UiO : Department of Geosciences
University of Oslo

On the Homogeneity of Arctic Mixed-Phase Clouds

Stian Leer-Salvesen Dammann
Master's Thesis, Spring 2023

Abstract

Mixed-phase clouds are substantial contributors to the modulation of precipitation and the radiation budget, particularly in the Arctic where they are ubiquitous throughout the year. One important characteristic of mixed-phase clouds is the degree to which they can actually be considered mixed, as opposed to having ice and liquid spatially separated in so-called pockets. Despite the wide range of implications, few attempts have been made to quantify the nature of such pockets. In this thesis, spatial heterogeneity is investigated through a case study of an Arctic mixed-phase cloud field occurring 12 November 2019 in Ny-Ålesund, Svalbard. The case study is modelled with the Weather Research and Forecasting model, constrained with representative ice nucleating particle concentrations (INPC) and cloud droplet number concentrations (CDNC) following measurements conducted during the Ny-Ålesund Aerosol Cloud Experiment field campaign. Cloud heterogeneity is evaluated from two perspectives – in terms of (i) droplet size and (ii) phase. For perspective (i) a machine learning algorithm is applied to detect heterogeneity of in-situ droplet size distribution measurements. Large variations in droplet size distributions are detected even on the meter scale, and a apparent relationship between droplet size and ice concentration is revealed. For (ii), two new methods are developed for quantification of cloud-top phase heterogeneity in the model. It is found that the spatial scales of genuine mixing at cloud-top are particularly sensitive to INPC and CDNC, and are heavily influenced by external forcing, particularly from topography. Furthermore, it is showed that WRF severely underestimates ice concentrations when constrained to the clean conditions often measured in remote, pristine environments such as the Arctic. Inclusion of several secondary ice production (SIP) pathways helped to ameliorate the discrepancy, although not sufficiently. Improved knowledge of the SIP mechanisms is needed to better represent Arctic mixed-phase clouds in models.

Acknowledgements

First and foremost, I owe a huge amount of gratitude to my supervisors for making this thesis so much more fun than I could have imagined. Trude, for setting aside so much time and effort for guidance, and for always doing so with your infectious positivity and genuine interest. Britta, not only for providing me with all of the simulation runs this thesis was built upon, but also for your incredible ability to steadily walk me through the dense modelling forest, and for joining my side in the argument of car vs. bus. And Rob, for your amazing dedication to the subject which I have benefited and learned so much from – I have thoroughly enjoyed all of our impromptu meetings throughout the year, and not least our long discussions extending far into the depths of KS-space.

Furthermore, I would like to thank Fabiola Ramelli and Jan Henneberger from the atmospheric physics group at ETH Zürich for all help with the processing of the HOLOIMO data. Thanks also to everyone on the ISLAS 2022 and MC2-ICEPACKS 2023 teams for amazing fieldwork experiences in terrible weather. And to the rest of MetOs for making my stay here not only academically interesting, but also social and comfortable. I have truly appreciated being part of the section over the past two years.

Thanks to Lesesal 2418b and the crew at ZEB for all of the pleasant breaks, both the needed ones and the not-so-needed ones. Particularly to Eskil for your magnificent proofreading efforts, enduring both abbreviations and a multitude of dimensions. Thanks also to my family and all my nieces and nephews at Nesodden for providing me with the best hiding place imaginable when the city gets too big. And of course to Johanne, for making even a big city feel like home to me.

Contents

List of Figures

List of Tables

PART I

Introduction & Background

CHAPTER 1

Introduction

Mixed-phase clouds present a complex system of water in all three phases as solid, liquid and gas. They are observed across all latitudes and seasons (Shupe et al., 2008) owing partly to the broad range of temperatures between approximately -38 and 0°C at which they can form and exist.

Mixed-phase clouds play an important role in the radiation balance, particularly in the Arctic and over the Southern Ocean where they are ubiquitous throughout the year (Dong and Mace, 2003). This includes a negative shortwave effect due to reflection of incoming solar radiation as well as a predominantly positive longwave effect as a result of intercepted outgoing thermal radiation from the ground. The net radiative impact depends on macroscopic factors such as cloud base altitude or cloud top temperature, but the radiative properties are also largely dependent on the clouds' microphysical composition. The balance between supercooled liquid and ice is important because a liquid cloud is more reflective than a glaciated cloud of the same given total water content due to the smaller and more numerous cloud droplets (Sun and Shine, 1994). The size distribution of cloud droplets is also central as a cloud of generally smaller droplets appears optically thicker than a cloud of larger droplets given the same total water content (Lohmann, Lüönd et al., 2016). Knowledge of what controls the ice-liquid-interactions and microphysical properties of mixed-phase clouds is thus, essential to ultimately understand the Arctic climate system.

In order for ice to form in mixed-phase clouds, a small subset of atmospheric aerosols called ice nucleating particles (INPs; e.g. Pruppacher and Klett, 1997; Vali et al., 2015) are required. As such, INPs play a critical role in the concentration of ice in mixed-phase clouds and more generally in weather and climate by facilitating the microphysical and dynamic processes that follow the production of ice (Burrows et al., 2022). However, the number of ice crystals in clouds has been observed to exceed the number of INPs by up to four orders of magnitude (Auer et al., 1969). This discrepancy is typically attributed to secondary ice production (SIP), describing all processes where already nucleated ice initiates further ice production (Ladino et al., 2017). Several SIP mechanisms have been studied over the past decades (e.g. Hallett and Mossop, 1974, Vardiman, 1978, Phillips et al., 2018, Field et al., 2017), but despite their proposed importance for the prevalence of ice in clouds, the physical basis remains poorly understood (Field et al., 2017).

Due to the difference in saturation vapor pressure over liquid and ice, the mixed-

phase cloud system is thermodynamically unstable and thus, ice crystals may grow at the expense of the supercooled liquid droplets in a process referred to as the Wegener-Bergeron-Findeisen (WBF) process (Bergeron, 1928; Findeisen, 1938; Wegener, 1911). This has significant impacts on the properties of mixed-phase clouds, such as lifetime, precipitation rates and composition. However, the WBF process only acts on the microphysical level and is thus dependent on the spatial distribution of liquid droplets and ice particles within the cloud. Mixed-phase clouds may either be homogeneously mixed with ice and liquid uniformly distributed throughout the cloud, or conditionally mixed if the ice and liquid is clustered into separated single-phased pockets (Korolev and Milbrandt, 2022a). Spatial separation of the phases would act to severely inhibit the efficiency of the WBF process. Therefore, the microphysical structure of the clouds is important for accurate simulations of cloud macrophysical properties, precipitation, and radiation in numerical weather prediction (NWP) models as well as earth system models (ESMs).

Both reanalysis products, NWP models, and ESMs frequently underestimate the amount of supercooled liquid water in mixed-phase clouds (Komurcu et al., 2014) causing large errors in annual mean downwelling shortwave radiation in both the Arctic and Southern Ocean (Bodas-Salcedo et al., 2016; Naud et al., 2014; Vergara-Temprado et al., 2018). The partitioning of ice and supercooled liquid within clouds is commonly represented in ESMs effectively only as a function of temperature, and the point of equal abundance differs between models by up to 40°C, contributing substantially to the uncertainties in high-latitude cloud feedbacks (McCoy et al., 2015). Furthermore, for models that do not employ simple temperature-dependent phase functions, parametrizations of the WBF process typically assume genuine homogeneous mixing of the phases within cloudy grid cells (Storelvmo et al., 2008). However, observations show that such mixtures rarely occur in nature (Korolev et al., 2003) and that the phases may be separated at scales much smaller than the typical grid resolution used in NWP (Korolev and Milbrandt, 2022b). This indicates that models may overestimate the efficiency of the WBF process, yielding a too rapid depletion of the supercooled liquid. A better understanding of the factors that influence and drive the mixed-phase processes is thus of vital importance to improve the representations of mid- and high-latitude clouds in models and to reduce uncertainties related to cloud feedbacks in future climate projections.

This thesis is centered around a case study from November 12, 2019 of an Arctic mixed-phase cloud, which was extensively measured during the Ny-Ålesund Cloud Experiment (NASCENT; Pasquier, David et al., 2022). The case is modelled with the Weather Research and Forecasting (WRF; Skamarock et al., 2019) NWP model on three domains of unequal extent and resolution, applying a state-of-the-art cloud microphysics parameterization scheme (Morrison et al., 2009) to simulate a wide range of cloud properties. Measurements of INPs and cloud droplets were conducted during the campaign and are used here to constrain the model with realistic aerosol concentrations. Furthermore, in-situ observations of cloud hydrometeors were obtained through digital in-line holography (Ramelli et al., 2020), providing high-resolution size distribution data of both cloud ice and liquid. The combination of model and in-situ observations allows for assessment of cloud microphysics at scales ranging from hundreds of kilometers down to a few meters.

The overarching goal of this thesis is to assess the spatial distribution of cloud ice and supercooled liquid within mixed-phase clouds during this case study. The spatial scales of single-phase pockets (conditionally mixed cloud) and mixed-phase pockets (genuinely mixed cloud) will be quantified in both the modelled and observed cloud field to evaluate the degree of which mixed-phase clouds are actually mixed. Through this process, the thesis will attempt to address the three main research questions summarized in Box 1 below.

- **Research Question 1**
  How often and under which conditions does a state-of-the-art NWP model produce genuinely mixed-phase clouds, and how well does it compare to in-situ observations?
- **Research Question 2**
  How does the model simulate the spatial distribution of cloud phases in an Arctic mixed-phase cloud, and what is the sensitivity to INP and cloud droplet concentrations and parametrizations of SIP?
- **Research Question 3**
  How homogeneous are observed cloud droplet size distributions on the meter- to kilometer-scale and how much are they influenced by the presence of ice?

1.1 Outline

The rest of the thesis is organised as follows:

Chapter 2 covers the main theoretical background relevant for this thesis.

Chapter 3 gives a brief description of the setting and location of the analysed case study.

Chapter 4 introduces the instruments from which measurements were collected, as well as the methods used for analysis of the observational data.

Chapter 5 gives a description of the model, experiment setup, and the algorithms developed for analysis in this thesis.

Chapter 6 presents the general results of the modelling study from a statistical point of view.

Chapter 7 features the results following the implementation of a new algorithm to quantify the spatial intermittency of the cloud phases.

Chapter 8 evaluates the variability of cloud droplet size distributions through application of machine learning techniques.

Chapter 9 yields a summary of the key points put forth in the thesis, and presents suggestions for further work on the subject.

Appendix A features results and sensitivity tests from validation of the machine learning algorithm used for size distribution analysis.

Appendix B gives a statement on the availability of the computer code developed for analysis in this thesis.

CHAPTER 2

Theory

2.1 Cloud microphysics

This section provides a brief description of the cloud physics background relevant to this thesis, including droplet and ice crystal formation, the Wegener-Bergeron-Findeisen process and secondary ice production.

2.1.1 Droplet nucleation and growth

Clouds are formed when air parcels containing water vapor are brought beyond saturation such that condensation into liquid droplets becomes the most stable energy state of the water. It can be shown that the equilibrium vapor pressure e_s over a planar surface of liquid water yields the (simplified) temperature-dependent Clausius-Clapeyron equation (Clapeyron, 1834; Clausius, 1850);

$$\frac{de_s}{dT} = \frac{L_v e_s}{R_v T^2} \tag{2.1}$$

where T is the air temperature, L_v is the specific latent heat of evaporation of water and R_v is the ideal gas constant of water vapor. Equation 2.1 can be solved explicitly for e_s analytically;

$$e_s = e_{s0} \exp\left(\frac{L_v}{R_v T_0} - \frac{L_v}{R_v T}\right) \tag{2.2}$$

where e_{s0} and T_0 are the saturation vapor pressure and temperature at some reference point (Ambaum, 2020). It follows that the amount of water vapor the air can hold before reaching its theoretical maximum increases exponentially with temperature. The strong dependence on air temperature explains why most clouds form only when sufficient cooling occurs, for instance through expansion following adiabatic lifting or through emission of longwave radiation.

Liquid water exhibits strong intermolecular hydrogen bonds, which gives rise to a surface tension energy barrier that must be overcome before water droplets can form from the vapor state as given by Equation 2.3 (Wallace and Hobbs, 2006),

$$\Delta E = 4\pi R_d^2 \sigma - \frac{4}{3}\pi R_d^3 nkT \ln \frac{e}{e_s} \tag{2.3}$$

where R_d is the radius of the droplet, σ is the interfacial energy between vapor and liquid, n is the concentration of water molecules and k is the Boltzmann constant. The first term is given by the surface tension of water, and is thus always positive and proportional to the total surface area of the droplet. The second term expresses difference in internal energy per water molecule of the liquid and gaseous phase, and is therefore dependent on the water vapor saturation. Solving for the critical droplet radius $R = R_c$ at which point the droplet is in (unstable) equilibrium with its surroundings yields the Kelvin equation (Thomson, 1872);

$$R_c = \frac{2\sigma}{nkT \ln \frac{e}{e_s}} \tag{2.4}$$

The Kelvin equation determines the droplet radius after which any further increase in supersaturation will cause the droplet to grow unconditionally and is said to become activated. The energy barrier is in practice large enough that water droplet embryos are never able to develop and grow homogeneously into cloud droplets in Earth's atmosphere (Wallace and Hobbs, 2006). The presence of foreign particles in the air acts to lower the critical supersaturation threshold and cloud droplets consequently form through heterogeneous nucleation in natural clouds. Cloud Condensation Nuclei (CCN), a small subset of typically soluble aerosol, serve as host surfaces on which water vapor can condense. Once dissolved, the dissociated ions of the soluble substance act to strengthen the molecular bonds within the liquid and therefore lower the saturation pressure over the droplet's surface. The fractional amount of lowered saturation water vapor pressure over the solute compared to a droplet of pure water $\frac{e'}{e}$ is given by Raoult's equation (Raoult, 1889);

$$\frac{e'}{e_s} = \Big[1 + \frac{imM_w}{M_s(\frac{4}{3}\pi R^3 \rho' - m)}\Big]^{-1} \tag{2.5}$$

where m is the mass, ρ' the density and i the Van't Hoff factor of the dissolved CCN, and M_s and M_w are number of moles of the CCN and pure water, respectively. Combining Kelvin's equation 2.4 and Raoult's equation 2.5 one obtains the Köhler equation (Köhler, 1921) describing the growth of a water droplet containing soluble substances;

$$\frac{e'}{e_s} = \exp(\frac{2\sigma'}{n'kTR})\Big[1 + \frac{imM_w}{M_s(\frac{4}{3}\pi R^3 \rho' - m)}\Big]^{-1} \tag{2.6}$$

The strong impact of CCNs on the activation and growth of cloud droplets means that the CCN concentration (CCNC) of the environment determines many cloud properties. Higher CCNC, such as typically found in continental air masses as compared to maritime air (Pruppacher and Klett, 1997), lead to more activated droplets and a stronger competition for water vapor (Lamb and

Verlinde, 2011). As a result, polluted air masses typically facilitate clouds with smaller and more numerous droplets compared to cleaner air masses, which can have significant implications for the cloud albedo, lifetime, and precipitation rates (eg. Albrecht, 1989; Barthlott et al., 2022; Lohmann, Lüönd et al., 2016; Twomey, 1974).

2.1.2 Ice nucleation and growth

Analogous to the Clausius-Clapeyron equation (Eq. 2.1) one can obtain explicitly an equation for the saturation ratio for supercooled water in the presence of ice (Ambaum, 2020);

$$\frac{e}{e_i} = \exp\left(\frac{L_{m0}}{R_v T_0} - \frac{L_m}{R_v T}\right)\left(\frac{T_0}{T}\right)^{\frac{c_{pl}-c_{pi}}{R_v}} \tag{2.7}$$

where L_m is the latent heat of melting and c_{pl} and c_{pi} the isobaric heat capacity of liquid and ice, respectively. As with liquid droplets, the nucleation of ice particles requires some excess energy barrier to be overcome, even if the solid state is thermodynamically favored, related to the rearrangement of its molecules into a lattice. Similar to cloud droplets ice particles are theoretically capable of nucleating homogeneously and directly from the vapor phase, but it requires temperatures below -65°C and supersaturations above approximately 1000% (Houze, 2014) and so does not occur naturally in the atmosphere. Instead homogeneous nucleation of ice occurs indirectly via the liquid phase through freezing of an already condensed cloud droplet. Theoretical descriptions of ice growth are more complicated than for droplets due to the irregular habits of ice crystals. One common simplification is to instead consider a smallest inscribed sphere of radius R_i around the crystal. Its properties are then modified by two habit-dependent and often empirically determined constants α_i and β_i to correct for irregularities in its volume and surface area, respectively. Similar to Kelvin's Equation (Eq. 2.4), the critical crystal radius for ice nucleation from the liquid phase is then given by Equation 2.8;

$$R_i = \frac{2\beta_i \sigma_i}{\alpha_i n_i k T \ln \frac{e_s}{e_{si}}} \tag{2.8}$$

Growth of an ice embryo to the critical nucleation radius requires temperatures below -38°C when freezing occurs homogeneously (Pruppacher and Klett, 1997), but ice nucleation may be initiated heterogeneously at much higher temperatures with the aid of ice nucleating particles (INPs) (Kanji et al., 2017). Heterogeneous nucleation is typically subdivided into three modes of action (Vali et al., 2015):

1. Deposition nucleation – nucleation occurring directly from the vapor phase as water vapor deposits onto the INP.
2. Freezing nucleation: immersion freezing – nucleation occurring from the liquid phase if the INP is already collected by the water droplet and ice starts to grow around it.

3. Freezing nucleation: contact freezing – nucleation occurring upon contact when an already existing droplet collides with an INP.

The main characteristic of a good INP is typically that to some extent it mimics the structure of the ice lattice and that it is virtually insoluble in water, such as some mineral dusts. Ice nucleation has been hypothesized to occur at impurities on the surface of the INP at so-called ice active sites (Kiselev et al., 2017; Pruppacher and Klett, 1997). The probability of nucleation by the INP then scales with the size of the particle, as the probability of the presence of an ice active site is greater (Kanji et al., 2017). However, ice nucleating activity has also been discovered in soluble and biological macromolecules from sources including fungal species, bacteria and pollen (Pummer et al., 2015). Studies suggest that ice nucleation at temperatures in the range -38 to -15°C is dominated by mineral dusts whereas freezing at higher temperatures requires the presence of INPs from biological sources (Murray et al., 2012). The number of particles that can act as INPs in a given volume of air is therefore first and foremost a function of air temperature as INPs become active at different sub-zero temperatures (DeMott et al., 2010).

The INP number concentration (INPC) is typically several orders of magnitude smaller than the CCNC, and varies greatly both spatially and temporally (Wallace and Hobbs, 2006). The production of ice in clouds can therefore be severely inhibited when the cloud temperature is above the homogeneous freezing range and few INPs are present. INPs are also efficiently removed from the atmosphere because of their ability to facilitate precipitation via ice nucleation. The lifetime of an efficient INP is therefore limited and the INPC is strongly dependent on the proximity to its sources (Leck and Svensson, 2015).

2.1.3 The Wegener-Bergeron-Findeisen process

Combining Equations 2.2 and 2.7 one can obtain the difference in saturation vapor pressure with respect to liquid e_{sl} and ice e_{si} shown in Figure 2.1. The saturation vapor pressure is greater over a plain surface of supercooled liquid than over ice for any given temperature below the melting point due to the stronger molecular bonds in the solid lattice. This difference in vapor pressures means that liquid droplets may evaporate and ice crystals grow simultaneously for certain relative humidities and temperatures, through what is called the Wegener-Bergeron-Findeisen (WBF) process (Bergeron, 1928; Findeisen, 1938; Wegener, 1911).

