOM are used to describe the vertical cloud structure of MCC observed during the identified periods (Table 2). Figures 2 – 4 display the reflectivity, Doppler velocity (hydrometeor vertical motions) and spectrum width parts of the master SOM. For a better comparison of the mean reflectivity profiles in Figure 2, they are displayed together in Figure S8.
Figure 3. As Figure 2 but for Doppler velocity (hydrometeor vertical motion). The white vertical dashed line indicates 0 m s\(^{-1}\) Doppler velocity. The percentages on top indicate the percentage of positive Doppler velocities in each CFAD.

The results of the SOM highlight distinct differences between different types of cells. Starting with the reflectivity part of the master SOM (Figure 2), the mean reflectivity profile has the largest values in the left column (Figure 2a,d,g) and the lowest values in the right column (Figure 2c,f,i). Node [2,1] has the largest mean reflectivity maximum (~18 dBZ) and node [1,3] the smallest (~4 dBZ). Despite having similarly large mean reflectivity values, the most frequent reflectivity values in node [1,1] (Figure 2a) are lower compared to the two other nodes in the same column. The two lower nodes in the left column and also node [3,2] (Figure 2h) have a structure indicative of precipitation with relatively large reflectivity values (> 10 dBZ) being frequent in the lower half of the cloud. However, for node [3,2] reflectivity values are overall lower, and the most frequent reflectivity values are between 0 – 10 dBZ in the upper half of the cells. The right column is characterized by frequent low reflectivity values (< 0 dBZ) in the upper half of the cells and fewer echoes near the surface, indicating the presence of non-precipitating cell parts. Especially, for nodes [1,3] and [2,3] the frequency of reflectivity values above 0 dBZ is very low close to the surface. Compared to the right column, nodes [1,2] and [2,2] (Figure 2b,e) have larger mean...
reflectivity values and the highest frequencies are closer to the middle of the cells. Nevertheless, these two nodes have weak or no surface precipitation as indicated by the low frequency of higher reflectivity values close to the surface.

Continuing with the Doppler velocity part of the master SOM, a distinct decrease in the strength of the hydrometeor vertical motions is evident from the top left to the bottom right (Figure 3). The comparatively strong vertical motions in nodes [1,1] and [2,1] (Figure 3a,d) are indicative of strong convection. By definition, convection is characterized by ascending hydrometeors during part of the cloud lifecycle or in a cloud region (Houze, 2014). While the overall fraction of upward hydrometeor motions is larger in node [2,1], stronger updrafts (> 1 m s\(^{-1}\)) are more frequent in node [1,1]. Hydrometeor vertical motions are somewhat weaker in nodes [1,2], [2,2] and [3,1] (Figure 3b,e,g), still indicative of convection, but overall the convection is weaker compared to the two nodes with strong convection. Even weaker are the hydrometeor vertical motions in node [3,2] and the right column (Figure 3c,f,h,i). If convection is at all present within these cells it is either very weak or limited to a small part of the cell. Notable is node [1,3] which has a strong increase in mean Doppler velocities near cloud top, possibly indicative of cloud top processes.

**Figure 4.** As Figure 2 but for spectrum width.
Lastly, the spectrum width part of the master SOM (Figure 4) shows a very similar pattern compared to the Doppler velocity: the largest spectrum width values are found in the top left and the lowest in the bottom right. Again, this is a result of the strength and presence of convective motions, which produce larger spectrum width values through the generation of turbulence. Comparable to Doppler velocity, node [1,3] has an increase in spectrum width near cloud top.

### 3.2 Cloud Morphological Characteristics

**Figure 5.** Fraction of cells, which are from open (red) and closed (blue) MCC periods at a specific SOM node is displayed by the bars. The sum of the two bars in each node equals 100%. The numbers listed in each bar indicate the percentages of the total number of open (left) or closed (right) MCC cells that are in each node: the sum of the nine numbers equals 100%.
To examine the characteristics of the SOM nodes further, other observations collected during the identified cells are collectively analyzed for each node. In this section, cell morphological characteristics are investigated. First, it is analyzed what fraction of cells in each node is from either open or closed MCC periods (Figure 5). For instance, 108 of the 109 cells in node [2,1] are from open MCC periods, accounting for 99.1% of the cells in this node and 15.0% of all 722 cells identified during open MCC periods. The left column of the SOM contains almost exclusively cells from open MCC periods (Figure 5a,d,g). Over 44% of all cells during open MCC periods are within these three nodes. The other six nodes show a mix of cells from open and closed MCC periods. Notable are nodes [1,3] and [3,2] (Figure 5c,h). Together they contain more than 62% of all cells from closed MCC periods. Node [1,3] is dominated by closed MCC periods as it has the largest (smallest) fraction of closed (open) cells of all nodes, while node [3,2] has an almost equal split between open and closed cells.