It is now known that the majority of precipitation reaching the surface originates from the ice phase (Mülmenstädt et al., 2015), and the WBF process, where ice grows rapidly at the expense of liquid, has long been proposed as a major contributor (eg. Findeisen, 1938; Lau and Wu, 2003; Wallace and Hobbs, 2006). By definition, the WBF process occurs only when $e_{sl} > e > e_{si}$ as the liquid and ice will both evaporate (grow) for vapor pressures above (below) the given range (Korolev, 2007). Korolev and Mazin (2003) showed that vapor pressures in this range can be related to updraft velocities using quasi-steady supersaturation approximations for mixed-phase clouds. The updraft must be strong enough (positive upwards) for ice to grow but not strong enough that a positive supersaturation with respect to liquid is sustained. Thus the updraft u_z,

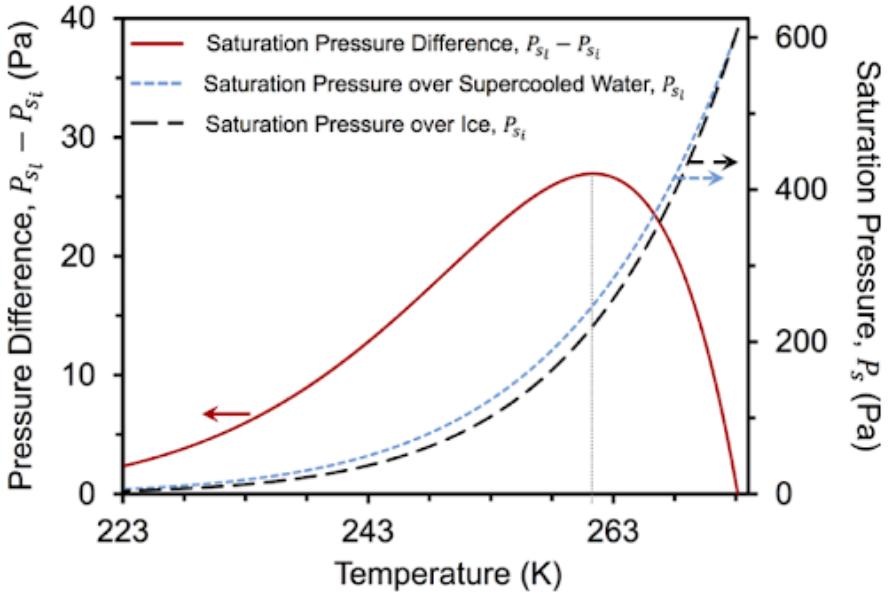

Figure 2.1: Temperature dependence of saturation vapor pressures in Pa over a plain surface of liquid water (light blue, dashed), ice (dark blue, dashed) and the difference between them (red, solid). Arrows indicate the associated ordinate axis. Figure from Nath and Boreyko (2016).

in combination with concentrations of ice and liquid, yields boundary conditions for when WBF may occur, which is useful for parameterizations of WBF in numerical weather and climate models (Storelvmo and Tan, 2015).

2.1.4 Secondary ice production

From field observations it has long been evident that there can be large discrepancies of several orders of magnitude between the observed ice crystal number concentrations (ICNC) and INPC in clouds (Auer et al., 1969; Kanji et al., 2017). The main mechanism proposed to bridge the gap from INPC to ICNC is secondary ice production (SIP), describing the process of producing new ice particles from already nucleated ice (Ladino et al., 2017).

Several SIP pathways have been proposed, of which *rime splintering*, or the Hallett-Mossop process (HM; Hallett and Mossop, 1974), has received the most attention (Field et al., 2017). HM is proposed to occur when a supercooled liquid droplet collides with a much larger rimed ice crystal causing the droplet to freeze and shatter, producing several new ice splinters that may subsequently grow. Physical descriptions remain elusive but the process has been well supported by in-situ observations and laboratory studies indicating that HM is most active in the temperature range of -8 to -3°C (Field et al., 2017).

Droplet shattering (DS) is another proposed mechanism for SIP wherein a large drizzle-sized liquid droplet nucleates from the outside for example via contact nucleation. As the outside freezes and expands, pressure builds in the still

liquid interior of the droplet causing it to eventually break and emit several new ice particles into the interstitial air (Keinert et al., 2020; Lauber et al., 2018). While the process is efficient in convective cloud regimes or in orographically induced clouds (Georgakaki et al., 2022; Lauber et al., 2021) where large frozen drops can be suspended in air for longer, previous modelling studies have found it to be less substantial in polar stratiform clouds (Fu et al., 2019; Sotiropoulou et al., 2020).

SIP has also been found to occur without the presence of liquid through *collision fragmentation*, also referred to as *mechanical breakup* (BR), which may occur when two larger ice crystals collide and break up as described in Vardiman, 1978. BR may operate at a wide temperature range with a maximum efficiency around $-15°C$ (Mignani et al., 2019), and is proposed to be particularly important when temperatures are too low for HM to occur (Sotiropoulou et al., 2020).

2.1.5 Some Common Cloud Measures

Mass concentrations of cloud water are usually quantified through the **mixing ratio** parameter q, which is computed as the mass ratio of ice, liquid, or both, to air. With a given dry air density ρ_d one can also compute the volume of air to get water contents (WC) in units of g/m^3. The WC of ice and liquid defines the **ice** and **liquid water content** (IWC, LWC, respectively) of the cloud, respectively. The sum of IWC and LWC is defined as the (condensed) **total water content** (TWC). **Ice** and **liquid water path** (IWP, LWP, respectively) are obtained through vertical integration of IWC and LWC, respectively. They yield a measure of the total ice or liquid cloud water content in a column of air, which is particularly relevant for the radiation balance.

Ice water fraction (IWF) is a commonly used measure of the balance between frozen and liquid water in clouds. It is calculated according to Equation 2.9 as the ratio of IWC to TWC;

$$\mathrm{IWF} = \frac{\mathrm{IWC}}{\mathrm{TWC}} \tag{2.9}$$

Depending on the focus of the study, the inverse version of Equation 2.9 defined by taking the ratio of LWC to TWC is also commonly used, and is referred to as the supercooled liquid fraction (SLF). This thesis will refer to IWF as the main quantity for ice-liquid-balance in the following analysis.

PART II

Methods

CHAPTER 3

Case Study Background

3.1 The NASCENT campaign

The data in this thesis was collected during the Ny-Ålesund Cloud Experiment (NASCENT; Pasquier, David et al., 2022), which was conducted between September 2019 and August 2020 in Svalbard, Norway. During the campaign, atmospheric measurements regarding aerosols, clouds, radiation, and meteorological properties were performed at five different sites around Ny-Ålesund. Ny-Ålesund is a former mining town now dedicated to Arctic research on the west coast of Spitsbergen. The town is situated at the South-Western coast of Kongsfjorden with several steep mountains surrounding it that significantly influences the observed wind directions close to the surface (Figure 3.1). Homogenized wind profiles obtained from radiosondes presented in Maturilli and Kayser (2017) show a clear effect of wind channeling through Kongsfjorden, with predominantly South-Easterly winds in the lowest 1000 meters contrary to the typically westerly winds higher up. Local weather patterns are also influenced by the proximity to large areas of open water off the south west coast of Svalbard providing a source of diabatic heating particularly in wintertime (Serreze et al., 2011).

The data used for analysis in this thesis, including INPCs from air samples and in-situ measurements of droplet and ice size distributions, are described in Chapter 4.

3.2 Meteorological setting

The case study considered in this thesis comprises measurements from 9-12 November, 2019 in the high Arctic that were made during the polar night with no direct incoming sunlight. The weather pattern during the period was shaped by a warm front that passed over Ny-Ålesund on 11 November shown in Figure 3.2. As the front passed over and left Ny-Ålesund in the warm sector behind it a pronounced air temperature increase was seen at the surface, with temperatures going from below -10°C on 10 November to relatively stable temperatures varying between -3° and 0°C on 12 November (Figure 3.3, upper panel). The ground was covered by snow throughout the period and the snow depth increased from about 11 cm on 9 November to 17 cm on 12 November. There were several snow showers on all days except 11 November. 12 November

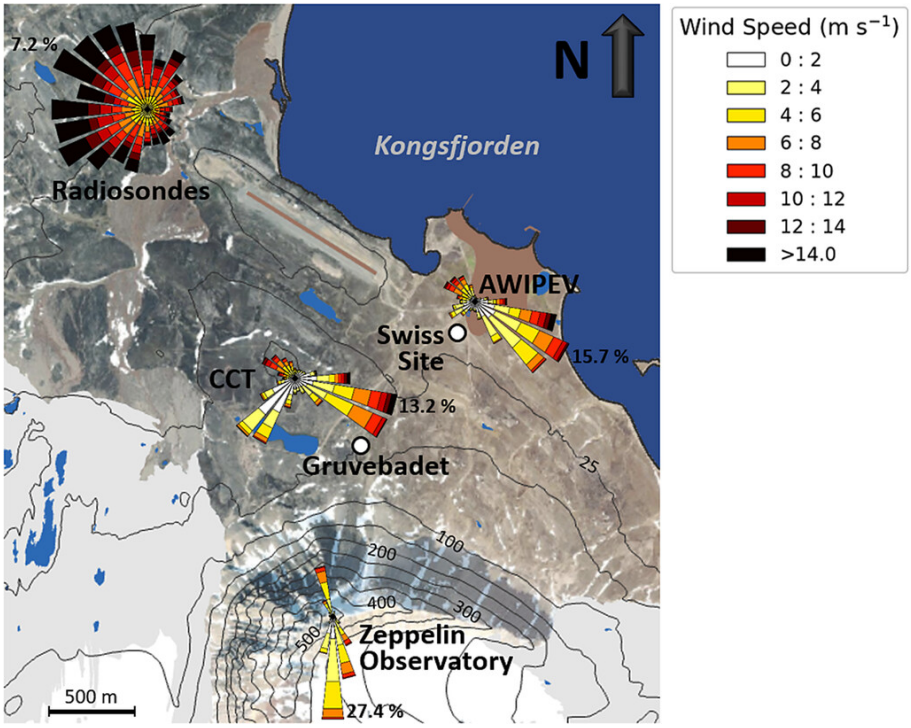

Figure 3.1: Topographic map of Ny-Ålesund with the measurement locations from the NASCENT campaign superimposed. Bar lengths and directions give the frequency of occurrence of each wind speed interval (colored) from each wind direction, respectively. Figure from Pasquier, David et al. (2022).

in particular saw many short but intensive showers, with a total of about 6 mm of liquid water equivalent precipitation adding about 4 cm to the accumulated snow depth.

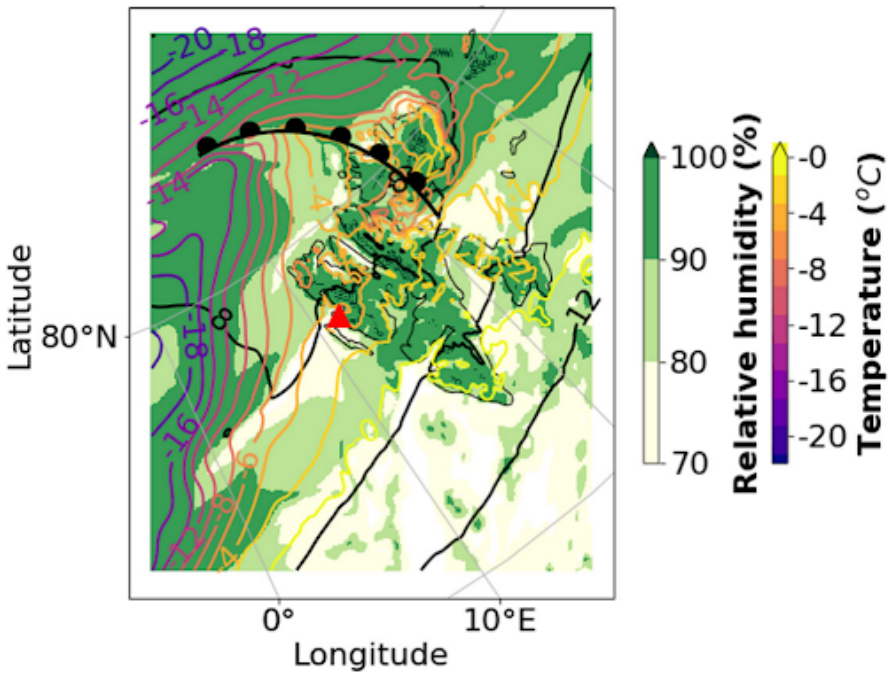

Figure 3.2: Weather map for Svalbard on 12 November 2019 at 06 UTC, showing the estimated location of the warm front relative to measurement location at Ny-Ålesund (red triangle). Also shown is relative humidity (green filled contours) and air temperatures at 1000 hPa (colored contours). Model data is from the MEPS weather maps (Hellmuth and Hofer, 2019), and figure is from Pasquier, David et al. (2022)

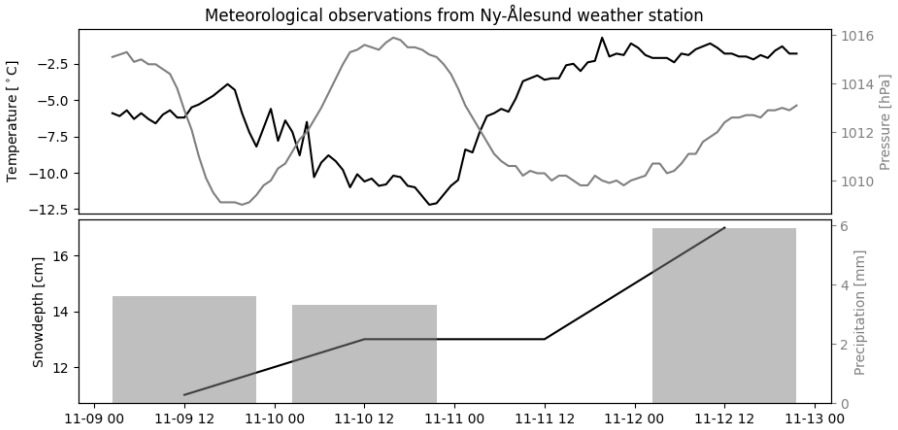

Figure 3.3: Weather observations from the Norwegian Met Office's meteorological station 'SN99910' at Ny-Ålesund during the case study. Upper panel shows hourly observations of air temperature at 2 meters above ground (black) and the surface air pressure reduced to sea level (grey). Lower panel shows daily snow depth (black) and 24-hourly accumulated precipitation (grey).

CHAPTER 4

Observational Methods

4.1 Instruments

4.1.1 Holographic imager

HoloBalloon (Ramelli et al., 2020) is a tethered balloon system designed to carry the HOLographic Imager for Microscopic Objects 3B (HOLIMO3B). HOLIMO3B provides high resolution droplet size distributions, which can be used to examine the spatial homogeneity of the cloud. It has a sample volume of approximately 18.6 cm^3 and uses digital in-line holography to size and count all of the hydrometeors larger than 6 microns within this volume. Due to the difference in shape between cloud droplets and ice crystals, all hydrometeors larger than 25 microns were classified as either cloud droplets or ice crystals using a convolutional neural network (Touloupas et al., 2020). The case study for this thesis took place on November 12 2019, when the HoloBalloon platform measured during three separate flights at altitudes up to about 850 m MSL. During these flights, the holograms were captured at a frame rate of 6 holograms per second.

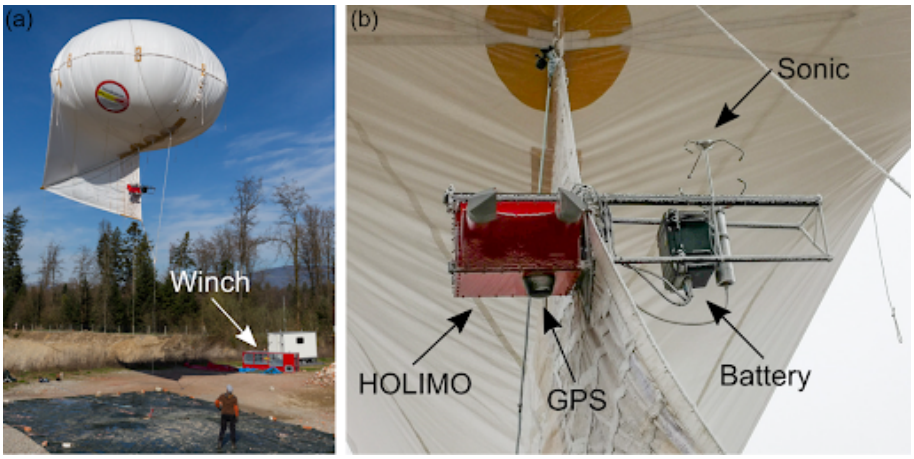

Figure 4.1: The HoloBalloon platform (a), with the HOLIMO3B instrument setup (b). Figure from Ramelli et al. (2020).

4.1.2 INP measurements

Air samples during the NASCENT campaign were taken with a high flow-rate aerosol-to-liquid impinger (Coriolis μ, Bertin Instruments, France). The impinger collects aerosols larger than 0.5 μm by rotating the air within a sterilized plastic cone with distilled water at high speeds, with a collection efficiency of 50%, 80% and 94% for 0.5, 2 and 5 μm particles, respectively (Wieder et al., 2022). It was active for one hour per measurement and operated at a flow-rate of 300 L min^{-}1, sampling a total volume of 18 m^3 each time. The INP concentrations within the air samples were measured with the Droplet Ice Nuclei Counter Zürich (DRINCZ) developed by David et al. (2019). DRINCZ consists of a back-lit temperature controlled ethanol bath, a submerged PCR tray slot, a USB web camera and a pump to keep the bath level constant (David et al., 2019). The sterilized PCR tray contains 96 wells each filled with 50 μL of the liquid sample and is covered by a sterilized plastic film to avoid contamination. The tray is lowered into the ethanol bath and the temperature is incrementally decreased from -2°C down to -32°C or until all wells are frozen, at a fixed rate of 1°C. The web camera takes a picture for every 0.25°C interval. As a well freezes the intensity of the light from the back-lit panel changes abruptly. By the end of the run, the two adjacent pictures that recorded the highest intensity difference for each well separately gives the freezing temperature of the given well. The INPCs (following Vali, 1971) for each temperature interval are then estimated based on the recorded freezing temperatures of all 96 wells.

4.2 The KS-clustering algorithm

The KS-clustering algorithm, henceforth referred to as the KSC algorithm, was first described by Allwayin et al. (2022) and developed specifically for data collected during the ACE-ENA campaign (Wang et al., 2022) with the aircraft-mounted HOLODEC instrument (Fugal et al., 2004). This thesis implements a revised version, which is adapted to fit holographic data from the NASCENT campaign. The data analysed in this thesis was obtained with the HoloBalloon instrument (Section 4.1.1), which stays at a roughly fixed location in the air and measures as the cloud passes through. This yields a spatial resolution dependent on the wind speed w and measurement frequency f, which in sum, is much higher than for research flights. For instance with the mean wind speed on November 12, 2019 of roughly 10 m/s, HoloBalloon yields a spatial resolution of 1.7 meters which is almost 18 times higher than during ACE-ENA.

The two major components of the KSC algorithm are the Kolmogorov-Smirov (KS) test and automated unsupervised clustering, which are described in detail below in Sections 4.2.1 and 4.2.2, respectively. The rest of this section covers details on the implementation of each component as well as the synthesis of the full algorithm and the choices made to get the results presented in this thesis.

4.2.1 The KS-test

The Kolmogorov-Smirnov (KS) test is a popular non-parametric statistical method that is frequently used to evaluate whether a given set of observations is plausibly derived from a known probability distribution (Kolmogoroff, 1941).

The two-sample variant of the test, which enables comparison of two independent sets of observations, is particularly advantageous as it does not require any a priori assumptions about the data (Young, 1977). The test entails the construction of the empirical cumulative distribution function (eCDF) for each dataset, which are then compared by means of the KS statistic k, measuring the maximum vertical distance between the two eCDF curves. Let C_a of size m and C_b of size n be the eCDFs of observations from sample a and sample b, respectively. Then k is given by Equation 4.1 (Gibbons, 1992).

$$k = \sup_x |C_a(x) - C_b(x)| \tag{4.1}$$

k thus falls between 0 and 1, whereby a high value indicates that the two samples are clearly different and a small value indicates a high degree of similarity. With Equation 4.2 one can compute the exact probability p of observing some KS-statistic $k_{m,n}$ larger than or equal to the observed k, under the null hypothesis H_0 that the two samples originate from the same parent distribution (Gibbons, 1992).

$$p = P(k_{m,n} \geq k) = 1 - \frac{A(m,n)}{\binom{m+n}{m}} \tag{4.2}$$

Here, $A(m,n)$ denotes the number of possible paths going from the point $(0,0)$ to (m,n) in a plane within the confines of a distance nk from the diagonal between the points. For large sample sizes, however, Equation 4.2 is computationally inefficient as the number of additions required for computing $A(m,n)$ grows (for fixed m) as $n^{\frac{3}{2}}$ (Hodges, 1958). Smirnov (1939) proved that for large sample sizes, p can be estimated by Equation 4.3, where $\sqrt{\frac{mn}{m+n}}$ is a normalizing factor for $k_{m,n}$ highlighting the dependence on sample sizes.

$$p = P(\sqrt{\frac{mn}{m+n}} k_{m,n} \geq k) = 1 - 2\sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2k^2} \tag{4.3}$$

It is common to define a significance level α that determines the grounds on which the null hypothesis is to be rejected or not. The k-value corresponding to $p = \alpha$ is referred to as the critical value k_c. Under the null hypothesis H_0 that the two samples originate from the same parent distribution, the p-value determines whether to keep H_0 if $k < k_c \Rightarrow p > \alpha$, or to reject it if the distributions are sufficiently dissimilar with $p < \alpha$.

The KS-test builds on the assumption that input observations are continuous. When applied to discrete observations, Noether (1963) showed that the test produces too conservative results, with critical values smaller than or equal to corresponding values for the continuous case. Since the droplet diameter observations analysed here are binned, two methods are proposed in this thesis to avoid the issue. The first proposed option, referred to here as the *FIT* method, is to fit each size distribution to a gamma curve prior to testing using the Maximum Likelihood Estate method from the `gamma.fit()` method in the `scipy.stats` module. Random samples are then drawn from the fitted gamma

distribution until the same number of droplets as before is obtained. This acts to smooth the eCDF of the observations creating artificially continuous samples based on the original binned data, but comes at the cost of introducing another source of uncertainty related to the fitting method.