Next, the horizontal width of the cells is examined. The horizontal width of the cells is determined by multiplying the temporal width of the cells at normalized height levels with the closest available INTERPOLATEDSONDE wind speed (Figure 6), i.e., it is the along-wind width only. Due to the standardized CFADs, the classification contains only height-relative information about the cloud width, e.g., upper vs lower levels. The absolute cloud width matters less, it depends in part on a choice in the watershed segmentation method. What matters more is the vertical structure of the cell width. There are substantial differences between the nodes in terms of the vertical structure of their width. Nodes [1,1] and [1,2] (Figure 6a,b) are the smallest in width on average. Additionally, node [1,1] shows a gradual decrease in width in the upper half of the cells, not seen for any other node. The nodes with the largest average widths are nodes [2,1] and [3,1] (Figure 6d,g). The right column and node [3,1] (Figure 6c,f,g,i) show a clear tendency to be wider in the upper half of the cells than closer to the surface, while for node [2,1] and the middle column (Figure 6b,d,e,h) width changes much less with height (except very close to cloud top).

Figure 7 displays CTH, cloud top temperature (CTT), and cloud base height (CBH) distributions for each node. A distribution of CTHs, CBHs and CTTs is available for each cell. Here, we use the 90\textsuperscript{th} percentile CTHs (10\textsuperscript{th} percentile CBHs and CTTs) of each cell to avoid cells being represented by outliers of these variables. The CTH of a cell is determined based on the cell mask from the watershed segmentation (see Figure 1d,h). The CTT is the INTERPOLATEDSONDE temperature at this height at the matching time. CBH estimates are from the ARSCLKAZRBNDIKOLLIS product, which is derived from ceilometer and lidar data. Since cells can overlap in time, cloud base heights are assigned to the cell that is the closest to the surface in any given KAZR profile. The distributions of CTH and CTT show a large spread within individual nodes (Figure 7), which can be explained by the variability of the mean CTHs of the MCC periods (see Table 2). However, the distributions also reveal differences between the nodes. Mean CTHs (CTT) reach from 1.9 km (-14 °C) in node [1,3] (Figure 7c) to 3.6 km (-32 °C) in node [2,1] (Figure 7g). These two nodes have other notable features compared to the rest of the nodes. Node [1,3] has very narrow distributions of CTH and CTT, while node [2,1] is the node that has more cells than any other node (~25%) reaching homogeneous freezing temperatures at cloud top [<-38 °C (Ickes et al., 2015)]. The rest of the nodes mostly have mean CTHs (CTTs) between 2 – 3 km (-20 – -30 °C). Mean CBH is between 0.8 – 1.1 km for most nodes, except nodes [2,1] and [3,1] for which mean CBH is ~0.5 km (Figure 7d,g). Thus, cells in these two nodes have
on average the largest cloud depths (CTH – CBH). The CBH distribution of node [3,3] is notable for having the largest spread of all nodes (Figure 7i).

**Figure 6.** Width of the identified cells at each node of the master SOM. The solid line indicates the mean and the shaded area the interquartile range.
Figure 7. Distributions of cloud top temperature, cloud top height, and cloud base height of the identified cells at each node of the master SOM. The box plots inside the violin plots indicate mean (diamond), median (solid black line), interquartile range (box), and 10th and 90th percentile (whiskers).

The last morphological characteristic that is analyzed is the horizontal size of the cells (area) and the location at which the KAZR transects the cells. This analysis is based on linking the vertical cell transects with horizontal cell maps (C-band radar echoes) as they pass over the AMF1 site (see Section 2.3). To analyze whether specific nodes are correlated with transects through specific parts of cells (cell cores vs. cell edges), a minimum relative distance $r_{\text{min}}$ from the cell center (geographical center of cell; see Supporting Information) is calculated for each cell. $r_{\text{min}}$ is
the closest distance of the KAZR to the cell center relative to the cell size. The calculation of $r_{\text{min}}$ is explained in the Supporting Information and is used instead of an absolute distance to account for the different sizes of cells. Note that this analysis applies to a height of 1 km since it is based on a CAPPI product at 1 km AMSL (see Section 2.3). The calculation of $r_{\text{min}}$ also assumes that cells are circular for simplicity. This simplification explains why $r_{\text{min}}$ can exceed 100 %, although this is not very common (Figure 8). If cells were all circular, then the average $r_{\text{min}}$ would be 50% and the distribution uniform, given that transect locations are stochastic. The average $r_{\text{min}}$ is <50% for most nodes, simply because very marginal cell transects are too small to be identified as individual cells.