The other proposed option, hereby termed the *NOFIT* method, is to use the KS-test statistic directly as the distance measure in the KS-matrix, instead of applying the p-value equation to produce a binary result of either failure or success. This avoids the p-value deficiencies under the discrete case, but the test may be less efficient at discriminating between samples as it becomes linearly dependent on the similarity between distributions. Both the *NOFIT* and the *FIT* method are tested and described in further detail in Appendix A, but the thesis is focused primarily on results achieved using the *NOFIT* approach.

4.2.2 Unsupervised classification

Unsupervised classification is a type of machine learning algorithm that does not require any prior labelling or 'human' knowledge of the data to be passed on (Marsland, 2014). The classifiers work by examining patterns or natural clusters in the input data based on various techniques and measures, and have the significant advantage over supervised learning algorithms, which rely on being fitted to labelled training data for initialization (Berry et al., 2020). The independence of labelling reduces selection bias as well as the number of assumptions made about the domain a priori, working in favor of more objective results. Unsupervised learners also rely in most cases on some initial subjective choices through the selection of a set of 'rules' for the classification, commonly referred to as hyperparameters. Common input clustering hyperparameters are measures such as the number of clusters expected, minimum cluster size, or the maximum distance two points can be apart to still be considered as part of the same parent cluster. The type of hyperparameters needed is determined by the type of clustering method to be applied, and so different types of algorithms are appropriate for handling different tasks depending on the properties of the data and the research questions to be addressed (Berkhin, 2006).

The *Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN; Ester et al., 1996) has the advantage of not requiring the number of clusters present in the results to be specified. It is therefore appropriate for cases where little or nothing is known about the structure of the data prior to classification. It originates from a group of unsupervised algorithms called density based classifiers, which have become increasingly popular for their ability to detect clusters of unequal sizes and arbitrary shapes (Amini et al., 2014). They also possess the ability to handle data containing noise, by classifying outliers that are not sufficiently similar to any surrounding points as not part of any class at all. DBSCAN utilizes only two hyperparameters, $\epsilon \in \mathbb{R}_{>0}$, a measure of the maximum size of the neighborhood defined around each point in the domain space, and `min_samples` $\in \mathbb{Z}_{>0}$, which determines the minimum number of samples within the ϵ-neighborhood around a given point required for a cluster to be defined around it (Ester et al., 1996).

Hierarchical Density-Based Spatial Clustering Algorithm with Noise (HDBSCAN; Campello et al., 2013) is an extension of DBSCAN which combines methods

from hierarchical clustering algorithms, which are able to find clusters of varying densities, with the DBSCAN advantages of identifying various shapes as well as not requiring a pre-determined number of clusters (Campello et al., 2020). At its simplest, HDBSCAN requires the selection of only one input hyperparameter 'minimum points', or K, as a measure of how conservative the algorithm should be in terms of required density thresholds surrounding the clusters. It avoids the use of ϵ by instead defining a *core distance* for each given point defined as the distance to its K'th nearest neighbor. Based on all computed core distances, a Minimum Spanning Tree (MST) is constructed such that all points in the domain are linked through minimizing the edge weight. The MST defines the basis for generating the cluster hierarchy, where each entry in the tree is sorted hierarchically based on the core distance weights (Campello et al., 2013). For practical purposes, HDBSCAN can also be initialized with another hyperparameter 'minimum cluster size', which sets how many samples a class must contain to be defined as its own cluster. This has the advantage of having an intuitive interpretation allowing the user to specify how many points are needed to yield a significant statistical basis for a cluster. Following the cluster hierarchy, HDBSCAN extracts the classes that are most 'stable' by iteratively removing branches of the MST (in decreasing order of distance) and recording how many iterations each class persists before dropping below the minimum cluster size. Thus, HDBSCAN ultimately returns the classes that are most independent of the distance metrics chosen (McInnes et al., 2017).

Here, the Python implementation in the HDBSCAN Clustering Library by McInnes et al. (2017) is applied. The major hyperparameters are `min_cluster_size` and `min_samples`. `min_cluster_size` determines the number of size distributions required to form a cluster, and was fixed to 10 in the results presented in this thesis. Under a mean wind speed of $10ms^{-1}$, this corresponds to approximately 17 meters of cloud, and was chosen as an attempt to find a balance between significantly large cluster sizes and cluster resolution. `min_samples` was found to be less crucial to the sensitivity of the results. Large values (much greater than `min_cluster_size`) tends to yield only one large cluster whereas low values (much less than `min_cluster_size`) returns mostly noise. Intermediate values were found to yield fairly stable results, and the parameter was left unchanged from its default value equal to `min_cluster_size` in the results presented here.

The HDBSCAN Clustering Library supports using a combination of DBSCAN and HDBSCAN through selection of an ϵ parameter, which in practice merges small clusters produced by HDBSCAN if they are closer to each other in the domain space than the threshold given by ϵ (McInnes et al., 2017). This can be included to avoid distinguishing between clusters of very similar properties. A lower threshold of $\epsilon = 0.001$ was used in this thesis, chosen after testing to yield a balance between cluster size distribution diversity and significance.

4.2.3 Algorithm methodology

The full KSC procedure is summarized in Figure 4.2 and elaborated below.

1. A minimum droplet number cutoff C_N is defined such that only holograms with a total droplet count above this threshold are considered for the

further analysis. This thesis follows the cutoff proposed in Allwayin et al. (2022) of $C_N = 0.7\overline{N_d}$, where $\overline{N_d}$ is the mean droplet count per hologram for the given segment. If $0.7\overline{N_d} < 100$ a minimum C_N of 100 is used instead to ensure a reasonable statistic for the size distribution.

2. All remaining holograms are randomly down-sampled to contain the same number of droplets equal to C_N. To reduce uncertainties related to resampling the procedure is repeated n_e times such that each hologram consists of n_e randomly down-sampled ensemble members containing C_N number of droplets. In this thesis, n_e was fixed to 20.

3. Each ensemble member in each hologram represents a size distribution of liquid cloud droplets in a binned format. In the FIT method, the discretized size distributions is smoothed by being fitted to a gamma distribution defined by a shape, scale and location parameter. This yields an artificially continuous set of diameter sizes, which is a prerequisite for the KS test. The algorithm is tested both with and without this step.

4. The KS matrix is now constructed by looping through each hologram and comparing to every other hologram, including itself. For row i of the KS matrix corresponding to hologram h_i in the defined cloud segment, each ensemble member in h_i is compared to every ensemble member in all other holograms. The entry on row i and column j in the KS matrix is then the average KS score between all ensemble members in hologram h_i versus all ensemble members in hologram h_j. The binary KS result based on the p-value equation (Eq. 4.2 or 4.3) is used for the *FIT* method, whereas the *NOFIT* method uses the KS-statistic directly. The KS matrix is now considered a distance matrix containing the distance between every pair of holograms in the 'KS space', where 'closeness' is indicated by a low average KS score.

5. The computed KS distance matrix is used as input to HDBSCAN for automatized clustering of holograms. The hyperparameters `min_cluster_size` and `min_samples` are determined based on the number of holograms included in the given cloud segment. HDBSCAN clusters holograms with similar distributions together based on closeness in the KS space.

The output from KSC are N_c separate groups of holograms where N_c is the number of detected classes. The average characteristic droplet size distribution is then calculated for each group individually, as well as statistics for ice water content, total water content etc. The algorithm has a total computational cost of the order $\frac{1}{2}n_H^2 * n_m^2$, where n_H and n_m are the numbers of holograms and ensemble members, respectively. The factor of one half appears due to commutativity of the two-sample KS test, making the KS matrix symmetric. Because of computational limitations, the high resolution severely limits the total segment lengths which can be analysed with the algorithm. Binning of holograms into 1 second time steps (6-by-6 holograms) was thus introduced to increase the segment lengths for analysis. This effectively increases the sampled hologram volume and produces much higher total droplet and ice crystal counts per distribution giving improved statistical significance.

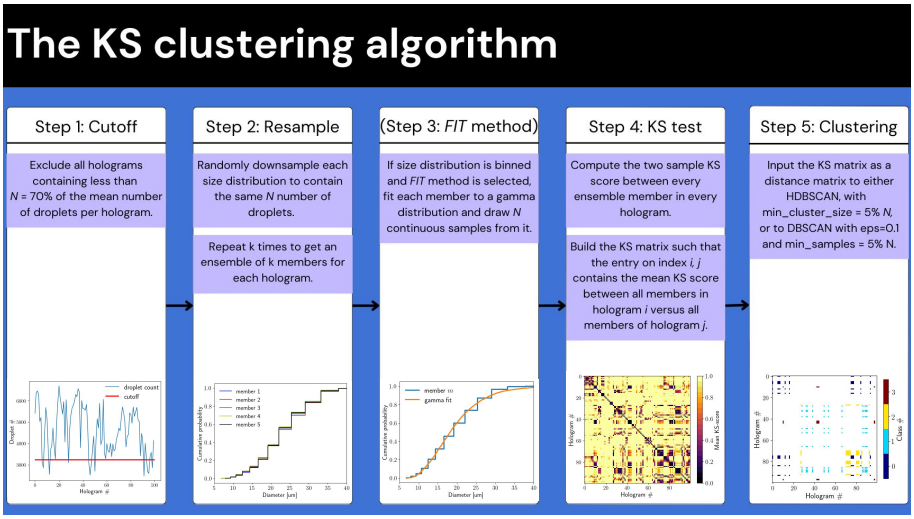

Figure 4.2: Flowchart of the KS clustering algorithm.

CHAPTER 5

Modelling Methods

5.1 Experiment setup

5.1.1 The Weather Research and Forecasting model description

The Weather Research and Forecasting (WRF; Skamarock et al., 2019) model is used to simulate this case study of mixed-phase clouds in Ny-Ålesund. WRF is a numerical weather prediction system developed by the National Center for Atmospheric Research, USA. Here, WRF is run with the Advanced Research WRF configuration version 4.0.1. The model is non-hydrostatic, fully-compressible and module-based, meaning it can be run with a wide range of physical parameterization schemes on scales ranging from large eddy simulations to the synoptic or global scale.

Here, WRF is initialized with input from the European Centre for Medium-Range Weather Forecasts (ECMWF) 5th generation reanalysis (ERA5) product. The model is nudged with the default value for nudging strength of $3 \cdot 10^{-4}$ s^{-1} every 6 hours at 06, 12, 18 and 00 UTC in terms of horizontal winds, temperature and specific humidity. The nudging process lasts 60 minutes and is gradually decreased towards the end of the period to avoid noise due to a shock as the nudging ends. The model time step is set to 30 seconds, and each simulation starts November 11 at noon, allowing 12 hours of spin-up before the start of analysis at midnight November 12. Boundary layer processes are handled by the Yonsei University scheme (Hong et al., 2006), and longwave and shortwave radiation is parameterized with the CAM scheme (Collins et al., 2004). Each simulation experiment is run on three nested domains of unequal size and resolution shown in Figure 5.1;

- Domain 1 (d01, outer) covers an area of 1800km (W-E) x 1350km (S-N) extending from the coast of Greenland in the west to Franz Josef Land in the north-east, with a spatial resolution of 15km. The model output is hourly.

- Domain 2 (d02, middle) extends an area of 515km (W-E) x 480km (S-N) covering most of the Svalbard archipelago and has a spatial resolution of 5km. The model output is hourly.

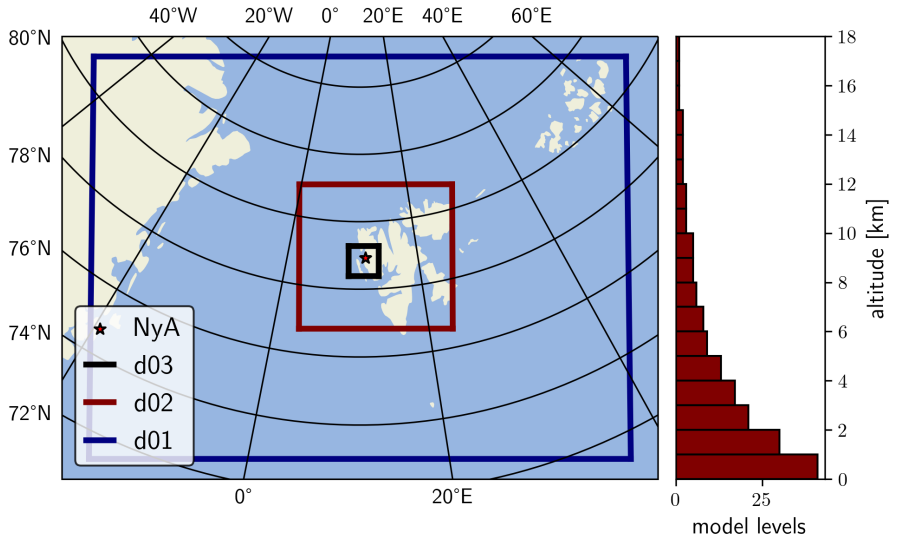

Figure 5.1: Extent of the three model domains (left panel) and the number of layers per kilometer of altitude (right panel). Star indicates the position of the measurement site in Ny-Ålesund.

- Domain 3 (d03, inner) covers an area of 100km (W-E) x 100km (S-N) on the west coast of Spitsbergen and has a spatial resolution of 1km. The output interval is 5 minutes.

The domains apply nesting such that the boundary conditions of the inner domains are provided by the next larger domain. The outermost domain d01 has boundary conditions supplied by ERA5. All three domains have the same vertical resolution with 172 layers extending up to a height of 17km. The levels are decreasing exponentially in density with altitude, and approximately half are located in the first 2500 meters where the low-level mixed-phase clouds are located.

5.1.2 Simulation experiments

WRF supports several different cloud microphysics parameterizations. Here, cloud microphysics are simulated with the double-moment bulk microphysics parameterization scheme following Morrison et al. (2005) and adapted in Morrison and Grabowski (2008), henceforth referred to as the 'Morrison scheme'. The Morrison scheme includes prognostic equations for both mass and number concentrations of cloud ice as well as the precipitating hydrometeor species rain, snow and graupel. Mixing ratios for cloud droplets are also simulated, but the cloud droplet number concentration (CDNC) is prescribed, and set to 250 cm^{-3} in the default version of the Morrison scheme. INPC is also prescribed, with deposition nucleation following Cooper (1986), and immersion freezing following Bigg (1953).

Ground based in-situ measurements of aerosols during the NASCENT campaign from both Gruvebadet, the Swiss site, and Zeppelin indicate that the default

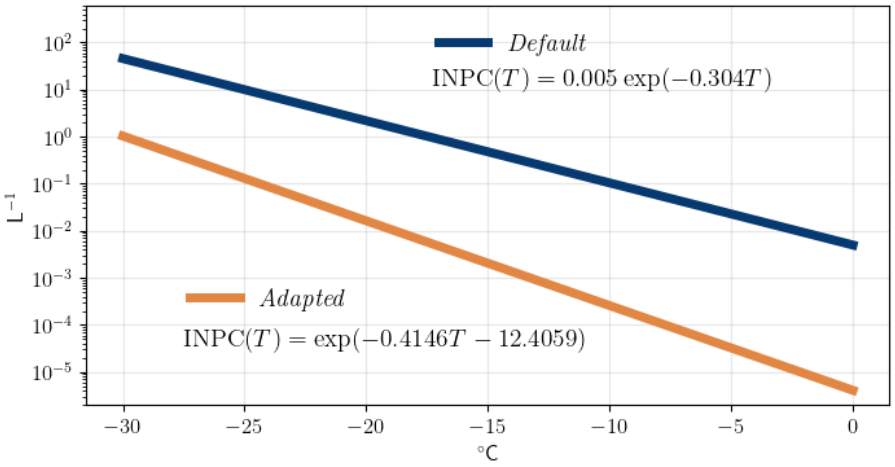

Figure 5.2: Temperature dependence of the INPC in the deposition mode as prescribed in the *Default* run (blue curve) and *Adapted* run (orange curve).

prescribed INP concentrations in the Morrison scheme are several orders of magnitude above realistic values for the Arctic at this time of year (Pasquier, David et al., 2022). The default value for CDNC of 250 cm^{-3} is also much higher than what is typically found in clouds formed in pristine environments such as the Arctic. Two separate experiments are thus run; one control run using the default Morrison scheme, and one adapted run where the INPC and CDNC are both constrained to measurements. The experiments are henceforth referred to as '*Default*' and '*Adapted*', respectively. The default configuration of the scheme uses a CDNC of 250 cm^{-3}, and an INPC (units L^{-1}) in the deposition mode dependent on the temperature T (units °C) as given by Equation 5.1 (Cooper, 1986).

$$\textit{Default}: \quad \mathrm{INPC}(T) = 0.005\exp(-0.304T) \tag{5.1}$$

In the simulations with adapted microphysics, CDNC was set to 10 cm^{-3} in agreement with in-situ measurements from HoloBalloon (Pasquier, David et al., 2022) on the same day. In the same run, the INPC (units L^{-1}), based on measurements from DRINCZ (see Section 4.1.2), was set as the temperature dependent exponential fit given by Equation 5.2 (Pasquier, David et al., 2022).

$$\textit{Adapted}: \quad \mathrm{INPC}(T) = \exp(-0.4146T - 12.4059) \tag{5.2}$$

The degree to which INPC is lowered going from *Default* to *Adapted* is dependent on temperature, and ranges from two orders of magnitude at about -25°C to more than three orders of magnitude close to 0°C. The full relation for temperatures between −30 and 0°C is shown in Figure 5.2.

HM is the only SIP mechanism included in the default configuration of the Morrison scheme. In the WRF implementation, HM is initiated when the

Table 5.1: Critical mixing ratios in units kg/kg of the default implementation of HM in the Morrison scheme and the modified thresholds applied in these experiments.

	Snow			**Graupel**		
	q^*_{snow}	q^*_{cloud}	q^*_{rain}	$q^*_{graupel}$	q^*_{cloud}	q^*_{rain}
Original	10^{-4}	$0.5 * 10^{-3}$	10^{-4}	10^{-4}	$0.5 * 10^{-3}$	10^{-4}
Modified	10^{-8}	10^{-6}	10^{-6}	10^{-8}	10^{-6}	10^{-6}

temperature is within the HM range of -8 to -3°C, and the following two conditions are met simultaneously;

1. Snow (graupel) mixing ratio q_{snow} ($q_{graupel}$) must be greater than some critical mixing ratio q^*_{snow} ($q^*_{graupel}$), and,
2. Cloud droplet mixing ratio $q_{cloud} > q^*_{cloud}$ or rain mixing ratio $q_{rain} > q^*_{rain}$.

Even though parts of the cloud in the case study were within the HM temperature range, the default critical mixing ratios were found to be too high for secondary ice to be produced in the default scheme. As an attempt to improve the agreement between simulated and observed ice crystal number concentrations, another set of simulations is included in the analysis, henceforth referred to as *Default+SIP* and *Adapted+SIP*. In order to increase the production of secondary ice in *Default+SIP* and *Adapted+SIP* the critical mixing ratios for HM to occur are significantly reduced, as summarized in Table 5.1. This adjustment is justified by the argument that parameterizations are developed based on laboratory experiments with varying cloud droplet size distributions that are not necessarily representative of realistic cloud water and ice mixing ratios (e.g. Hallett and Mossop, 1974; Mossop, 1978). Yet the thresholds for the initiation of the HM process are implemented in models for continental and maritime cloud droplet size distributions, which are significantly different than what is observed in the pristine Arctic environment (Field et al., 2017). The low CDNC in the Arctic allows for the same width in cloud droplet size distribution with a much lower liquid mixing ratio, potentially allowing for HM initiation at lower mixing ratios thresholds compared to less pristine environments. Furthermore, additional SIP processes are accounted for through the inclusion of parameterizations for DS as well as BR, following the WRF implementation by Sotiropoulou et al. (2021).

Thus, a total of 4 WRF simulations are analysed and compared in this thesis, as summarized in Table 5.2.

5.2 Analysis methods

Two new methods are developed for quantification of the spatial scales at which modelled clouds appear liquid, glaciated or homogeneously mixed. The

Table 5.2: Summary of the key differences between the WRF experiments analysed in this thesis.

	Default	***Default+SIP***	***Adapted***	***Adapted+SIP***
INPC	Unchanged	Unchanged	Modified	Modified
CDNC	250cm^{-3}	250cm^{-3}	10cm^{-3}	10cm^{-3}
HM	Unchanged	Modified	Unchanged	Modified
BR	No	Yes	No	Yes
DS	No	Yes	No	Yes

analysis can be applied to any level or part of the cloud, but in this thesis the algorithms are applied with respect to the cloud-top. Cloud-top analysis allows for comparison with remote sensing by satellite in potential future work, as the cloud-top is observable from space.