Mean cell sizes are $\sim80$ – $110$ km$^2$ depending on the node with smaller (larger) mean cell sizes found in the top (bottom) row. However, there is a large spread of cell sizes with the smallest cells being only a few km$^2$ and the largest exceeding 200 km$^2$. Based on satellite imagery Wu & Ovchinnikov (2022) report mean cell sizes $\sim300$ km$^2$ in the vicinity of the AMF1 site for the two March 2020 open MCC periods during COMBLE. This discrepancy is likely due to our radar-based approach that may distinguish multiple cells that may appear as a single cloud in satellite imagery (see Figure 1). Moreover, the cell sizes apply to a height of 1 km instead of cloud top.

Most nodes have a large spread of where cells are transected (Figure 8). Cell cores ($r_{\text{min}}$ < 25 %) and edges ($r_{\text{min}}$ > 75 %) are transected in each node. Notable are nodes [2,1], [2,2], [3,1] and [3,2], which are more related with transects of cell cores, as over 50 % of cell transects are through the cell core (Figure 8d,e,g,h). Barely any transects (< 5 %) within these nodes only transect cell edges. On the other hand, nodes [2,3] and [3,3] are the two nodes that appear to more frequently (~25 %) have transects that only sample cell edges (Figure 8f,i). To summarize there are three groups of nodes based on where cells are transected: the top row contains transects of all parts of cells, the left two nodes in the bottom two rows mostly contain cell transects of cell cores, and most transects only going through cell edges are found in the right nodes of the bottom two rows. These groups can also be justified by comparing the $p$-values (based on a t-test) of the $r_{\text{min}}$ distributions. Distributions that have large $p$-values (> 0.2) when compared to each other are in the same group.
Figure 8. Distributions of cell size and minimum relative distance of the KAZR from the cell center based on horizontal C-band scanning radar data at each node of the master SOM. The displayed cell sizes are at the time of the minimum relative distance. For a description of how the minimum relative distance is calculated, see Supporting Information. Box plots are as in Figure 7.

3.3 Cloud Microphysical and Dynamical Characteristics

In this section, microphysical and dynamical characteristics of the SOM nodes are analyzed, specifically four variables: microwave radiometer liquid water path (LWP) (Figure 9), a cloud phase classification from the THERMOCLDPHASE product (Figure 10), surface precipitation rate, and surface virtual potential temperature ($\theta_v$) anomaly (a measure of cold pool strength) (Figure 11). Since multiple cells can be identified in a single vertical column, it is not
always possible to unambiguously assign a LWP measurement (a vertically integrated quantity) to
a single cell. Therefore, LWP is only assigned to a cell if that cell makes up more than 80% of
cloud in a vertical profile and is the closest cell to the surface. Data points from the 2D
THERMOCLDPHASE product are assigned to the cell they cover. Precipitation measurements and
the surface $\theta_v$ anomaly are assigned to the cell closest to the surface. The minimum (including
negative values) difference between the instantaneous and the six-hour mean surface $\theta_v$ (centered
on the instantaneous value) during each cell is used as a measure for cold pool strength.

The SOM nodes show strong contrasts in terms of presence of liquid water (Figure 9). The
largest LWP values are found in node [1,1] with a mean value (90th percentile) of 0.33 kg m$^{-2}$ (0.86
kg m$^{-2}$) (Figure 9a). Other nodes with substantial but lower LWP values are nodes [1,2], [1,3],
[2,1], and [3,2]. The lowest mean (90th percentile) values of LWP are in node [3,3] with 0.05 kg
m$^{-2}$ (0.12 kg m$^{-2}$) (Figure 9i) and nodes [2,3] and [3,1] have only slightly larger values. Noticeable
for almost all nodes is that the mode of the distribution is at low LWP values at or below 0.05 kg
m$^{-2}$, indicating that the presence of liquid is confined to specific parts of the cells. The exceptions
are nodes [1,3] and [3,2]. While the latter has a secondary mode at larger values (Figure 9h), node
[1,3] completely lacks the mode at low values (Figure 9e). Instead, the mode is similar to the mean
value, indicating a constant presence of liquid within cells associated with this node. These two
nodes are the same nodes that have the largest fraction of cells from closed MCC periods (Figure
5).

Figure 9. Distributions of liquid water path (LWP) of the identified cells at each node of the master
SOM. Box plots are as in Figure 7.
Figure 10. Fraction of retrieved cloud phase of the identified cells at each node of the master SOM. Note that R+D refers to rain and drizzle combining the rain, drizzle, and liquid + drizzle classification of the THERMOCLDPhase product (see Table 1). Note that this product is only available for cells identified after 11 February 2020 (441 of 889).

The THERMOCLDPhase product corroborates these findings (Figure 10). The largest fraction of mixed phase and liquid in any node is found in node [1,3] (Figure 10c), the node associated with a constant presence of liquid in cells. Furthermore, the node with the largest mean LWP, node [1,1], has the second largest fraction of combined mixed and pure liquid phase (Figure 10a). Generally, the other nodes follow this pattern of a correlation between the frequency of larger LWP path values and the fraction of mixed phase. Two notable deviations from this pattern are nodes [2,1] and [3,2]. Node [2,1] has a relatively low fraction of mixed phase (~15 %, Figure 10d),
but much larger LWP values as compared to nodes with a similar mixed phase fraction. The low 
fraction of mixed phase is likely related to much larger ice water contents within this node than 
other nodes, as indicated by the large mean reflectivity values (Figure 2d). For node [3,2], the 
mixed phase fraction is very low (~7 %, Figure 10h), however, there is a comparatively large 
fraction of rain and drizzle (~ 6 %, Figure 10h) which is the reason for the secondary mode with 
larger LWP values (Figure 9h). The rain is associated with cells from closed MCC periods in this 
node.

**Figure 11.** Distribution of precipitation rate and virtual potential temperature anomaly of the 
identified cells at each node of the master SOM. Box plots are as in Figure 7.
Lastly, surface precipitation rates and cold pools are analyzed. Precipitation rates and (negative) $\theta_v$ anomalies increase steadily in each column from top to bottom (Figure 11). In the top row, precipitation rates are negligible and the mean surface $\theta_v$ anomaly is close to 0 K, and the distribution has few cells with larger anomalies. On average the strongest precipitation rates (0.7 mm hr$^{-1}$) and surface $\theta_v$ anomalies (-0.8 K) are found in node [3,1] (Figure 11g). The distributions of precipitation rates for all nodes have a mode at very low values (< 0.1 mm hr$^{-1}$), indicating that precipitation associated with the analyzed cells is usually caused by brief showers. Overall, the precipitation rates and the occurrence of cold anomalies correlate well. This indicates that cold pools are likely precipitation induced.

4 Discussion

4.1 Interpretation of the Cell Classification

Our analysis shows that a SOM-based cell classification can determine distinct characteristics of open and closed MCC clouds observed during MCAOs. For instance, the vertical distributions of the radar variables differ significantly between nodes. Furthermore, observations made independent from the KAZR such as LWP, surface precipitation, and surface $\theta_v$ anomalies show strong differences between the nodes. These distinct nodes reveal information about dynamical/microphysical characteristics of cell transects (the “signal”), but also about varying locations of the transects in evolving three-dimensional cloud structures. In this section, an interpretation of the individual nodes is provided concerning what possible types of cells they might describe. Cell transect location can be considered “noise” in the SOM classification, however neighboring cells (possibly in multi-cellular cloud systems) may interact, e.g., a decaying cell may trigger a new cell. Such interactions are not examined here: each cell is examined in isolation.

The left column of the SOM has two major things in common: these three nodes have the largest mean reflectivity profile values (Figure 2), and they are almost exclusively from open MCC periods (Figure 5). Thus, they describe the most intense cells observed during open MCC periods. There are major differences between these three nodes. Node [1,1] has strong convection as indicated by the vertical distributions of Doppler velocity and spectrum width (Figures 3a, 4a), but has on average shallower cloud depths, smaller cell widths and sizes than the bottom two nodes (Figures 6 – 8). Based on these characteristics, node [1,1] is interpreted as representing developing intense cells, that have strong convection, developing in depth and size. For node [2,1] convection is similarly strong (Figures 3d, 4d), but mean reflectivity values and cloud depth are the largest (Figures 2, 7). Moreover, cell widths and sizes are larger (Figures 6, 8). Thus, the interpretation of node [2,1] is that it contains mature intense cells. Lastly, compared to the mature intense cells, node [3,1] has substantially weaker convection (Figures 3g, 4g), cells are slightly shallower in depth and have an increased width in the upper half of cells. These characteristics indicate that node [3,1] generally contains weakening intense cells. To summarize, these three nodes describe a convective lifecycle of intense cells with a developing, mature and weakening stage.