The modelled cloud-top phase and its variability are analysed from two different viewpoints – temporally and spatially, and the two new algorithms developed are hereby termed the 'temporal' and the 'spatial' algorithm, respectively. In the temporal analysis the cloud is described as it flows through each vertical column in the model, thus following the Eulerian description of fluid motion. The spatial description, on the other hand, looks at each timestep of the model output separately and quantifies how the cloud phase is distributed spatially, and thus 'tracks' single-phase pockets from the Lagrangian point of view.

The procedures of the two methods are described in detail below, following a brief description of the main calculations and post-processing applied before analysis.

5.2.1 Calculations

Output from the Morrison scheme in WRF are mixing ratios q (units kg kg^{-1}) per grid cell of two liquid hydrometeor classes, cloud droplets and rain, and three solid classes, ice, snow and graupel. Mixing ratios are converted to water contents (units g m^{-3}) via the air density ρ, which is calculated using virtual temperature T_v, air pressure P, and the ideal gas constant of dry air R_d following Equation 5.3 (Wallace and Hobbs, 2006).

$$\rho = \frac{P}{R_d T_v} \tag{5.3}$$

The relationship between water content WC and mixing ratio q is then given by Equation 5.4

$$WC = 1000\rho\, q \tag{5.4}$$

All classes are included in the calculation of IWF (Eq. 2.9), such that LWC and IWC are given by Equations 5.5 and 5.6, respectively.

$$LWC = LWC_{cloud} + LWC_{rain} \tag{5.5}$$

$$IWC = LWC_{ice} + LWC_{snow} + LWC_{graupel} \tag{5.6}$$

In the following analysis, a cloud volume is defined as mixed-phase based on the upper and lower thresholds for IWF of 0.1 and 0.9, respectively, following Korolev et al. (2003). A cloud fraction (CF) threshold is applied such that only grid cells with CF > 0.5 are included in the calculation of IWF. All other grid cells are given a non-numerical placeholder as the IWF. This is done to avoid out-of-cloud precipitation or volumes of very small WC to influence the subsequent analysis. The edges of the two inner domains are also clipped before analysis to avoid issues related to artefacts following the nesting procedure (Chavez and Barros, 2023). The 10 outermost grid cells from each edge is thus excluded from all analysis, reducing the total domain extents from 100x100km and 515x480km to 80x80km and 415x380km for *d03* and *d02*, respectively.

5.2.2 The temporal algorithm

For the temporal analysis of the cloud-top phase, each horizontal grid cell is treated separately and cloud-top pockets are counted as the cloud is advected with the wind through the given column of the model. The temporal algorithm described can be applied to quantify both mixed-phase and single-phased cloud-top segments, but the focus here will be on mixed-phase. The cloud-top is located based on the apical grid cell with a cloud fraction above 50%, and a mean value for the IWF is calculated based on the upper 7 layers of cloud below cloud-top. For each time step and horizontal grid cell the cloud-top is then characterized as mixed-phase based on the IWF thresholds stated in Section 5.2.1, or liquid (ice) if the IWF is below (above). The temporal algorithm then converts the data into N timeseries of length n_s with a continuously mixed-phased cloud-top, where n_s is the number of time steps in the segment and N is the total number of segments. Based on the mean cloud-top wind speed w_s in meters per second during the segment occurrence, and the temporal resolution of the model output r_T in seconds, the length of each mixed-phase cloud-top segment L_s in meters is computed according to Equation 5.7.

$$L_s = r_T n_s w_s \tag{5.7}$$

The temporal cloud-top segment lengths are compared between the two smallest domains d02 and d03 to investigate the influence of model resolution on cloud phase heterogeneity. The 5 minute output of d03 is coarsened to hourly to match the temporal resolution of d02 before the algorithm is run. The output of the algorithm for d03 is subsequently regridded by averaging 5 by 5 grid cells together to match the 5 times lower spatial resolution of d02.

5.2.3 The spatial algorithm

In the spatial analysis, or Lagrangian view, of the cloud-top phase distribution, the mixed pockets are counted spatially for each timestep t in the model output. At every t the mean cloud-top horizontal wind direction is calculated for the whole domain. A set of transects spanning the whole domain are defined and oriented along the mean cloud-top wind direction at the given time, each separated by 5 grid cells as exemplified in Figure 5.3. The cloud-top is located

along each of the transects, and ice and water contents are averaged across the upper 7 in-cloud layers at every point. The mean cloud-top IWF is calculated for all points and the cloud-top is classified as liquid, mixed or ice based on the same IWF thresholds of 0.1 and 0.9 as before. The spatial algorithm counts the number n_s of consecutive cloud-top grid cells classified as mixed-phase along the interpolated transect through the model domain. The length of the pockets are now determined by the spatial resolution along the transect r_{tr}, given by the horizontal cloud-top wind angle θ_w (0° N) and the horizontal spatial resolution of the model r_H, according to Equation 5.8.

$$r_{tr} = r_H \cos(\theta_w) \tag{5.8}$$

The length L_s of a given pocket is then given by Equation 5.9.

$$L_s = n_s r_{tr} \tag{5.9}$$

For each transect, mean and total pocket lengths, number of pockets and total length of the segment are recorded. Values are added together for all transects at each given timestep to yield a timeseries of pocket statistics across the whole domain. This process is repeated also for single-phased pockets with respect to both liquid or ice.

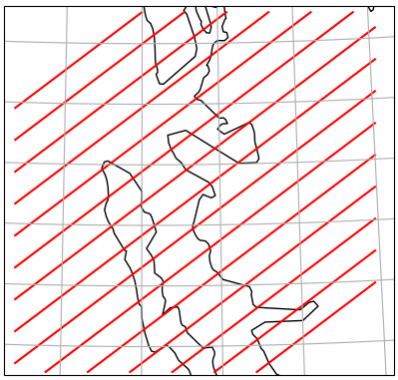

Figure 5.3: Illustration of the transects (red lines) at some given timestep for domain d03 along which the cloud-top spatial pockets are counted.

PART III

Results & Discussion

CHAPTER 6

Statistics of Cloud Phase Mixing

An important characteristic of mixed-phase clouds, one that is perhaps under-communicated in some scientific research, is the degree of homogeneity of mixing between liquid droplets and ice particles (Korolev and Milbrandt, 2022b). Whether or not a cloud should be classified as mixed-phase is in reality determined not only by how much ice and liquid it contains, but is also a question of the spatial and temporal scales that the cloud is evaluated at. Mixed-phase is often defined by the IWF interval between 0.1 and 0.9 (e.g. Korolev et al., 2003), but unless well mixed, this definition becomes largely dependent on the size and location of the considered cloud volume.

In this chapter, the degree of phase-mixing is evaluated statistically by quantifying how often, where, and under which conditions homogeneous mixing is produced in the modelled cloud field. When evaluating the homogeneity of the cloud field, the minimum cloud volumes considered are set by the model resolution of the smallest domain *d03*, with 1-by-1 km horizontally and between 20-40 m vertically, depending on altitude. Statistics of the cloud, occurring as either pure liquid, pure ice, or mixed-phase, are presented as a function of time, temperature, altitude and TWC. Lastly, the LWC, IWC and ICNC as simulated by the model runs over Ny-Ålesund are compared to the HoloBalloon measurements from November 12 to examine the agreement between model and observations for this case study.

6.1 Domain-averaged cloud phase statistics

The cloud phases are evaluated as a function of time, to examine the evolution of the cloud throughout the day of 12 November. From Figure 6.1, showing IWC, ICNC and IWF as averaged over *d03*, the higher INPC in *Default* is evident through its high ice concentrations, exceeding those of *Adapted* by about an order of magnitude on average. Without added SIP processes, the variations in both IWC and ICNC are very small. When additional SIP is included, however, the ice contents are at times increased by up to two orders of magnitude. Both *Default* and *Adapted* show a large impact from SIP toward the end of the simulation period, starting at approximately 18:00. *Default+SIP* and *Adapted+SIP* show a close agreement in both number and mass concentrations of ice during this high SIP event, but with a somewhat longer lasting effect in *Default+SIP*. *Adapted+SIP* predicts a strong impact from secondary ice

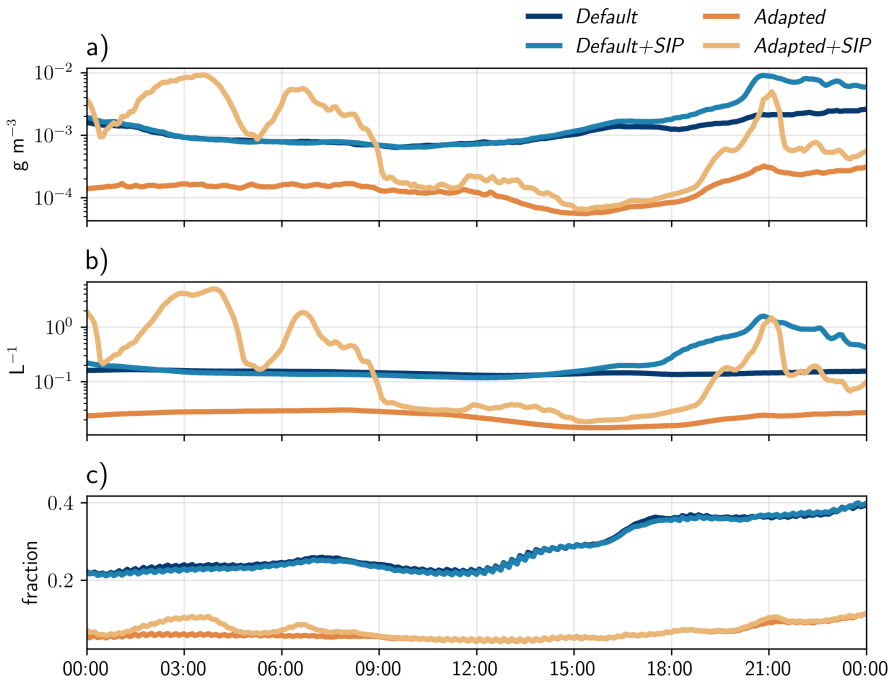

Figure 6.1: Simulated (a) IWC in g m^{-3}, (b) ICNC in L^{-1}, and (c) IWF for November 12, 2019, as averaged over the whole *d03*. Blue colors show the *Default* runs whereas orange colors show the *Adapted* runs. Lighter colors indicate the SIP runs.

also during the night and morning until about 09:00, which does not occur in *Default+SIP*. During this time, *Adapted+SIP* produces both higher IWC and ICNC than what is seen in the *Default* runs, despite the much lower INPC. It is unclear why this strong signal is simulated in *Adapted+SIP*, whereas no change is seen in *Default+SIP*. Examination of the SIP tendencies did not yield any immediate explanations (not shown). Ascertaining the source of this surprising increase in ICNC in the *Adapted+SIP* simulation is beyond the scope of this thesis, but should be investigated further in future modelling studies.

It is evident in Figure 6.1c that the *Default* runs have an average IWF that is 0.2 greater than that of *Adapted*. The lowering of INPC effectively removes the cloud from the mixed-phase zone, as defined earlier, when considering the average cloud spanning the whole domain of *d03*. Perhaps surprisingly, the inclusion of additional SIP processes does not substantially contribute to close this gap. Even though the additional SIP processes cause sudden increases in the IWC, almost no difference is seen in the IWF between simulations with and without SIP during those events. When the IWF remains constant while the IWC is increased, it follows from the definition of IWF that one of two situations are possible. Either, 1) the TWC is increasing proportionally with the IWC thus leaving the balance between them unchanged, or 2) the regions of increased SIP already had a high IWF before and thus an increase in ice mass cannot increase the IWF further. The first possibility can be supported by

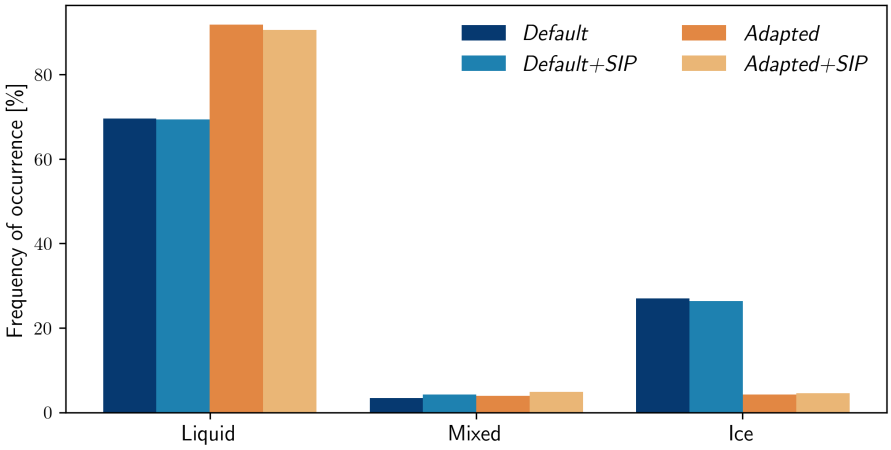

Figure 6.2: Frequency of occurrence in percent of ice, mixed-phase, and liquid out of all cloudy model grid cells on 12 November.

considering the HM process, the efficiency of which is dependent on the presence of both smaller and larger liquid droplets (Hallett and Mossop, 1974). Thus, having more liquid water in the cloud would potentially enhance the production of ice through HM, without large altercations to the IWF. On the other hand, the second possibility may be supported by enhancement of BR in high ice conditions, as the likelihood of ice-ice-collisions scales with the concentration of ice particles (Phillips et al., 2017; Vardiman, 1978; Yano et al., 2016)

It is worth noting that although the *Default* have a domain-averaged IWF well within the mixed-phase zone throughout the day, this is not at all the case when considering each grid cell separately as in Figure 6.2. Here it is evident that liquid is by far the most dominant of the phases. Following the definition of IWF < 0.1 for liquid clouds, liquid makes up at least 90% of the mass in approximately 70% of the grid cells in the *Default* runs and close to 90% in the *Adapted* runs. Mixed-phase, on the other hand, occurs only in about 5% of the grid cells for both the *Default* and *Adapted* runs. This once again highlights the importance of scale when defining mixed-phase clouds.

Figure 6.3 shows the probability density functions (PDFs) of the occurrence of ice, liquid, and mixed-phase versus temperature of the entire simulated cloud deck on November 12. Each PDF is here normalized to itself, first and foremost highlighting the differences in shape between the runs. However, it should be mentioned that a direct comparison in magnitudes between each of the runs is not achievable from this type of analysis, due to the different scaling factors applied to each. Considering the shapes, it is evident that the agreement between runs with and without SIP is much closer than between experiments with modified and default aerosol concentrations alone. From the probability density of any cloud phase shown in Figure 6.3d, the PDFs are very similar across the simulations, indicating that the experiments seem to agree on the temperatures required for cloud formation. Agreement is also high for the liquid phase, although shifted slightly to higher temperatures for the *Adapted* runs (Fig. 6.3c). The difference between *Default* and *Adapted* is

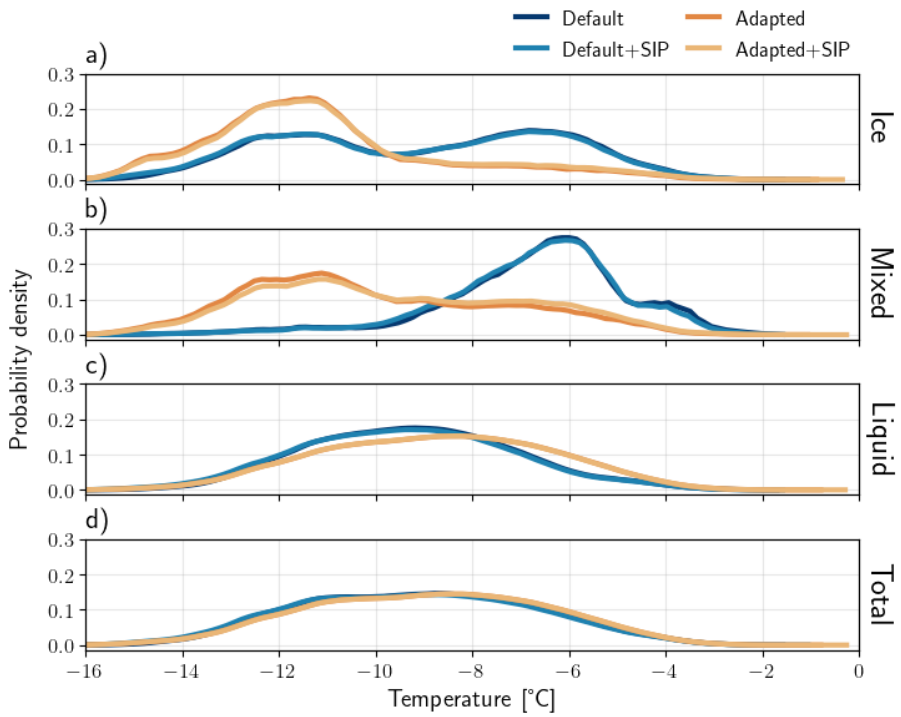

Figure 6.3: Probability density of occurrence of ice (a), mixed-phase (b), and liquid (c) versus temperature, each normalized to 1. Panel d) shows the probability density of any cloud phase versus temperature. Color-coded as in Figure 6.1.

greatest for the mixed-phase and ice cloud occurrences, and the *Adapted* runs both appear to require colder temperatures for ice to be present (Fig. 6.3a and b). The *Default* runs show a bimodal tendency for ice clouds, with peaks in occurrences at temperatures around both -12 and -7°C. The warmer peak is also evident for the mixed-phase clouds. The *Adapted* runs show some of the same bimodal shape for the mixed-phase, but not as much for ice clouds. The lower temperature peak could indicate the range where the ice is primary nucleated in the cloud during these simulations. The baseline of ICNC seen in Figure 6.1b stays fairly constant slightly above 0.1 L−1 for the *Default* runs and about an order of magnitude lower for the *Adapted* runs. From Equations 5.1 and 5.2 stated in Chapter 5 regarding the prescribed INPC in *Default* and *Adapted*, a temperature of -12°C would yield an INPC of approximately 10^{-1} and 10^{-3} L^{-1}, respectively. This corresponds very well particularly for the *Default* runs, and could be an indication of the ice production in the *Default* runs being INP-limited even with the much higher prescribed INPC. For the *Adapted* simulations, however, the baseline of mean ICNC is in fact greater than the assumed INPC at $T = -12$°C even without additional SIP processes included. This means that the ice particles must either have been transported from other, lower temperatures at which nucleation occurred, or that some ice particles may in fact have been produced through the unmodified HM process

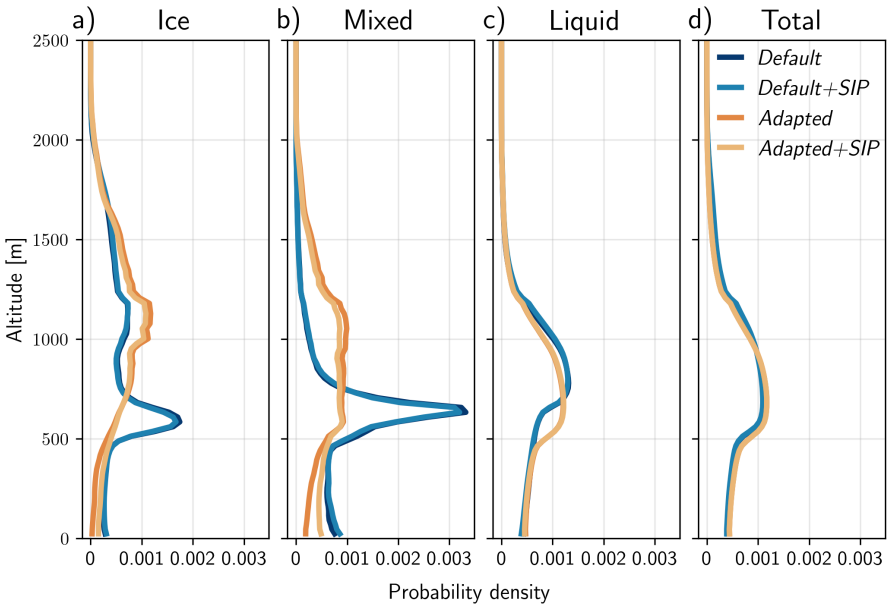

Figure 6.4: Same as in Figure 6.3, but as a function of altitude.

at higher temperatures in the *Adapted* run.