The microphysical and dynamical characteristics of these nodes highlight important processes associated with this lifecycle. The presence of liquid decreases substantially as the lifecycle progresses (Figure 9, 10), while cold pools and surface precipitation rates intensify (Figure 11; also shown by the reflectivity distributions in Figure 2). Thus, the glaciation of cells at
the mature lifecycle stage, the ensuing precipitation formation, and the development of cold pools
appear to play a key role in the cell lifecycle. At the developing and mature stages, CTTs can be
sufficiently cold for homogenous freezing of supercooled liquid droplets for some cells, but in
most cases CTTs are higher, requiring heterogenous freezing (Figure 7a, b).

There is one other difference between these nodes that needs to be mentioned. Nodes [2,1]
and [3,1] tend to sample mostly cell cores, while node [1,1] does not show a clear trend in terms
of cell center proximity. It is unlikely that transect location variations significantly impact the
lifecycle stage interpretation: the transects of cell edges in node [1,1] are likely caused by some
intense updrafts developing on the edges of pre-existing cells. If these updrafts are not sufficiently
isolated, their horizontal reflectivity structure might not be separable from the cell edge on which
they develop. Furthermore, a newly developing updraft might make up a whole cell such that it
can be observed at the edge of that cell. Yet, the strong precipitation cores of the mature and
weakening intense cells are unlikely to be located on cell edges. This raises the question concerning
which nodes contain the edge transects of these lifecycle stages. This is discussed in the
interpretation of nodes [2,3] and [3,3].

Node [1,3] is dominated by cells from closed MCC periods. This node is characterized by
rather stratiform cloud profiles with frequent low reflectivity values near cloud top and weak
vertical motions (Figures 2 – 3). Near cloud top, an increase in upward hydrometeor motions and
spectrum width is found (Figures 3 – 4). This cloud top turbulence may be attributed to generating
cells (e.g., Hobbs & Locatelli, 1978; Rosenow et al., 2014) driven by cloud top radiative cooling
(Keeler et al., 2016). Such cells, with a reflectivity and Doppler velocity structure similar that of
generating cells at the top of frontal clouds (e.g., Plummer et al., 2015; Zaremba et al., 2022),
become apparent in a zoom-in of these cells, e.g., around 10:00 UTC in Fig. 1e (not shown). These
traits are indicative of stratiform clouds, with hydrometeors initiated at cloud top and slowly
growing as they fall through the cloud. Based on these stratiform characteristics and the lack of
surface precipitation, this node is interpreted as non-precipitating stratiform cells. Furthermore,
these cells have the shallowest cloud depth and the lowest spread in CTHs (Figure 7), and it is the
only node with a presence of liquid in all parts of cells (Figure 9). At the same time, there are no
cold pools (Figure 11). Since these cells are rather horizontally homogeneous, the transect location
is of little consequence (Figure 8). Very few cells from open MCC periods are in this node because
the relatively high $M$ values, strong surface heating, and intense convection during such periods
prevent the development of long-lasting stratiform clouds. The non-precipitating stratiform cells
are likely the “base state” in a closed MCC environment. They dominate in the closed MCC
example in Figure 1h. Yet, this cloud regime exists in an $M > 0$ environment and contains
occasional convective updrafts (Figure 1f), classified in different nodes. Hence, we still refer to
this as mesoscale convective updrafts (MCC) (consistent with McCoy et al., 2017 and others), even
though stratiform processes may dominate clouds and precipitation.

In terms of the vertical structures of the radar variables, cloud morphological and
microphysical characteristics, the comparison between nodes [1,2] and [2,2] is very similar to the
comparison between the developing and mature intense cells. The major difference is that
reflectivity values are lower, the convection is less intense (Figures 2 – 4), and the cells are
shallower (Figure 7). Thus, these two nodes are interpreted as developing ([1,2]) and mature ([2,2])
moderately convective cells. Unlike the intense convection, the moderate convection can occur in
the continuous cloud deck of closed MCC periods (Figure 5).

Node [3,2] is comparable to [3,1], but with lower reflectivity values and even weaker
convection (Figures 2 – 4). The cells are also slightly shallower (Figure 7). Thus, this node can be
interpreted as containing weakening convective cells as well, which can be of an intense or
moderate convective origin. In addition to these convective cells, it is notable that node [3,2] has
the second most cells from closed MCC periods (Figure 5). These cells appear to have different
development mechanisms than the convective cells in the same node. For instance, some cells in
Figure 1h are classified in this node but are rather stratiform, while two adjacent cells observed
around 13 March 2020 1040 UTC are also classified in this node and appear to be the remnants of
active convection (Figure 1d). This overlap is remarkable: cells with very different development
mechanisms (convective vs. stratiform) result in very similar vertical cloud structures at some
point during their development. Compared to the non-precipitating stratiform cells (node [1,3]),
the stratiform cells in this node produce surface precipitation (rain) (Figures 10, 11) and are
dominated by in-cloud ice as opposed to liquid (Figures 9, 10). This could indicate that the
formation of ice is important for the development of significant precipitation in closed MCC. To
summarize, node [3,2] contains a mix of precipitating stratiform cells and weakening convective
cells.

This leaves nodes [2,3] and [3,3] to be interpreted. Cell transects included in these two
nodes relatively frequently occur near cell edges (Figure 8). This is interpreted as these two nodes
being the only nodes that are not predominantly associated with a lifecycle stage and cell intensity,
but also the relative location of the transects. These two nodes include transects near the cell center;
these are interpreted as dissipating convective cells at the very end of the convective lifecycle. This
interpretation is based on the rather stratiform vertical cloud structures, cells being mostly from
open MCC periods, much larger cell widths near cloud top than closer to the surface, and often
elevated cloud bases (Figures 2 – 7). These two nodes also contain transects near the cell edge.
Such transects may be associated with any intensity or lifecycle stage. Especially when cells reach
their maximum size during the mature and weakening stages, cell edges likely lack precipitation
and have weak vertical motions making their vertical structure comparable to that of dissipating
cells. This leads to the SOM classifying cell edges from all lifecycle stages and intensities in these
two nodes instead of their corresponding node, and answers the question concerning the lack of
cell edge transects in specific nodes. In other words, these cloud-edge transects add noise,
revealing no information about intensity or lifecycle stage.

Lastly, it should be mentioned that the two nodes are not entirely the same: node [2,3] has
slightly stronger convective motions. This might be due to the development of new updrafts in
dissipating cells or edge transects through a convective updraft. The reason for the development
of new updrafts in node [2,3] might be related to the lack of surface precipitation and associated
cold pools (Figure 11). To summarize, nodes [2,3] and [3,3] contain a specific lifecycle stage
(dissipating convective cells), but also edge transects from all lifecycle stages.

4.2 Study Limitations and Future Work
A few things need to be considered when interpreting the results of the cloud classification. While the SOM yields discrete classifications, it should be remembered that these clouds occur on a continuous spectrum in terms of their development stage, the intensity of convection, or stratiform precipitation growth. The SOM with more nodes provided in the *Supporting Information* shows this continuous spectrum in more detail, but, as previously argued, the main information is captured by the 3x3 node SOM. Also, an individual cell defined objectively (watershed segmentation method) may consist of multiple smaller convective turrets at different life stages, e.g., the blue cell around 12:20 UTC in Figure 1a-d. The analysis provides a comparison of the horizontal and vertical structures of cells, but it is not possible to examine the temporal evolution of the three-dimensional structure of individual cells. Rather, the statistical analysis of a large sample of cells allows differentiation in terms of intensity, lifecycle stage, and transect location. The results provide strong evidence that a convective lifecycle dominates in open MCC environments. This sets the stage for further work that examines cell interactions and mergers in MCAO convection.

The approach to objectively classify clouds helps to conceptualize the defining characteristics of convective cells during MCAOs, and to elucidate the driving dynamical-microphysical processes. While buoyant ascent (possibly forced by outflow boundary convergence) clearly is the driving mechanism during relatively intense MCAOs with strong surface heating, stratiform cloud and precipitation processes play a role as well, especially under low $M$ value conditions. This may result in a closed MCC organization, which was less common during COMBLE, and observed only late in the field campaign (April-May).

Our analysis is limited to the comparatively short data record of COMBLE. While a certain range of MCAO intensities is captured, it is not necessarily representative of the long-term climatological conditions over the high-latitude oceans, and certainly not elsewhere. The approach presented here, using the COMBLE dataset, could be used to analyze MCC clouds at observational sites with a much longer data record such as ARM's North Slope of Alaska and Eastern North Atlantic (ENA) sites. On average, $M$ values are lower in the area of the ENA site during MCAOs (or post-frontal periods) (Fletcher et al., 2016), and the amount of time during which open and closed MCC occurs is much more equal (Jensen et al., 2021). A comparison with the ENA site allows examination of the importance of the ice phase in MCAO convection dynamics: the ENA site is considerably warmer, resulting in less common mixed-phase MCC.