Figure 6.4 shows the probability density of phase occurrences as a function of altitude. As in Figure 6.3, all four experiments agree closely on the total vertical distribution of the cloud (panel d), whereas the largest differences are seen in the ice containing states. Where both *Default* runs exhibit a strong peak in probability of occurrence for mixed-phase clouds at an altitude of about 600 meters, the *Adapted* runs yield mixed-phase clouds distributed more evenly across a wider range of altitudes. For the *Default* runs, ice clouds primarily occur around the same altitude as mixed-phase clouds, whereas the occurrence of liquid clouds is predominantly at higher altitudes. This may be indicative of more liquid-topped mixed-phase clouds with precipitating ice falling through closer to the cloud base, consistent with previous studies of Arctic mixed-phase clouds (eg. Boer et al., 2009, Morrison et al., 2012). However, similar to Figure 6.3a, the ice clouds in the *Default* runs show a bimodal tendency with altitude, with a smaller peak in occurrences also evident above the dominant liquid cloud layer. This shows that the clouds in the *Default* runs have a predominantly liquid interior with a base of typically either mixed-phase or ice, but that the top can be either liquid or glaciated depending on additional factors. In the *Adapted* runs the structure is almost inverted with most of the pure ice or mixed-phase cloud occurrences located above the dominant liquid cloud altitude. This is likely due to the much lower ice concentrations in *Adapted*, making the ice-containing phases unable to compete with liquid in terms of the IWF. As such, the characterization of ice and mixed-phase in the *Adapted* runs appears to be mostly dependent on the LWC being very low, for instance close to the cloud edges. However, as was seen in Figure 6.2, the occurrence of mixed-phase and ice in the *Adapted* runs is very low, and so the statistical significance of

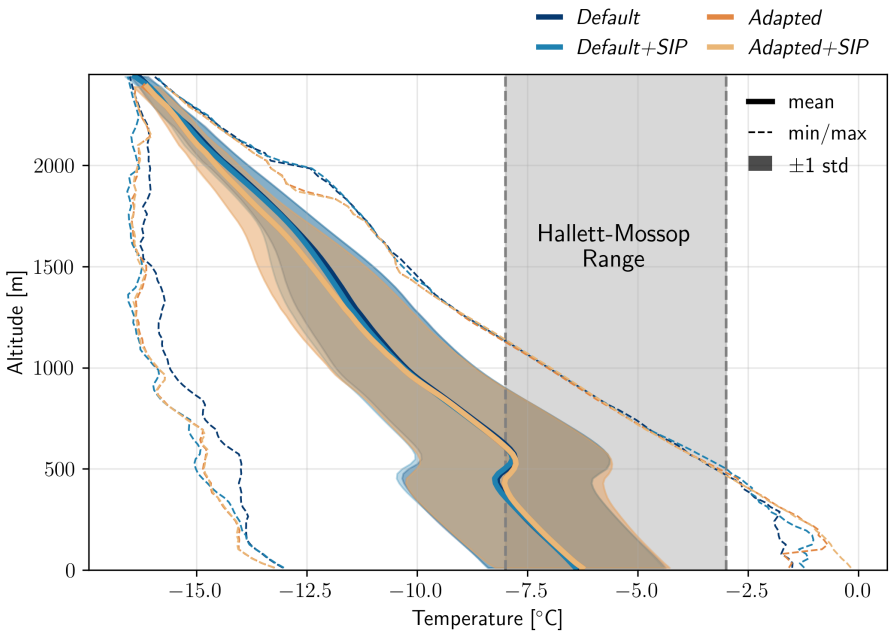

Figure 6.5: Solid lines show the simulated temperature profiles from the four experiments (colored) calculated as the mean for 12 November across the horizontal domain extent. Color-shaded indicate a distance of one standard deviation from the mean, whereas dashed lines show the span between minimum and maximum temperatures produced at each altitude per model. Grey, shaded region represents the temperature range where HM is activated in the model.

their respective vertical distribution is likely small.

The differences between runs with and without SIP appeared close to negligible in terms of determining the phase occurrences as a function of either temperature and altitude. One possible explanation for this can be found in Figure 6.5 showing the mean lapse-rates of each model simulation. All four experiments predict approximately the same average lapse rates with similar degrees of variability. Superimposed is also the temperature range over which HM is assumed active in the model. Only the part of the cloud located below ∼600 meters was on average within the HM temperature range. This could potentially explain the small increase in mixed-phase occurrences in the *Adapted* run at altitudes lower than ∼500 meters, although it would not be expected to have a large impact as this region was for the majority was below cloud-base. The experiments do, however, show large spread between the highest and lowest recorded temperatures across the domain and through the day, and so HM was theoretically possible up to an altitude of about 1200 meters. Any ice production in the model above this level must therefore have occurred through heterogeneous nucleation in the runs without added SIP processes, or possibly through BR and/or DS in the runs with SIP.

The effect of SIP processes are much more apparent when considering IWF as a function of TWC, as shown through a 2-dimensional probability density

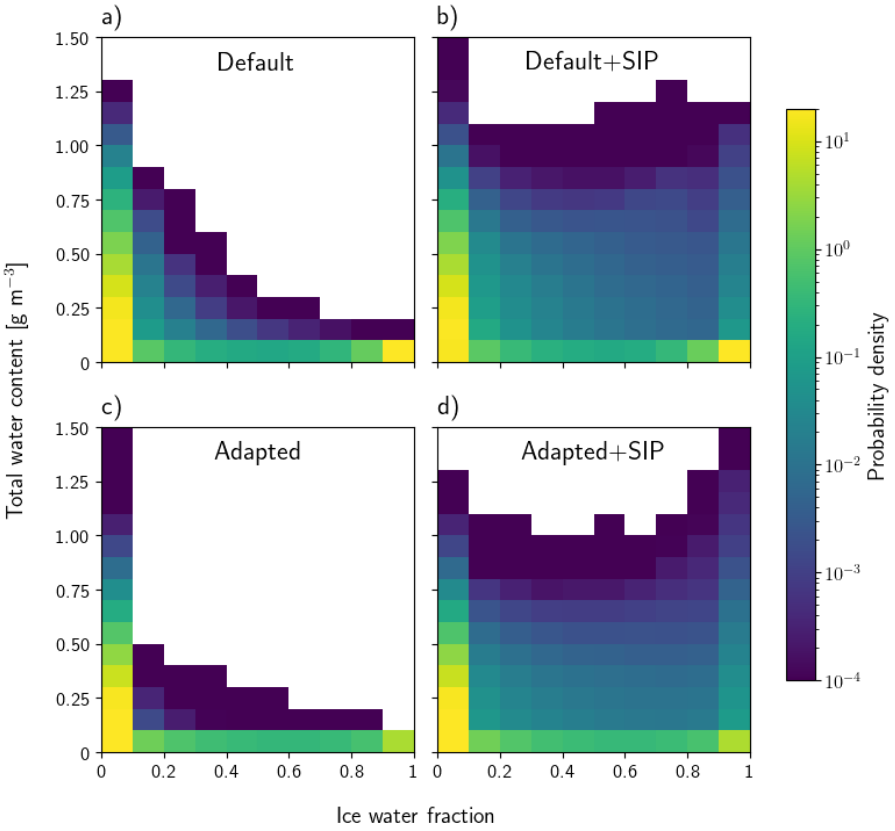

Figure 6.6: 2-dimensional probability density of IWF versus TWC. Yellow colors indicate a higher probability of occurrence.

histogram in Figure 6.6. Liquid clouds (IWF < 0.1) are the most frequent cloud type across most TWC in every experiment, with them most common at low TWC and becoming less frequent with increasing TWC. The effect of using representative aerosols is evident in that higher IWFs occur more frequently across a broader range of TWC in *Default* compared to *Adapted*. *Adapted* also has a much lower probability density of clouds at the glaciated stage of IWF > 0.9. This suggests that by reducing the concentration of INPs to observed values, the model produces complete glaciation of the cloud much less frequently. The effect of including additional SIP processes acts to dramatically increase the probability of occurrence of the ice-containing states at IWF > 0.1 over a much broader range of TWCs. Whereas glaciated clouds occurred only at very low TWC in the runs without additional SIP, IWF > 0.9 are observed also at much higher values following the inclusion of SIP. LWC is by definition small at the glaciated state. Therefore, it follows that the higher frequency of occurrence of IWF > 0.9 at high TWC is due to a large increase in IWC with SIP.

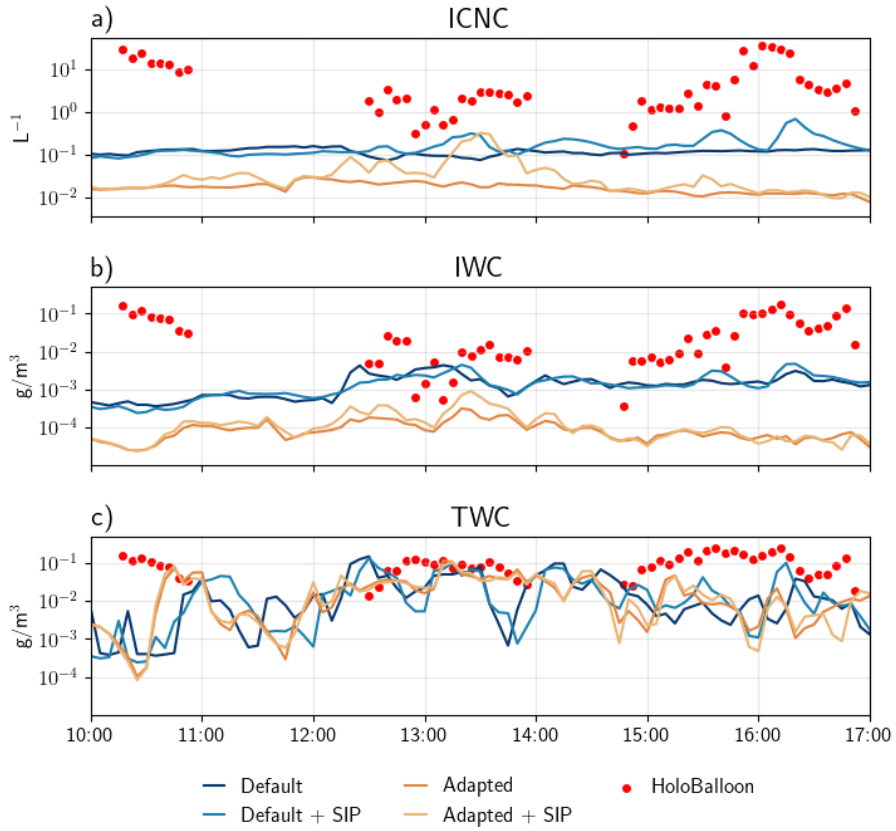

Figure 6.7: Simulated ICNC (a), IWC (b), and TWC (c) averaged over a box surrounding the location of HoloBalloon for each of the model experiments (colour lines). The confines of the box is defined horizontally by the four grid cells closest to the location of HoloBalloon, and vertically by the lowest 7 layers of the simulated cloud cover. HoloBalloon measurements (red dots) from the three flights on November 12 are averaged over 5-minute time steps to match the temporal output of the model.

6.2 Comparison with HoloBalloon

The ICNC, IWC, and TWC as simulated by each of the model experiments are compared to the measurements from HoloBalloon in Figure 6.7. Because the exact position of the simulated and observed cloud may be somewhat offset, particularly in terms of the cloud base altitude, a volume surrounding the position of HoloBalloon was defined for the comparison. This step involves averaging over a box confined horizontally by the four grid-cells closest to the HoloBalloon location, and vertically by the lowest 7 layers of cloud counting from cloud-base. Thus, only in-cloud grid-cells are included in the comparison, and the averaging acts to reduce the potential influence of small spatial or temporal offsets between the simulated and observed cloud.

All simulations severely underestimate the ICNC on average compared to

measurements from HoloBalloon, and whereas HoloBalloon shows large degrees of variability between the three separate flights, the model experiments exhibits an almost entirely constant ICNC throughout the day. The inclusion of SIP processes increases the variability somewhat, causing for instance a spike in the ICNC during the second flight around 13:30 in both *Default+SIP* and *Adapted+SIP*, and during the third flight in *Default+SIP*. However, even during these spikes, the ICNC as simulated by both *Default+SIP* and *Adapted+SIP* only get close to the observed values during a short period of the second flight (Fig 6.7a). The effect of SIP is less apparent in terms of the IWC, which is expected, as the processes act primarily to multiply ice particles, sometimes at the expense of their relative size such as through the BR mechanism. Still, there may be a contribution from SIP in terms of mass through the combination of other processes, such as WBF or deposition growth in general, if smaller and more numerous ice crystals are favorable for the growth rates. This effect is only apparent here during a shorter period around 13:00 in the *Adapted* run.

As seen in Figure 6.7b-c, simulated IWC is within the range of observational variability only during the second flight, through which liquid is the dominant phase in both the observed and simulated TWC. It is evident from the comparison that both *Default* runs achieve a much closer agreement with observations in terms of both ICNC and IWC from HoloBalloon than does *Adapted*. However, as is known from measurements of INPC, the default Morrison scheme assumes aerosol conditions which are unrealistic in pristine environments such as the Arctic. When the model is constrained to observed concentrations of INPs and cloud droplets, the *Adapted* runs show that both the IWC and ICNC are decreased by up to an order of magnitude, increasing the gap from observations even further. As such, it appears that the *Default* runs may perform better when evaluated purely in terms of number and mass concentrations of ice, but likely for the wrong reasons. This is an indication that additional and/or enhanced parametrizations of SIP processes are necessary in model simulations in order to reduce the discrepancy in simulated ice concentrations under realistic aerosol conditions for the Arctic.

CHAPTER 7

Cloud-Top Phase Variability

The cloud-top is a key factor in the dynamics of the mixed-phase cloud system, and its phase is of vital importance particularly for the lifetime of the cloud. Supercooled liquid is a strong driver of radiative cooling when present near the cloud top, enhancing in-cloud turbulence which in turn acts to resupply moisture to balance the constant loss of water to the ice phase (Morrison et al., 2012). When the cloud-top contains ice, however, the WBF process may cause a rapid depletion of supercooled liquid which has important implications for both lifetime, dynamics, precipitation rates, and radiative properties of the cloud (Korolev and Milbrandt, 2022b). Additionally, ice particles formed at the cloud-top may grow rapidly through the WBF process as this is typically the region where stratiform clouds exhibit the highest liquid water contents (Carey et al., 2008). Upon sufficient mass gain through deposition, the ice particles may achieve high enough fall speeds to sediment through the cloud's interior, potentially depleting more supercooled liquid water on its way through either riming, the WBF process, or both (Kumjian et al., 2014). Thus, the microphysical processes of the cloud-top have the potential to influence the cloud phase of the entire cloud, and the balance between ice and liquid at cloud-top is therefore often a good indicator of the state the cloud is in.

Here, results from the analysis using the two new methods for quantification of cloud-top phase variability are presented. Cloud-top pockets are quantified in terms of mean length, total extent, and number of occurrences, and the results are focused on the smallest domain *d03*. The chapter concludes with results regarding the influence of model resolution on pocket structure by investigating differences arising from applying the temporal algorithm on both *d02* and *d03*.

7.1 Spatial pockets

Shown in Figures 7.1, 7.2, and 7.3 are the results of applying the spatial algorithm on *d03* with respect to ice pockets, mixed-phase pockets, and liquid pockets, respectively. As the whole domain was mostly cloudy throughout 12 November, the three figures are closely related, since the occurrences of ice, mixed-phase or liquid cloud-tops are mutually exclusive if a cloud is present.

The first half of the day is characterized by a few small and isolated glaciated patches (Figure 7.1) in an otherwise liquid- or mixed-phase-topped cloud cover (Figures 7.2 and 7.3). After noon, an abrupt increase in both the length and

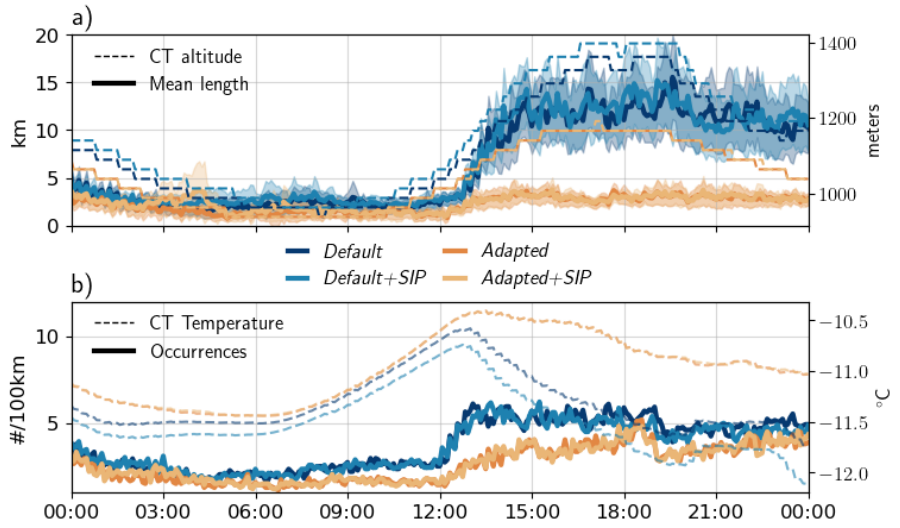

Figure 7.1: Spatial distribution of glaciated cloud-top pockets. Upper panel shows the mean length of each pocket (solid line, left axis) with its associated 95% confidence interval for the mean (shaded area, left axis), and the evolution of the mean cloud top altitude in model levels (dashed line, right axis). Bottom panel shows the number of transitions between glaciated and non-glaciated cloud tops (solid line, left axis) and the evolution of cloud-top temperature (dashed line, right axis). Line colors represent the different model experiments. Cloud-top is defined here as the apical 7 layers of the cloud.

number of glaciated pockets is seen in the *Default* simulations. With an average of 5 pockets per 100km and a mean length of each pocket extending to almost 15 km, it follows that approximately 75% of the examined cloud-top transects are glaciated after noon. Whereas a small increase in pocket numbers is seen also in the *Adapted* runs, the mean length of each glaciated segment here remains nearly constant and below 5 km throughout the day. Preceding the cloud-top glaciation in the *Default* runs at noon is a period where mixed-phase cloud-tops dominate (Figure 7.2). These are the only cloud-tops where the WBF process can act since both liquid and ice must be present. Due to the potential impact of WBF, the mixed-phase cloud-tops are inherently unstable as the ice may rapidly deplete the liquid and consequently glaciate the cloud-top. This is evident in both *Default* runs where the period of long mean mixed-phase cloud-top segment lengths around noon is transitioned into a state where glaciated cloud-tops dominate in a matter of only a few hours (Figure 7.1). Around noon, an abrupt change is seen also in the cloud-top altitude (dashed lines, panel a) and temperature (dashed lines, panel b). The *Default* runs predict a rapid rising of about 400 meters from an altitude of 1000m to 1400m ASL. The *Adapted* runs predict a much less pronounced cloud-top rise with only about half of what is seen in *Default*, causing the cloud-top to reach altitudes of about 1200 mASL. The cloud-top rising is preceded by a distinct increase in the cloud-top temperature seen in both *Default* and *Adapted*, starting shortly after 06:00 and leading up to noon. As the cloud-top ascends, the temperature is consequently

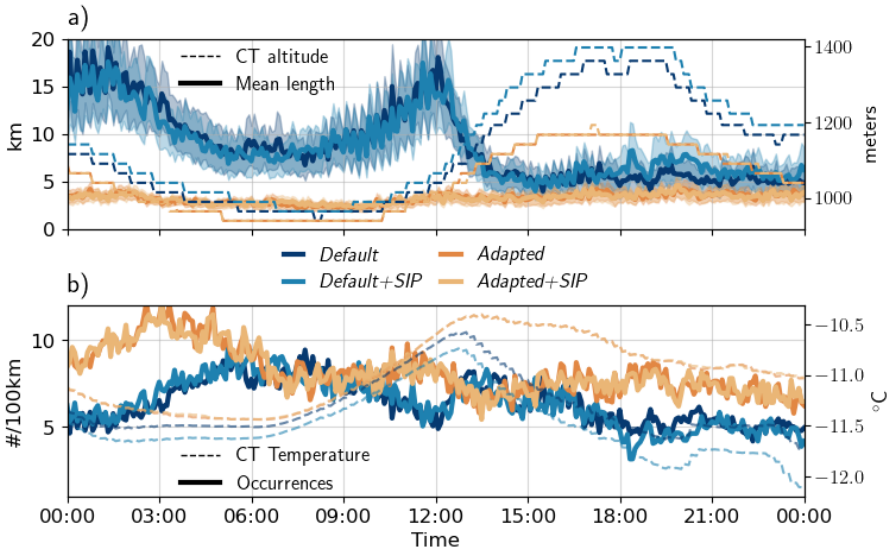

Figure 7.2: Same as in Figure 7.1, but for mixed-phase cloud tops.

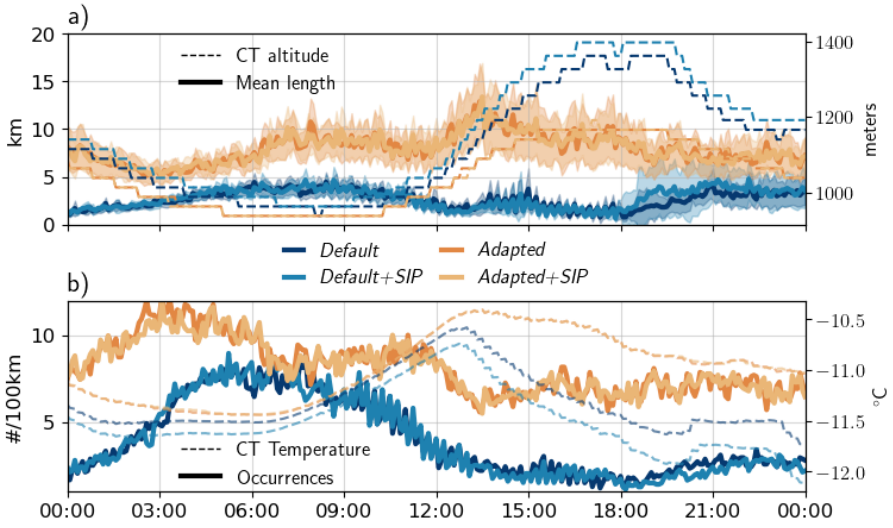

Figure 7.3: Same as in Figure 7.1, but for liquid-topped cloud segments.

decreased again in *Default*, whereas *Adapted* maintains a somewhat higher temperature as the cloud does not ascend as far.