In future work, we will analyze the MCAOs listed in Table 2 using a large-eddy simulation (LES), starting with the 13 March 2020 case (Kosović et al., 2023), following a similar approach as presented in this study. A radar simulator such as the Cloud Resolving Model Radar Simulator (Oue et al., 2020) will be used to create synthetic radar data from the LES that is comparable to KAZR observations. A comparison of the observational SOM classification to an LES-output based SOM classification may reveal whether the model reproduces MCC with the characteristics observed in this study. Another benefit of conducting this analysis with LES is that the three-dimensional spatial structure and the temporal evolution of the variables presented in this study can be analyzed, giving further insight into the impact of locations of cell transects on the cell classification. The LES further allows analysis of a large number of cells given that the simulation covers a large domain, stretching from north of the ice edge to south of the COMBLE observation site (~1,200 km distance).
5 Conclusions

In this study, the characteristics of open and closed mesoscale cellular convection (MCC) during marine cold-air outbreaks (MCAO) are investigated. The data were collected as part of the COMBLE field campaign at a coastal site in northern Norway. Unambiguous periods of open and closed MCC are determined within MCAOs, based on radar and satellite data. Periods of open and closed MCC are associated with different MCAO intensities and display distinct cloud top height differences. Watershed segmentation and self-organizing map (SOM) algorithms are used to objectively identify and classify individual cells observed during these periods. The identification and classification of cells are based on the composite frequency-by-altitude structure of the cells as observed by a high-resolution vertically-pointing radar. Furthermore, the horizontal structure associated with the vertical cell transects is identified with a similar watershed segmentation in scanning radar data. The profiling radar’s time-height transects across advecting cells may occur anywhere relative to MCC, which often has a complex, multi-core footprint as documented by the scanning radar. Invariably, the SOM nodes contain some distinguishing information about transect location, which adds noise to a process-focused classification. A sample size of 889 cell transects is large enough for a physically meaningful cell classification to emerge from this stochastic noise.

The main findings are as follows:

- The SOM algorithm yields nine nodes with distinct vertical distributions of the three radar moments, i.e., reflectivity, Doppler velocity, and spectrum width. Furthermore, there are distinct differences between the cloud morphological characteristics of the individual nodes. Some nodes group vertical transects of convective cells in distinct lifecycles stages and at different intensities, while others describe rather stratiform clouds. Such interpretation of the nodes is corroborated by independent microphysical and dynamical observations.

- A group of three nodes is interpreted as intense convection. These cells only occur during open MCC periods and are separated into developing, mature, and weakening intense cells. Developing cells are associated with strong updrafts, high spectrum width, shallower cloud depth, pockets of large liquid water path, and minimal precipitation. As the convection matures and then weakens, updrafts decrease in strength, cloud tops are higher and colder, the amount of liquid decreases indicating that cells glaciate, surface precipitation intensifies and cold pools develop below the cells. Two additional nodes contain cell transects in the decaying stage of the convective lifecycle. New updrafts can develop below these dissipating cells.

- Cells with more moderate convection compared to the intense cells are observed during open MCC and occasionally during closed MCC. This moderate convection has a similar lifecycle.

- During closed MCC periods, cells with stratiform characteristics are observed most frequently. These stratiform cells are separated into non-precipitating and precipitating cells. Non-precipitating stratiform cells are relatively shallow with cloud tops mostly >-15°C, have weak vertical drafts, somewhat enhanced spectrum width near cloud top, and mostly mixed-phase
cloud. The precipitating stratiform cells are dominated by ice, and are structurally very similar to weakening convective cells.

Acknowledgments: The COMBLE campaign was enabled by an ARM Mobile Facility deployment proposal to the Office of Science of the U.S. Dept. of Energy (DOE). This research was supported by DOE Atmospheric System Research (ASR) Grants DE-SC0018927 and DE-SC0021151. We appreciate the assistance from several ARM instrument and data mentors, especially Scott Giangrande. The National Center for Atmospheric Research is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977.

Open Research: COMBLE campaign data at the Andenes site (ANX) used in this study are available through the references found in Table 1. C-band equivalent radar reflectivity factor is obtained from the Norwegian Meteorological Institute (2020). MODIS satellite data used in the Supporting Information is from the MYD021KM (MODIS Characterization Support Team (MCST), 2017a) and MYD03 products (MODIS Characterization Support Team (MCST), 2017b). The watershed segmentation was implemented in python code using the sci-kit image package (Van Der Walt et al., 2014) and the self-organizing map algorithm was implemented in python code with the minisom package (Vettigli, 2018).