Consistent with findings in Chapter 6, liquid is present in the cloud-top as the dominating phase far more often in *Adapted* compared to *Default* (Figure 7.3). With much fewer INPs available to initiate the primary nucleation of ice, it is expected that the ICNC is consequently decreased in the *Adapted* runs, which was also seen in Chapter 6. However, a reduction in the number of ice particles in the cloud would also act to decrease the efficiency of the WBF, which reduces the sink of liquid water to the ice phase. This is evident by the

much more constant mean length and number of pockets for all three phases for *Adapted* compared to *Default*. The nucleation and subsequent growth of ice in clouds cause a release of latent heat related to the phase change, which can potentially increase the instability and aid in additional vertical growth of the cloud. When the ice nucleation in the model is hampered, for instance through the lowering of the INPC such as in *Adapted*, the latent heat release is much lower, which could explain the greatly reduced cloud-top rising compared to *Default*. This hypothesis is strengthened also by the small differences seen between *Default* and *Default+SIP*, where the latter extends about 40m higher up at the maximum between 15:00 and 18:00. The enhanced production of ice through SIP, as seen during some periods of the day in Chapter 6, is associated with a further increase in the latent heat release, which again may cause the additional cloud-top rising seen in *Default+SIP* (Seifert and Beheng, 2006). This could indicate that the cloud microphysics, here in terms of ice production, may be an important driver for the larger scale dynamical changes to the cloud seen in *Default* through the sudden rising of the mean cloud-top. Such a difference is not seen between the *Adapted* runs, likely because the ice production is already too low to begin with.

Other than the small additional increase in cloud-top height in *Default+SIP*, there are few differences between the run with and without extra SIP processes included. SIP appears to have a much smaller impact on the cloud-top phase distributions compared to the domain-averaged IWC seen in Chapter 6, likely explained at least in part by the cloud-top temperatures falling outside of the HM temperature range. Therefore, any additional production of ice through SIP at cloud-top must come from either BR or DS, but their contributions alone appear largely insignificant for the cloud-top phase analysis in this case.

Spatial intermittency of mixed-phase clouds was studied through in-situ measurements by aircraft in Korolev and Milbrandt (2022b). Aircraft measurements yield a similar data structure as the spatial algorithm presented here, where analysis is carried out along a straight segment over a time span short enough to assume an approximately stationary cloud. They found that genuinely mixed cloud segments may vary in length on a cascade of scales, ranging from 100 km down to at least 100 m. They calculate an average length below 1 km at temperatures around -12°C, which is less than the minimum detection length in the spatial algorithm. The *Default* runs substantially exceeds this estimate, with mean lengths > 5 km throughout the day (Figure 7.2a). The cloud-tops in the *Adapted* runs are dominated by liquid (Figure 7.3), reducing the prevalence of mixed-phase pockets, but even so the mean length of genuinely mixed segments typically computes close to 5 km. This is consistent with discussion presented in Korolev and Milbrandt (2022b), arguing that even cloud-resolving models with horizontal resolutions close to 1 km tend to be biased toward the mixed-phase, due in part to difficulties with representing sub-grid scale heterogeneity.

7.2 Temporal pockets

The results from applying the temporal algorithm on *d03* are shown in Figure 7.4. It is worth noting that the temporal approach causes some inherent

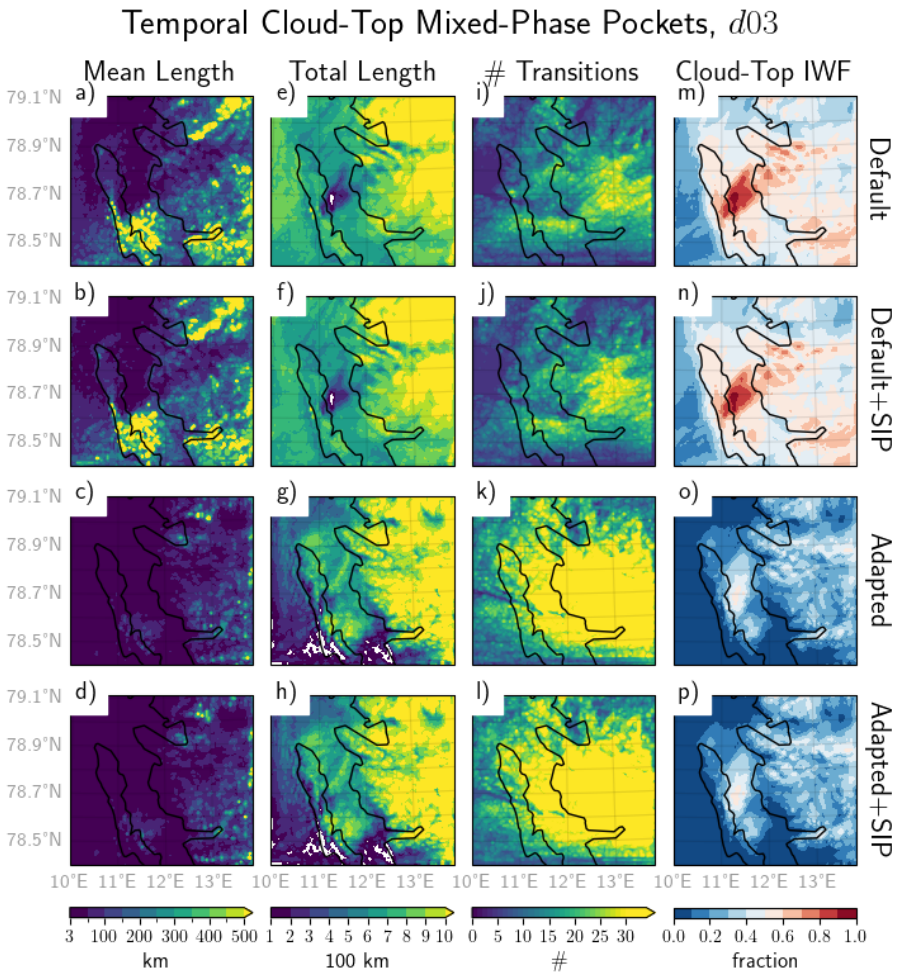

Figure 7.4: Mean length in km (panels a-d), total length in 100 km (panels e-h), and total count (panels i-l) of mixed-phase cloud-top pockets per grid-cell. Panels m-p show the mean cloud-top IWF. Rows correspond to the various model runs. Black contours show the outline of the west coast of Svalbard. Cloud-top is defined here as the mean of the upper 7 model layers of the cloud.

differences from the spatial approach, perhaps most importantly regarding the resolution. Since the pocket lengths are calculated according to Equation 5.7, the minimum pocket length is now a function of both the wind speed and the temporal resolution of the domain. With a mean wind speed at cloud-top of 10 m s^{-1} and the 5 minute output of *d03*, the length of a cloud-top segment lasting only one output time step in the simulation is 3km, which is up to 3 times more than the minimum length in the spatial algorithm.

Consistent with previous results, the differences between runs with and without additional SIP processes are small when looking at the cloud-top. The impact from SIP on the cloud-top pocket structure thus appear to be close to negligible in the simulations for this case study. From Figures 7.4m-p, showing the mean

Table 7.1: Domain-averaged mixed-phase pocket statistics corresponding to Figure 7.4.

	Default	***Default+SIP***	***Adapted***	***Adapted+SIP***
Mean Length (km)	147	150	47	46
Total Length (km)	806	792	603	601
Transitions	10	10	20	20
Mean IWF	0.41	0.41	0.15	0.16

cloud-top IWF it is evident that the *Default* runs contain much more ice at cloud-top compared to *Adapted*. As a result, the pocket structures look very different between the *Default* and *Adapted* runs. Although *Default* and *Adapted* yield similar total lengths of mixed-phase pockets over land, as seen in Figures 7.4e-h, the mean lengths (Figures 7.4a-d) and the number of transitions (Figures 7.4i-l) reveal that mixed-phase pockets in *Adapted* runs tend to be smaller and more intermittent. Whereas the mean pocket lengths averaged across *d03* in the *Default* runs are approximately 150km, the mixed-phase cloud-top segments in the *Adapted* runs are only about 1/3 of this as summarized in Table 7.1.

Evident in all panels of Figure 7.4 is the effect of topography on phase occurrences. Both *Default* and *Adapted* show a clear tendency towards increased occurrence of mixed-phase cloud-tops over land – particularly in terms of total mixed-phase pocket lengths. Since the pockets are evaluated at stationary points through the Eulerian view in the temporal algorithm, the effect of topography is much more pronounced compared to the spatial algorithm. Points located in the vicinity of mountains or steep hills will be under topographic influence throughout the analysed period, and therefore potentially get much different results than for instance oceanic points. The difference between land and ocean is most pronounced in *Adapted* where the total length of mixed-phase pockets are several 100km less than over land. Gierens et al. (2020) showed in their observational study of mixed-phase clouds in Ny-Ålesund that orographically driven local wind patterns can have profound impacts on the properties and occurrence of the low-level mixed-phase clouds. In our case, this is perhaps particularly evident in Figures 7.4m-p where the mean cloud-top IWF is much higher than the domain average for both *Default* and *Adapted* runs in a region over the island Prins Karls Forland to the south-west of Ny-Ålesund. This region of enhanced cloud-top IWF marks the location of the highest mountain of the island, Monacofjellet, with an altitude of 1084 m MSL. The mean wind direction on November 12 was south-westerly, causing updrafts to the south-western side of the mountain as the air is forced upwards, followed by a region of downdrafts on the lee-side of the island. The sinking motion brings the air into the sub-saturated regime where both ice and water may evaporate, but as the liquid evaporates at a faster rate compared to ice, the IWF increases (Heymsfield and Miloshevich, 1993; Lohmann, Henneberger et al., 2016). The orographic effect is seen through an increase of the mean cloud-top IWF in this region of up to 0.5 in both *Default* and *Adapted* runs. The effect is more confined to a smaller region surrounding only the very tallest mountains in the *Adapted* runs, which could be due to the cloud droplets in *Default* evaporating faster due to their smaller size, as discussed in Chapter 6.

7.3 Effect of model resolution on pocket structure

Previous studies have highlighted the influence of model resolution on the simulated pocket structure of mixed-phase clouds, with its associated implications regarding the WBF process (Henneberg et al., 2017; Korolev and Milbrandt, 2022b; Tan et al., 2016). Here, this resolution effect is investigated on the transition between the two smallest domains, *d03* and *d02*, through comparison of the temporal algorithm results as applied on each separately. Shown in Figure 7.5 are the resulting differences between the two domains. Note that the results presented in Section 7.2 for *d03* are not exactly the same as those applied here. For a direct comparison between the two domains, the results from *d03* were regridded to match the resolution of *d02* after the calculations were performed. However, since the temporal resolution also differs between the two domains, the data from *d03* was coarsened to the same hourly output as in *d02* before calculation of the pocket statistics. This was done to avoid resolution effects in the results, which could emerge due to the dependence on temporal resolution in Equation 5.7.

Once again, effects of topography are apparent, similar to the results seen in the previous section. The representation of topography is differing between the two domains due to the 5 times lower resolution in *d02* compared to *d03*. With lower resolution it becomes more difficult for *d02* to pick up features of particularly steep terrain since each 5km^2 area is represented only by a single altitude instead of the 25 in *d03*. This becomes apparent especially in the total pocket length of the *Adapted* runs shown in Figures 7.5g-h. Here, the orographic effects on the total mixed-phase pocket lengths seen in *d03* are not produced as close to the windward side of the coast in *d02*, causing a difference of more than 400km between the domains in some regions. Due to the much lower INPC in *Adapted*, the cloud-tops are in this run mostly liquid, and without the presence of topography to perturb the updrafts, ice is not produced at high enough rates to produce mixed-phase cloud-tops. The effect is not as apparent along the coasts in the *Default* runs, likely because these runs contain enough ice to have cloud-tops well within the mixed-phase regime. Thus, any further perturbation of the vertical motion does not cause severe impacts on the mixed-phase cloud-tops, unless the wake effects cause strong enough downdrafts for the liquid to be depleted appreciably.

Interestingly, as seen in Figures 7.5m-p, there are large differences in the produced IWF at cloud-top between the two domains across all simulations analysed. The *Adapted* runs on *d02* produce a cloud-top IWF about 0.12 higher than on *d03* on average, whereas both *Default* runs on *d02* have an average IWF at cloud-top more than 0.3 above that of *d03*. The difference is seen most clearly over land, but is also present to some degree over ocean without the influence of topography. This effect of increased ice production by the coarser domain could be an indication of an overestimation of the WBF process efficiency in the model. Since ice and liquid is assumed homogeneously mixed within each respective grid-cell of the domain, the 5-times lower resolution of *d02* may cause unrealistically large interfaces between the two phases, whereas the inherent thermodynamical instability of their coexistence would in reality act to separate them (Korolev and Milbrandt, 2022b). In the case of *d03* the phases are still assumed homogeneously mixed within a grid-cell, but the higher

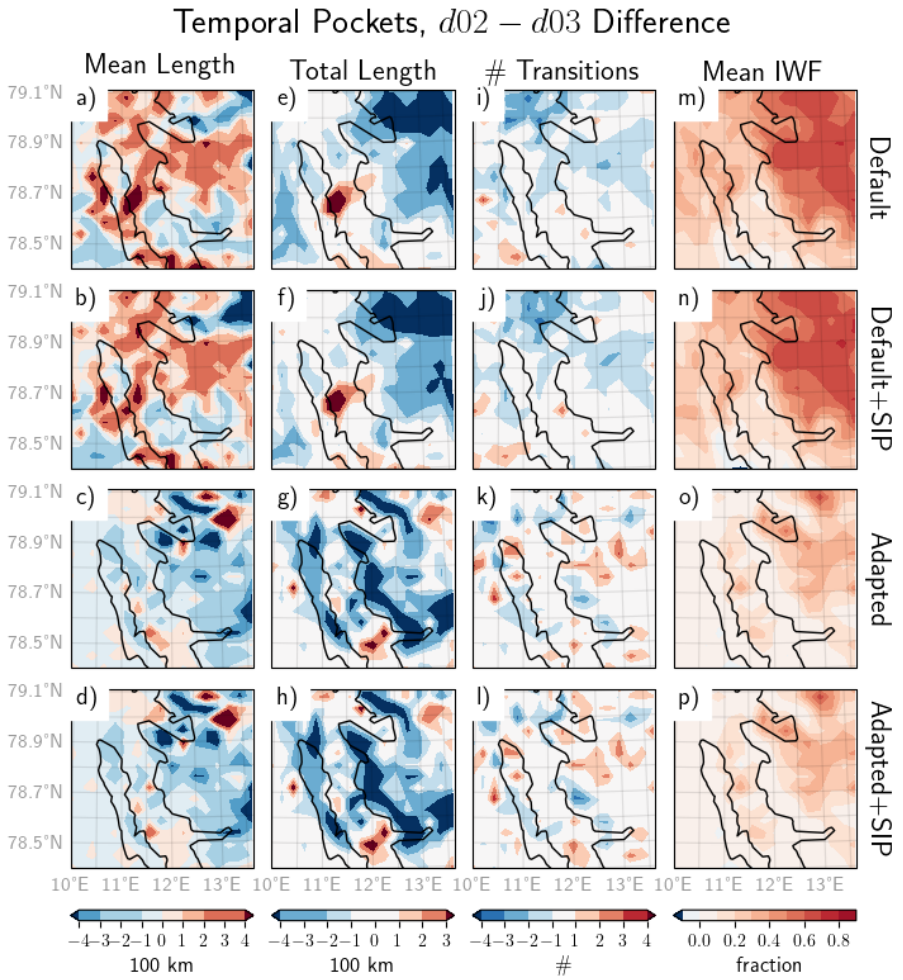

Figure 7.5: Same as in Figure 7.4 but now showing the difference between the clipped domain of *d02* and the temporally and spatially coarsened domain of *d03*. *d03* was coarsened temporally before and spatially after the application of the temporal algorithm.

spatial resolution enables the separation of ice and liquid to a greater extent. As the phases are better separated, the phase interface on which the WBF process can act is reduced and the efficiency may be lowered purely as a result of the higher resolution (Henneberg et al., 2017). This resolution effect on the WBF process has been found to cause large underestimations of supercooled liquid water in clouds represented by global climate models of much coarser resolutions (Komurcu et al., 2014), but clearly applies also to NWP models as shown here.

CHAPTER 8

Variability of Droplet Size Distributions

Cloud microphysics refers to the small-scale processes driving the formation and evolution of individual particles in a population of cloud hydrometeors. Due to the large number of droplets contained within a typical cloud, theoretical descriptions as well as numerical simulations of each individual droplet's life-cycle is impossible (Morrison et al., 2020). Instead, cloud droplet populations are described from a statistical point of view through size distributions. Cloud droplet size distributions are important for a wide range of cloud processes, such as the formation of precipitation (Barthlott et al., 2022), and is closely linked to cloud phase for instance through SIP (Barthlott et al., 2022; Pasquier, Henneberger et al., 2022).

In this chapter, variability of the cloud droplet size distributions are investigated using machine learning to extract cloud segments that contain statistically significant differences from the mean cloud droplet population. The procedure, termed the KSC algorithm (Chapter 4), is applied to observations of cloud droplet size distributions collected with HoloBalloon from the two first flights on 12 November.

The KSC algorithm can be applied on a wide range of scales by merging adjacent holograms together. For instance, running KSC on the original holograms yields a spatial cloud resolution of around 1.7 m, with an assumed wind speed of 10 m s^{-1}. Merging of holograms into 1s time steps gives six times higher total sampled cloud volume, but consequently a lower spatial resolution of around 10 m. Although merging comes at the cost of a lower spatial resolution, it has some significant advantages, such as more statistically robust size distributions due to higher mean droplet counts. Importantly, merging also lowers the computational cost for clustering of a given cloud segment as the number of holograms are reduced.

Section 8.1 presents the results from a case study of a shorter segment of cloud where KSC is applied with both the original high resolution holograms, and with the 1s merged holograms for comparison. Section 8.2 shows the results of clustering droplet size distributions on the whole cloud segments from the flights performed on 12 November, enabled by using the 1s merged holograms. Finally, Section 8.3 features a comparison between the cloud droplet size distributions in the WRF simulations, and the KSC results from Section 8.2.

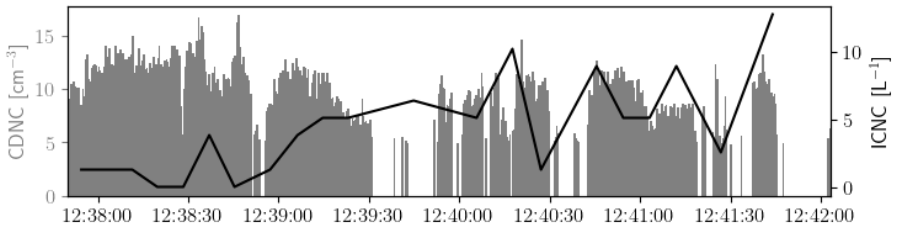

Figure 8.1: CDNC (grey bars, left axis) and 30-second averaged ICNC (solid black line, right axis) measured by HoloBalloon through the case study segment. Holograms with total droplet counts less than the droplet cutoff of 70% of the mean are not included in the analysis and thus not shown.

8.1 Case study: shorter cloud segment

Before applying the KSC algorithm to HoloBalloon data, the technique was validated using a synthetically produced dataset with size distributions drawn randomly from three different gamma functions of known shape and scale. Additionally, the dataset included some size distributions also of random shapes and scales to mimic noise. The algorithm successfully identified the all the size distributions as belonging to each of their parent classes, and most of the noise samples were correctly labelled accordingly. Results from the validation process and a more thorough discussion of the KSC performance can be found in Appendix A.

KSC is applied to a case study of a cloud segment from the second flight with HoloBalloon on 12 November, using both 1.7 m and 10 m resolution to examine the effect of lowering the original resolution. The period from 12:37:49 to 12:42:17 was selected based on a number of factors. The segment was sampled shortly after HoloBalloon entered the cloud, therefore capturing a transect ranging from the cloud-base to 40 meters into the cloud's interior at an altitude of 670 m MSL. As can be seen from Figure 8.1, the CDNC was relatively constant throughout the segment with a mean CDNC = 9.3cm^{-3}, although there were some periods of droplet counts falling below the cutoff threshold at 6.5cm^{-3}. The segment also contained periods with varying degrees of measured ice concentrations, ranging from close to 0 in the beginning of the segment to more than 10L^{-1} towards the end (black line, Figure 8.1).