References


Introduction

This document provides supporting information about the methodology, concerning the performance of the self-organizing maps (SOM), the a priori selection of the number of nodes of the SOM, and how the distance of the AMF1 to the cell center is calculated. Furthermore, additional figures show examples of cloud morphology and an example of an identified cell in the horizontal C-band radar data. Another additional Figure (S8) provides an easier comparability of the mean reflectivity profiles seen in Figure 2.

Text S1. Sammon map of the master SOM

One way to test the performance of a SOM is to display the Euclidean distances between the nodes in two dimensional space and connect bordering nodes with lines. This display is a so-called Sammon map. Good performance is indicated by nodes being roughly equidistant from each other and the connecting lines not intersecting with other lines.

The Sammon map of the master SOM (Figure S5) does not have intersecting lines and the distances between nodes show no major outliers. Somewhat of an outlier is node [1,3]. This is likely explained by this node having a majority of cells from closed MCC periods as opposed to all other nodes having a majority of cells from open MCC periods.

Text S2. Testing of SOMs with more nodes

The selection of the number of nodes is mainly driven by the goal of classifying individual nodes. Thus, keeping the total number of nodes low is preferable for an easier physical interpretation of the nodes and discussion of their potential classification.
Furthermore, having too many nodes could lead to lengthy and redundant descriptions of similar nodes. Selecting an SOM with more nodes would need to be weight against the additional information these nodes provide.

In order to highlight that the main results would not be impacted by choosing a SOM with more nodes, Figures S6 and S7 show a master SOM for 5x5 nodes. A 4x4 nodes SOM was also trained and conclusions are similar (not shown) The data used to train these SOMs is the same as in Figures 2 – 4. Learning rate and number of training steps ($t_n$) remain the same for the training of the SOMs with the larger grids. The standard deviation ($\sigma(t)$) of the neighborhood function $n(t)$ is modified to account for the larger grid. For the SOM with 5x5 nodes, $\sigma(t)$ at each training step ($t$) is the following:

$$\sigma(t) = \frac{7}{1 + \frac{9t}{t_n}}$$

Overall, the SOMs with more nodes are comparable to the 3x3 SOM and show the same characteristics with a more gradual transition between nodes. Spectrum width is omitted here, but shows the same comparability to the 3x3 SOM. The 3x3 SOM condenses the information in the SOMs with larger grids by essentially grouping similar nodes while retaining the key information needed to differentiate between cells. Thus, choosing the larger grids would not provide much additional information while making the interpretation of the SOM more lengthy.

**Text S3. Calculation of the minimum relative distance to the cell center.**

The minimum relative distance ($r_{min}$) between the KAZR and the cell center is determined based on the minimum absolute distance ($d_{min}$) and the assumption that cells are circular with an area $A$. The area of each cell can be determined from the number of pixels it consists of in the CAPPI radar data (1 pixel = 1 km$^2$). Thus, the minimum relative distance is:

$$r_{min} = \frac{d_{min}\sqrt{\pi}}{\sqrt{A}}$$

The cell center is the geographical center of all the radar pixels that make up the cell with no weighting by reflectivity (black x in Figure S3). If the same cell is identified in multiple time steps, the location of the cell center and its size is determined in 10 seconds intervals by linearly interpolating the location and size of the cells, captured by the Met Norway composite at a 5 minute interval.
Figure S1. Infrared satellite image showing an example of a closed MCC period on 13 March 2020 at 11:10 UTC. The red star indicates the location of the AMF1 site. Data source: MODIS Aqua Band 31.

Figure S2. Infrared satellite image showing an example of a closed MCC period on 25 April 2020 at 11:30 UTC. Cloud tops are much lower and warmer compared to the open MCC period. Data source: MODIS Aqua Band 31.
Figure S3. Example of an identified cell in the horizontal C-band radar data. The cell itself is displayed with fully saturated colors, while the remaining data is displayed more transparent. The center of the cell is marked with a black $\times$, and the location of the AMF1 site with a black square.
Figure S4. Quantization error (QE) for different numbers of training steps. The QE starts to converge around 10,000 training steps, the selected number of training steps.

Figure S5. Sammon map showing the Euclidean distance between the nodes of the master SOM.
**Figure S6.** As Figure 2, but Reflectivity part of the master SOM with 5x5 nodes.

**Figure S7.** As Figure 3, but Doppler velocity part of the master SOM with 5x5 nodes.
**Figure S8.** Mean profiles of reflectivity at each node of the master SOM from Figure 2. Dashed lines are used to help differentiate nodes.