Shown in Figure 8.2 are the results from KSC applied on the case study segment with 10 m resolution. In the KS-matrix (Figure 8.2a), the dark diagonal running from the upper left corner shows the low KS-score that results from comparing a hologram to itself, and is therefore inherent in all the KS-matrices produced by KSC. This diagonal can be interpreted as the 'timeline' of the cloud segment, showing increasing time downwards along the diagonal. At least two clusters are clearly discernible by eye, with a small one located in the first part of the segment (upper left corner of the KS-matrix), and a second large one spanning almost the entire rest of the segment. HDBSCAN successfully extracts the smaller cluster, labelled Class 1 (blue) with 118 holograms, as seen in Figure 8.2b. The second large cluster is also detected and labelled Class 2 (green) with 708 holograms, but HDBSCAN classifies also a third cluster, Class 3 (red),

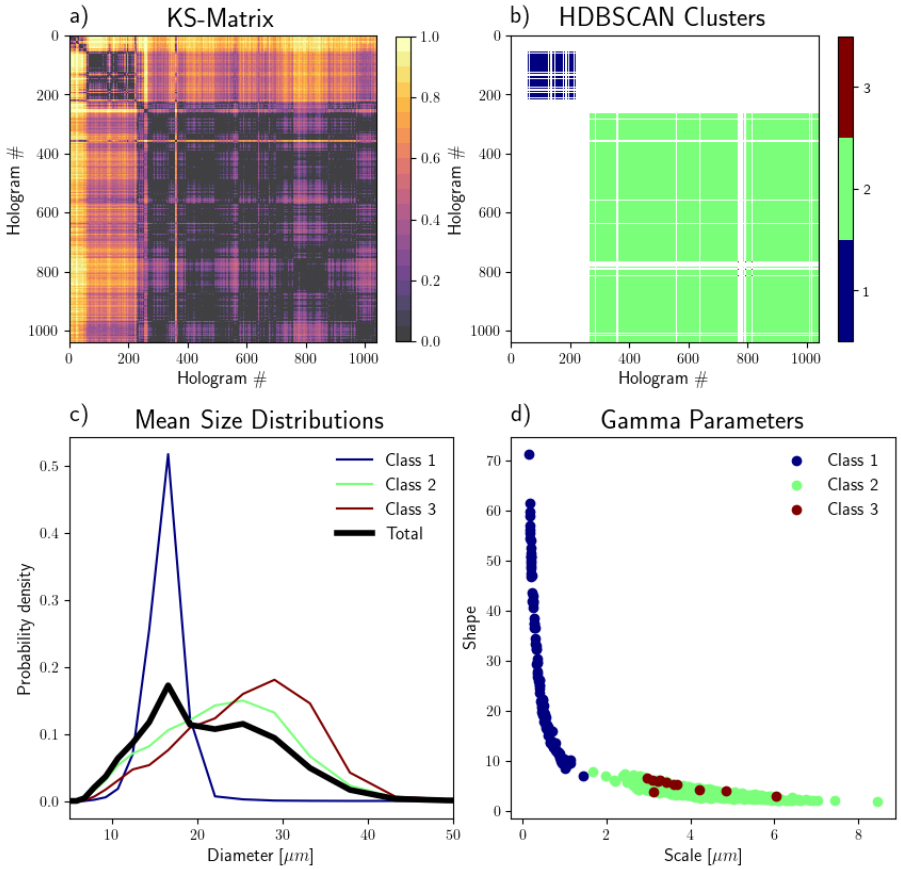

Figure 8.2: Results from the KS-clustering of the case study segment, using the *NOFIT* method and HDBSCAN with `min_sample_size`=10 and `min_samples`=10. Shown in panel a is the KS-matrix with the mean KS-score (colors) between all holograms as indicated by the axes. The cluster matrix (panel b) show the classes (colors) that each pair of holograms are predicted to belong to. The mean size distributions (panel c) and the fitted shape and scale parameters for all class members (panel d) are computed for each of the identified classes individually. Classes are labelled according to increasing average ICNC.

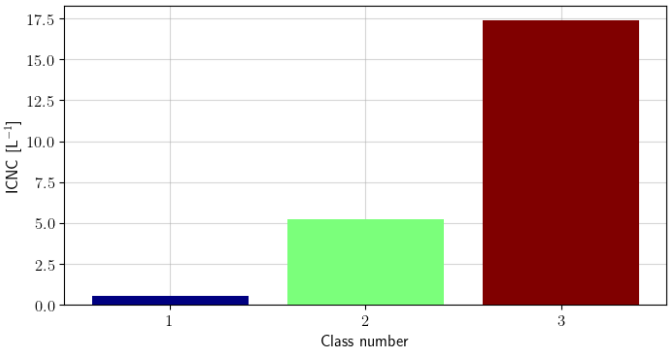

Figure 8.3: ICNC in L^{-1}, averaged across each of the respective classes as detected by the KSC algorithm in Figure 8.2.

encompassed by Class 2 and consisting of only 11 holograms.

From Figures 8.2c-d it is evident that the mean cloud droplet size distributions are clearly different between the classes, although Class 2 and 3 look somewhat more similar in both shape and median size. Class 1 contains on average much smaller cloud droplets and has a distinct peak in probability density at a diameter of approximately $17\mu m$. Class 2 and 3 each have their maximum probability density at approximately 25 and $29\mu m$, respectively. In addition, Class 2 and 3 both have a much broader distribution shape compared to class 1. Where almost all droplets in Class 1 are confined between 10 to $22\mu m$, Class 2 and 3 both have droplets ranging from the minimum detection size of HOLIMO3B of 6 μmat the smallest, to above 40 μmat the largest.

Shown in Figure 8.3 are the mean ICNCs averaged across each of the detected classes for the case study. There are clearly large differences between the three classes. Where Class 1 contains only about $0.5L^{-1}$, the largest cluster, Class 2, contains almost 10 times as much ice with $5.2L^{-1}$. Class 3 has even more ice with $17.4L^{-1}$. Class 3, which is by far the smallest and fully contained within Class 2 in Figure 8.2b, seems to represent the extremities of Class 2 since it consists of distributions with similar shapes and scales but is shifted somewhat towards larger diameters and more ice. As previously mentioned, the cloud segment was sampled shortly after HoloBalloon entered the cloud, and the first part of the segment, represented well by Class 1, is thus collected close to cloud-base. This could explain the high prevalence of small cloud droplets found in Class 1, as the cloud-base marks the first point when rising air achieves supersaturation and liquid droplets start to nucleate. Following the evolution along the updraft from cloud-base, droplets are given more time to grow through condensation and so broaden the size distribution spectrum. The transition from the narrow mean size distribution seen in Class 1 to the much broader mean distributions of Class 2 and 3 occurs in only a matter of a few tens of meters of gained altitude. The rapid transition may be aided by the very clean conditions under which the cloud is forming (Pasquier, David et al., 2022), acting to reduce the competition for water vapor and allowing each droplet to grow larger before a quasi-steady supersaturation is reached (Lamb and Verlinde, 2011). Further growth and broadening of the distribution

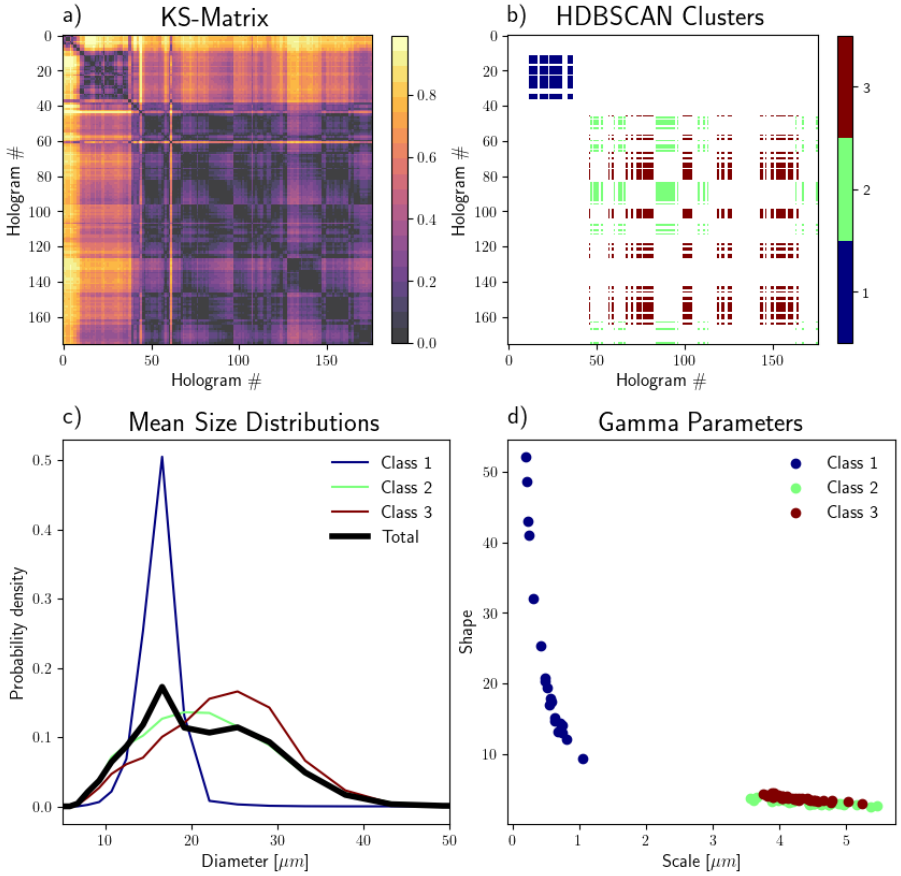

Figure 8.4: Same as in Figure 8.2, but with holograms merged in 1-second time steps.

is then possibly also enhanced through collision-coalescence processes, as some of the fewer but larger droplets produced by clean-air condensation growth may act as coalescence embryo (Lamb and Verlinde, 2011). However, since a large contribution from collision-coalescence would cause a significant decrease in the CDNC, the most heavily influenced droplet size distributions would be expected to fall below the cutoff of the algorithm at 70% of the segment's mean CDNC. Although a small decrease can be seen during the second part of the segment, it is far from enough to explain the broadening of the size distribution seen here. Thus, collision-coalescence was likely not a dominant process during this case study, consistent with results in Motos et al. (2023).

The same cloud segment is analysed with KSC (Figure 8.4) using also a coarsened resolution of 10 m to investigate the sensitivity to spatial resolution, motivated by the potential for using KSC on longer flight segments by lowering the computational cost. Merging of holograms into 1-second time steps involves a 6-fold lowering of the spatial resolution such that each hologram now represents the probing of a 10m long cloud segment. Thus, the total number of holograms included in the KS-matrix is much lower, whereas the total droplet count per

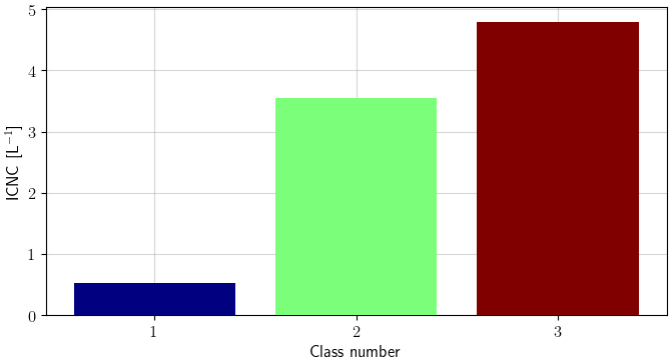

Figure 8.5: Mean ICNC in L^{-1} per class corresponding to Figure 8.4.

hologram is correspondingly increased.

The KS-matrix in Figure 8.4a is visually very similar to the KS-matrix in Figure 8.3a, although the lower resolution is apparent. Class 1 in Figure 8.4b is detected as before, now containing 20 merged holograms corresponding to 120 of the original holograms. Class 2 and 3 look somewhat different from the earlier case in that the distribution of holograms among them is more even. The mean cloud droplet size distributions for each class show the same picture as earlier, with a narrow Class 1 of generally smaller droplets, and the broader Class 2 and 3 containing much larger droplets. However, the differences in the maximum probability density between the classes are somewhat smaller than in the high resolution case, with peaks now at 17, 19, and 25 μm for Class 1, 2 and 3, respectively. Compared to the high resolution analysis, Class 2 now appears more as an intermediate step between Class 1 and 3. Following the merging of 6-by-6 holograms and given a constant value for `min_cluster_size`=10, the small clusters in the high resolution case, such as the previous Class 3 in Figure 8.2b with only 11 holograms, would no longer be detected by HDBSCAN. As a result, Class 3 is now produced instead as a larger cluster with a mean size distribution shifted less towards larger droplets compared to the high resolution clusters. This demonstrates how the lowering of the resolution to some extent acts to even out extremities found on the very smallest scales.

Figure 8.5 shows that the differences in mean ICNC between the three classes are still detected. Class 1 still contains by far the least ice, as expected, since the cluster appears very similar across the two resolutions in Figures 8.2b and 8.4b. Class 3 still has the most ice, but on average much less than before seeing as the cluster is larger and the very high ice contents of the 11 holograms in the previous Class 3 are averaged out across a lot more holograms. The application of merging appears to produce a more linear transition between the classes, rather than picking up smaller classes with very different size distributions. Merging also increases the sample size of size distribution, in this case from about 100 to 600 cloud droplets, which greatly reduces the risk of clustering on insignificant differences in the size distributions arising from too small sample sizes. This is particularly important when using the *NOFIT* method, where the p-value equation taking into account the sample size is not applied.

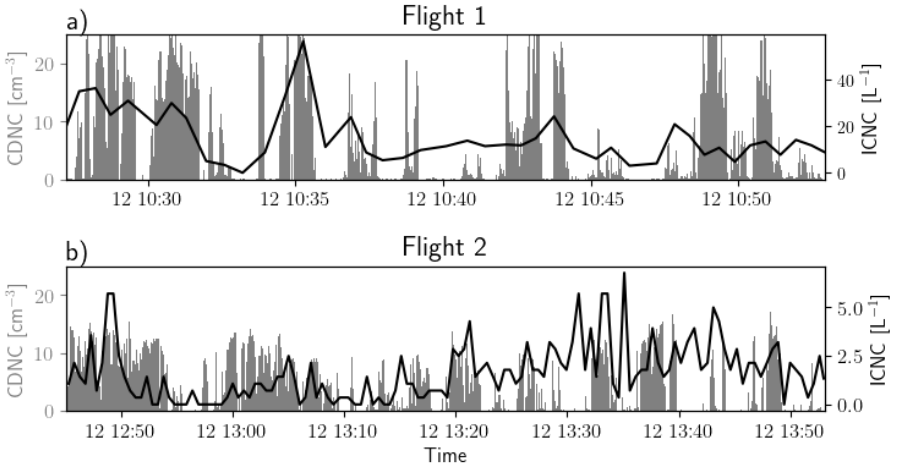

Figure 8.6: CDNC (grey bars, left axis) and 30-second averaged ICNC (solid black line, right axis) measured by HoloBalloon through the entirety of Flight 1 (panel a) and Flight 2 (panel b).

Table 8.1: Cluster statistics for Flight 1 corresponding to Figure 8.7.

Class Number	**1**	**2**	**3**	**4**	**5**	**Noise**
Hologram Count	10	39	19	33	13	80%

8.2 Clustering of entire flights

The KSC algorithm was applied on the 1-second merged cloud segments from the entirety of Flight 1 and 2 with HoloBalloon on 12 November. Flight 1 contained on average almost 10 times more ice than Flight 2, and exhibited much more variation in the CDNC throughout, shown in Figure 8.6. Flight 1 was measuring in cloud for a duration of 25 minutes, whereas Flight 2 lasted longer with a total duration of 70 minutes. Only the parts of the flights measured at a relatively constant altitude were analysed here, so the first and last 2 minutes of in-cloud measurements were excluded. This was done to avoid clustering of the size distribution differences arising mainly from the probing of different stages in the cloud evolution as shown in the previous section with the differences between cloud base and interior.

Shown in Figure 8.7 are the results from clustering on 1-second merged holograms on Flight 1. The KSC algorithm extracts a total of 5 different clusters from the cloud segment, of sizes given in Table 8.1. Worth noting is the significant amount of detected noise, which is likely arising from the use of the *NOFIT* method described in Section 4.2.1. As *NOFIT* does not apply the p-value equation for the KS-test, size distributions are discriminated based on the linear KS-test statistic rather than the binary 'success' or 'failure' variant, potentially making it harder for the clustering algorithm to find regions of higher density in the data. Further considerations on the use of the *NOFIT* method can be found in Appendix A. The two dominant clusters found from Flight 1 are Class 2

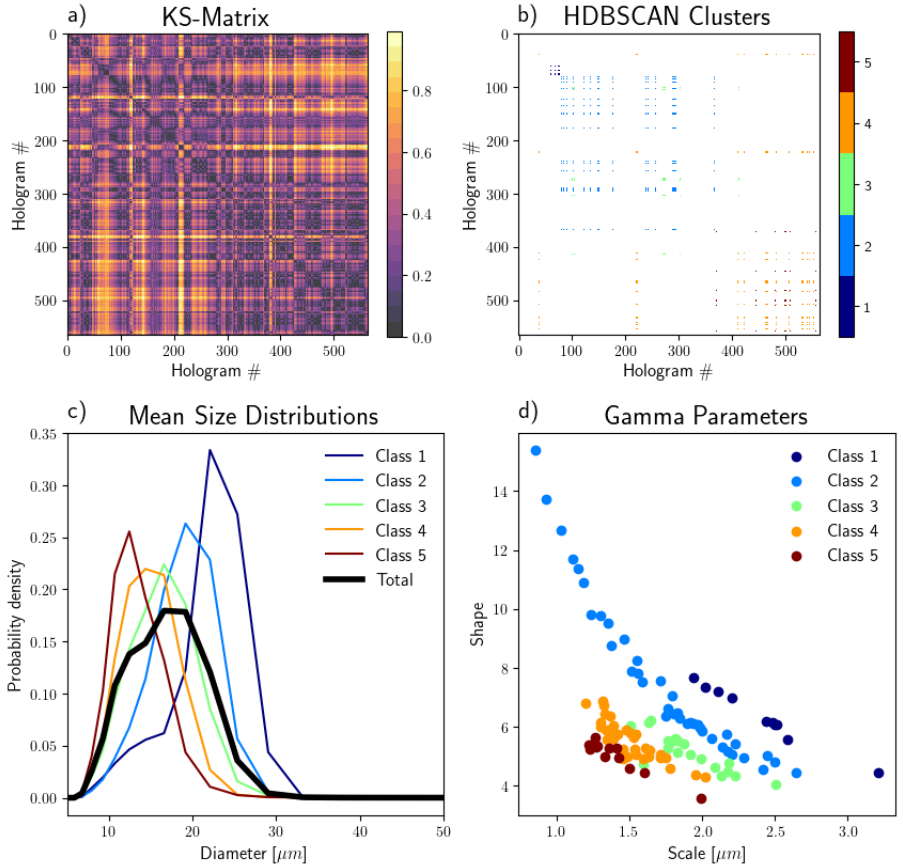

Figure 8.7: Results from applying KSC on the entire cloud segment of Flight 1 with HoloBalloon on 12 November.

(light blue) and Class 3 (orange) consisting of 39 and 33 holograms, respectively, which corresponds to roughly 390 and 330 meters of cloud in total. The mean size distributions of the classes vary greatly in terms of location, ranging in peak probability density from 13 μm(Class 5) to 24 μm(Class 1). All clusters exhibit size distributions of fairly similar shapes, perhaps apart from Class 2, which also contains distributions extending to higher values for the fitted gamma shape parameter (Figure 8.7).

The cluster variability was large also for Flight 2, the results of which are shown in Figure 8.8. A total of 7 size distribution clusters were detected from the cloud segment, with one cluster much larger than all other, labelled Class 6, as shown in Table 8.2. For Flight 2, the clusters consist of both broad and narrow mean size distributions, similar to what was seen in the results from the case study (Figures 8.2 and 8.4). For Flight 1, however, all clusters have relatively narrow mean size distributions, with no broad distributions resembling those found in both the case study and in Flight 2. As can be seen from Figure 8.6a, Flight 1 sampled a cloud segment that contained large variations in CDNC as well as a high ICNC compared to the case study segment taken from Flight 2

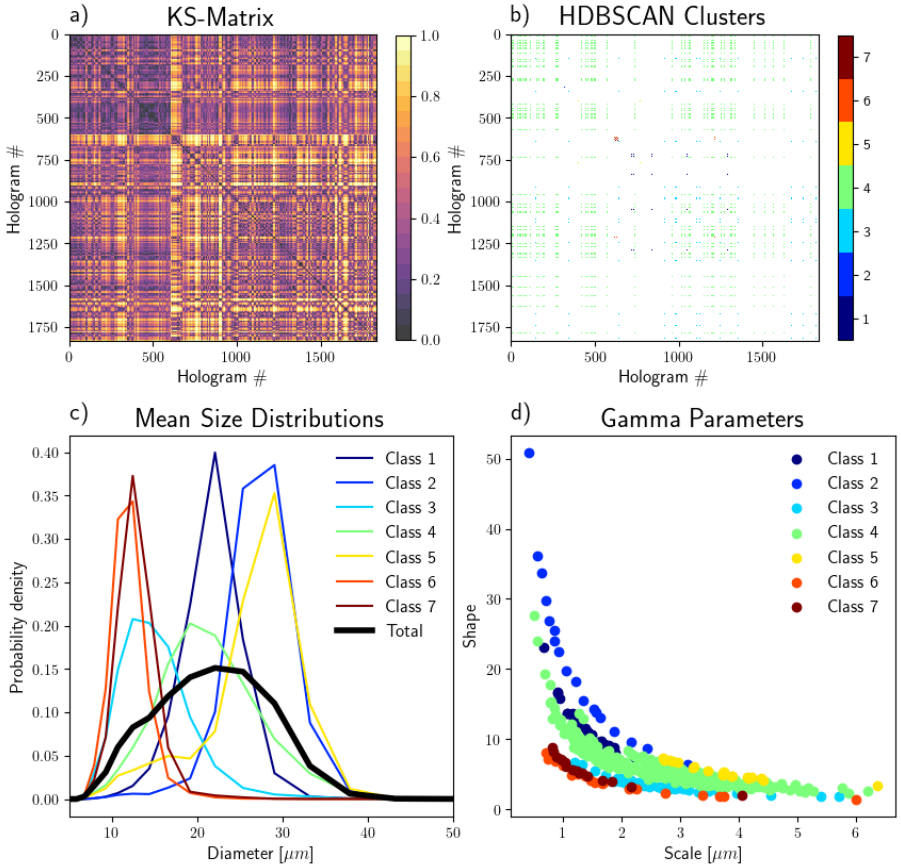

Figure 8.8: Results from applying KSC on the entire cloud segment of Flight 2 with HoloBalloon on 12 November.

Table 8.2: Cluster statistics for Flight 2 corresponding to Figure 8.8.

Class Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Noise**
Hologram Count	18	20	19	16	48	304	32	75%

(Figure 8.1). This could indicate that Flight 1 was more influenced by turbulent mixing, consistent also with updraft velocity measurements (Pasquier et al., 2020), and thus appeared more heterogeneous than what was seen in the case study from Flight 2.

The mean ICNC per class was calculated for both Flight 1 and 2 and is shown in Figures 8.9a and 8.9b, respectively. Classes containing more ice appear to follow the same tendency seen in the case study analysis of having generally larger droplets. Figures 8.7d and 8.7d reveal in combination with Figures 8.9a and 8.9b that higher ice contents are also positively correlated with the fitted gamma shape parameter of the size distributions. The effect is examined further in Figure 8.9c where the ICNC is shown as a function of the mean droplet diameter per class. In Flight 1, the increase in ICNC with increasing mean

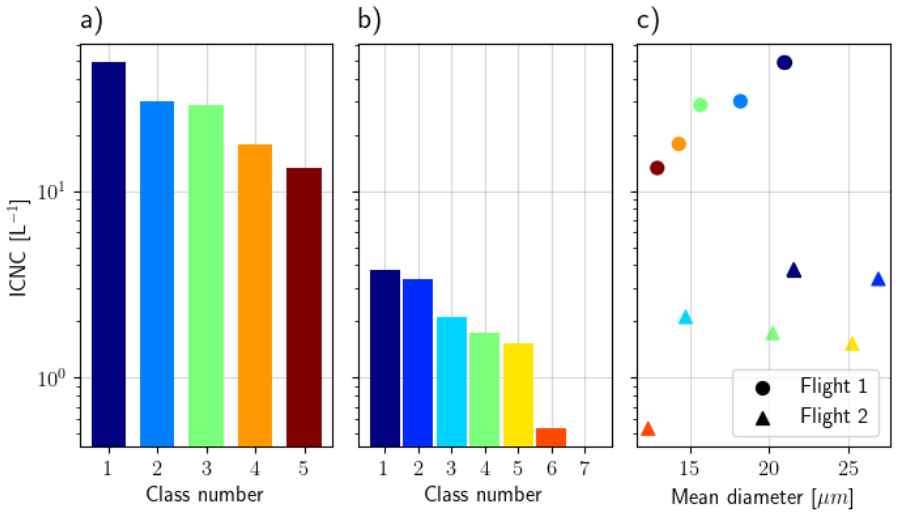

Figure 8.9: Mean ICNC in L^{-1} per class in Flight 1 (panel a) and Flight 2 (panel b), corresponding to Figures 8.7 and 8.8. Panel c) shows the mean ICNC as a function of the mean droplet diameter per class detected in each flight.

droplet diameters is monotonic and grows approximately exponentially. In Flight 2 the tendency is more varied, but a positive relation is still evident.

As with any two correlated variables there are three possibilities regarding the apparent relationship between droplet size distribution shapes and ice contents:

1. the presence of ice has a causal effect on the size distribution shape,
2. the size distribution shape has a causal effect on ice contents, or
3. both are affected simultaneously by secondary conditions, and there is no direct causal relationship between them.

In this case, the three types of relationships are all possible and may exist in combination with each other.

One possible explanation for the evident correlation is offered by the WBF theory. Growth or evaporation rates of liquid droplets are determined by the saturation vapor pressure of the interstitial air, giving rise to a quasi-steady equilibrium between water in the liquid and gaseous phase. If ice is present, however, vapor may be lost through depositional growth of the ice crystal, thereby altering the balance between the liquid droplets and the vapor, ultimately resulting in evaporation of liquid to restore the equilibrium (Lamb and Verlinde, 2011). If enough ice is present in proximity to the liquid under the right conditions, some smaller droplets may lose enough mass to disintegrate completely. This could potentially explain why the classes with high ICNC have a much lower frequency of occurrence of smaller droplets, as seen in both Figure 8.7 and 8.8.

The complete loss of smaller droplets to the ice phase would also contribute to the higher mean droplet diameters seen in Figure 8.9. However, it does not

explain why the larger droplets are more frequently observed under high ice conditions. This could, however, possibly be explained through the second causal relationship, wherein the size distribution shape affects the ICNC. SIP by the HM process has been found through laboratory experiments to require the presence of droplets with radii both below 12 μmand above 24 μm(Korolev and Leisner, 2020; Mossop, 1978). This condition is clearly met for instance in Class 1 of Flight 1, which has a secondary plateau in probability density for droplets below 16 μmin addition to the primary peak at $25\mu m$. On the other hand, the classes with ICNC close to zero, such as Class 4 and 5, contain almost no droplets large enough for HM to occur. For Flight 2 only two classes are clearly outside the HM size thresholds, Class 6 and 7, which are also the classes containing by far the least ice. This is particularly significant as Class 6 is also the largest of the detected classes, consisting of more than 300 merged holograms.

Another potential mechanism to explain the presence of high ICNC with larger droplets is through the DS process. Recent studies have suggested that droplet shattering may occur when supercooled liquid droplets larger than 50 μmare present (Lauber et al., 2018) and that the probability of DS occurrence increases with increasing diameters (Keinert et al., 2020). Analysis conducted by Pasquier, Henneberger et al. (2022) following the NASCENT field campaign reveals evidence of a high SIP event during Flight 1 on 12 November. They find that large supercooled liquid droplets were present in this time span and that conditions were favorable for SIP by the DS mechanism. Although hardly visible in the probability density distributions in Figures 8.7 and 8.8, owing to their low relative frequency of occurrence, cloud droplets above 50 μm were detected during both flights (not shown), thereby likely contributing at least in part to the positive dependence of ICNC on droplet size found in Figure 8.9c.

8.3 Comparison with size distributions from WRF

Cloud microphysics in NWP and climate models can typically be sub-categorized into two types regarding how they handle droplet size distributions (Morrison et al., 2005). Bin models explicitly simulate the evolution of cloud particles and are thus, able to predict the behavior of the size distributions, at the cost of requiring substantial amounts of computing resources. Bulk models represent a simplified alternative where the size distributions are assumed to have a shape, scale and/or intercept based on a variety of thermodynamic conditions. Under this simplification, size distributions are no longer simulated explicitly in the cloud evolution, but bulk models have been found able to replicate most features predicted by bin models in cloud-resolving models, while requiring several orders of magnitude less model integration time (Jiang et al., 2000; Morrison et al., 2005). The Morrison microphysics scheme is a bulk model and assumes cloud droplet size distributions based on the simulated air temperature, pressure, and mixing ratio of cloud droplets. It also depends on the CDNC, which is constant in the Morrison scheme and set in this case to 250 and 10cm^{-3} for *Default* and *Adapted*, respectively.

The cloud droplet size distributions as simulated in the model experiments during the two flights are shown in Figure 8.10 together with the observed and

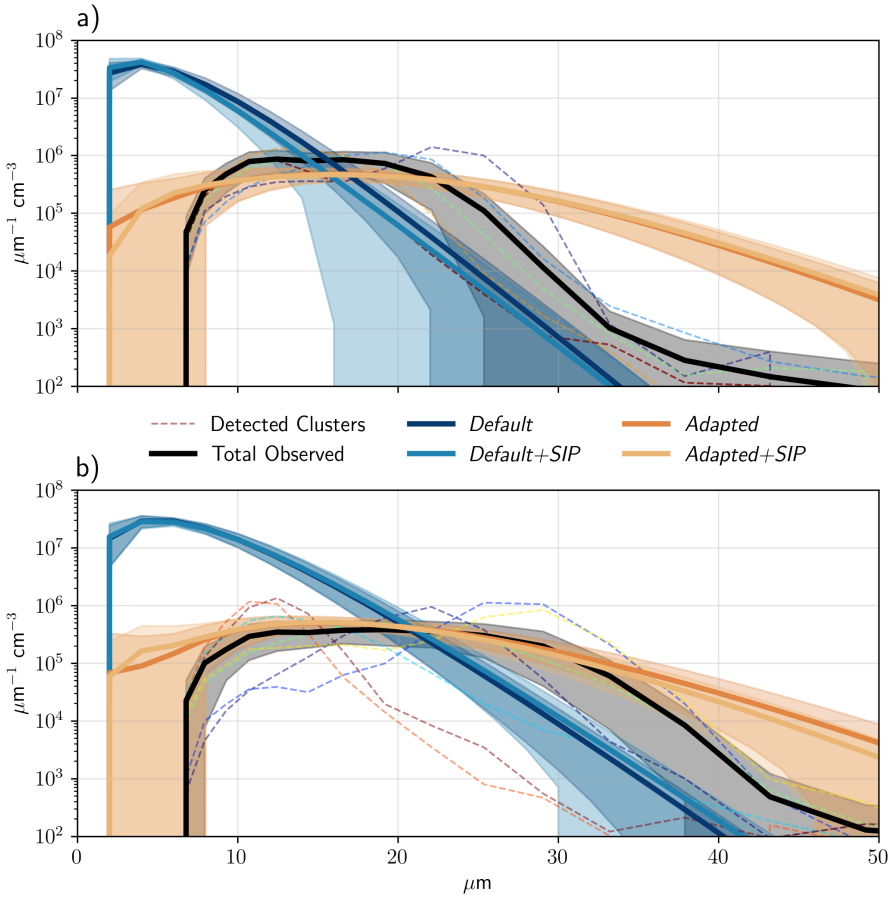

Figure 8.10: Simulated and observed bin-weighted cloud droplet size distributions in $\mu m^{-1} cm^{-3}$ from Flight 1 (panel a) and Flight 2 (panel b) on 12 November. Solid, coloured lines are the mean simulated size distributions within a box surrounding the location of HoloBalloon, averaged over the duration of the flights. Shaded coloured area shows the spread represented by a distance of one half standard deviation from the mean. Solid, black curve shows the average observed size distributions from HoloBalloon, and the coloured, dashed lines represent each of the classes as detected by the KSC algorithm corresponding to Figures 8.7 and 8.8 for Flight 1 and 2, respectively.

classified size distributions from the KSC algorithm. Size distributions from the model are calculated as the average over a box surrounding the location of HoloBalloon, similar to what was done in Chapter 6, and the shaded region represents one half standard deviation of the the spread within this box during the flights. The lower droplet size detection limit with HOLIMO3B of 6 μm is evident in the figure, and explains the discrepancy between the model runs and HoloBalloon at these small diameters. Large differences in shape between the *Default* and *Adapted* runs are apparent during both flights, where the *Adapted* runs show much larger concentrations for larger droplet sizes. As previously discussed, this is typical for clean air conditions where the competition for water vapor between droplets is greatly reduced due to the scarcity of available CCN (Georgakaki et al., 2021; Twomey, 1974). Through the lowering of the CDNC, acting as a proxy for CCNC, in the *Adapted* runs, the model achieves much better agreement with observed concentrations of cloud droplets within the range of about 10-35 μm. With the configuration of CDNC in the *Default* runs, the assumed size distribution substantially overestimates the presence of small cloud droplets (less than 20 μm while larger droplets are underrepresented. The *Adapted* simulations, however, greatly overestimate the occurrence of large droplets during both flights. The overestimation is particularly evident in Flight 1, as the *Adapted* runs fail to capture the observed narrowing of the mean size distributions compared to Flight 2. The *Default* runs, on the other hand, do capture an increase in the occurrence of larger droplets from Flight 1 to 2 similar to what was observed. The drop-off in concentration with increasing droplet sizes also corresponds better between the *Default* runs and observations. Inaccurate representations of cloud droplet size distributions may be problematic for the simulation of a wide range of macrophysical properties, such as radiation transfer (Wang et al., 2022) and precipitation efficiency (Liu et al., 2020), and microphysical processes, such as DS (Phillips et al., 2018) or riming (Saleeby and Cotton, 2008). These results indicate that the Morrison scheme may assume shapes of the cloud droplet size distributions that do not represent the cloud droplet populations well under Arctic conditions.

PART IV

Conclusion & Outlook

CHAPTER 9

Conclusion

9.1 Summary

This thesis was centered around a case study of Arctic mixed-phase clouds on 12 November 2019, which was comprehensively observed during the NASCENT field campaign (Pasquier, David et al., 2022). A combination of cloud-resolving modelling and in-situ measurements was used to evaluate some important yet unanswered questions regarding the characteristics of mixed-phase clouds. These are summarized through three overarching research questions stated in Chapter 1, which loosely correspond to the three parts of the results featured in Chapters 6, 7, and 8, respectively.

Firstly, **Research Question 1** asked the question of how often and under which conditions mixed-phase clouds tend to be genuinely mixed. In Chapter 6, it was found that although the modelled cloud contained an average IWF well within the mixed-phase zone, only about 5% of the total cloud volume could be considered genuinely mixed. The use of representative aerosols in the simulation caused a shift from ice to liquid, with the occurrence of liquid increasing from 70% to about 90%, however the cloud maintained the same percentage of genuinely mixed cloud volume at 5%. Furthermore, when constrained to observed INPC and CDNC the model was found to severely underestimate the prevalence of ice compared to in-situ measurements. The default configuration of the model was found to perform better in terms of simulated ice concentrations, although likely for the wrong reason through overestimating INPC. Including parameterizations of BR and DS as additional SIP pathways, as well as lowering the thresholds required for HM initiation, showed the potential of SIP to reduce the discrepancy between modelled and observed ice. Further research on the efficiency of these processes, for instance regarding the number of splinters produced by HM, may yield further improvements to the simulation of ice under realistic Arctic conditions.

Secondly, **Research Question 2** regards the simulated spatial distribution of cloud phase. Chapter 7 addressed this from the perspective of cloud-top through two new analysis algorithms developed in this thesis. It was found that the spatial scales at which a mixed-phase cloud can be considered genuinely mixed is highly influenced by the representation of INPC in the model, but largely insensitive to the SIP parameterizations used in this thesis. Other external factors, such as topography and cloud-top lifting, were also found

to have a substantial impact on the cloud-top phase variability. Effects of model resolution were revealed to influence not only the mean IWF but also the spatial scales of genuine mixing at cloud-top, and may thus have important implications for the efficiency of the WBF process. One possible solution to mitigate the resolution issue includes the representation of subgrid variability when simulating the WBF process. This has previously been done for large scale climate models of much coarser resolution (e.g. Storelvmo et al., 2008), but could potentially yield an impact even for high resolution models as indicated by the results presented here.

Lastly, **Research Question 3** considered variability in cloud droplet size distributions on various scales, and the relationship between size distribution characteristics and ice. In Chapter 8, an adapted version of an automated clustering algorithm following Allwayin et al. (2022) was implemented and applied to in-situ observations of droplet diameters from the NASCENT campaign. The clustering algorithm detected several groups of significantly different size distributions for all analysed cloud segments, exhibiting large spread in both shape and median diameter. A positive relation was found between mean ICNC and median droplet diameter per distribution. Several possible explanations are proposed for the apparent relationship, such as enhanced SIP through HM or DS, but assertion of their relative importance requires further research. Observations were also compared to the implicitly simulated droplet size distributions from the model, upon which a large overestimation of small droplets was revealed using the default model configuration. Much better agreement with observations was achieved by constraining the model to observations of CDNC, although this instead caused an overestimation of larger cloud droplets. This shows that the representation of droplet size distributions in the model is, as of now, likely not ideal for use under clean Arctic conditions, and should be evaluated in future modelling of Arctic clouds.

9.2 Outlook and concluding remarks

In this thesis, two new methods for the quantification of modelled cloud-top heterogeneity were developed, which are particularly useful for comparison with satellite products, aircraft measurements, or stationary in-situ observations of cloud phase. One possible path forward entails the implementation of more sophisticated methods for detection of single- or mixed-phase cloud pockets. Unsupervised machine learning would likely prove particularly useful for this purpose, as it could, with minimal subjective influence, detect irregularly shaped pocket structures and provide more thorough statistical descriptions of pocket characteristics, such as lifetime, size, composition, and, perhaps most importantly, evolution. This improvement to the pocket analysis algorithms could potentially contribute to a more thorough understanding of how the model handles spatial phase heterogeneities at cloud-top.

Furthermore, the apparent relationship that was found between ice contents and the characteristics of the droplet size distributions should be investigated further. The clustering algorithm, first described in Allwayin et al. (2022) and adapted here, proved useful for the analysis carried out in this thesis,

but further improvements are possible. Specifically, the algorithm applies the Kolmogorov-Smirnov test as the metric for distribution dissimilarity, which has a null distribution that is strictly not valid for discrete data (Noether, 1963). This thesis presented two possible workarounds, which both yield satisfactory results. However, there exists alternatives to the Kolmogorov-Smirov test which are designed or modified to work also under the discrete case (Dufour and Farhat, 2002), some of which could potentially yield performance improvements when the algorithm is generalized to discrete data.

Finally, it has been evident through this thesis that the WRF model with microphysics parameterized with the Morrison scheme struggles to accurately represent mixed-phase clouds under the clean Arctic conditions that were observed during the NASCENT campaign. This is clearly a problem that requires further attention in future modelling research. Several studies have pointed to the role of SIP being underrepresented in models, and called for the inclusion of more SIP processes in addition to HM, in particular BR and DS (eg. James et al., 2021; Pasquier, Henneberger et al., 2022; Phillips et al., 2017; Sotiropoulou et al., 2020). Here it was found that, although the representation of BR and DS and the modification of the HM initiation thresholds improved simulated ice concentrations to some extent, their relative efficiency was not sufficient to achieve agreement with observations. This was presumably due to shortcomings in the SIP parameterizations arising from insufficient knowledge of their mechanisms. A combination of theoretical research striving for a more a complete description of the underlying physics, and experimental research to better quantify these processes, is likely needed to improve the representation of several SIP pathways in models. These advancements have proven necessary to reduce the discrepancy between simulated and observed ice mass and number concentrations in Arctic clouds, and to ultimately improve our understanding of the complex Arctic climate system.

Bibliography

Albrecht, B. A. (15th Sept. 1989). 'Aerosols, cloud microphysics, and fractional cloudiness'. In: *Science* vol. 245, no. 4923. Publisher: American Association for the Advancement of Science, pp. 1227–1231.

Allwayin, N., Larsen, M. L., Shaw, A. G. and Shaw, R. A. (1st July 2022). 'Automated Identification of Characteristic Droplet Size Distributions in Stratocumulus Clouds Utilizing a Data Clustering Algorithm'. In: *Artificial Intelligence for the Earth Systems* vol. 1, no. 3. Publisher: American Meteorological Society Section: Artificial Intelligence for the Earth Systems.

Ambaum, M. H. P. (2020). 'Accurate, simple equation for saturated vapour pressure over water and ice'. In: *Quarterly Journal of the Royal Meteorological Society* vol. 146, no. 733. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3899, pp. 4252–4258.

Amini, A., Wah, T. Y. and Saboohi, H. (1st Jan. 2014). 'On Density-Based Data Streams Clustering Algorithms: A Survey'. In: *Journal of Computer Science and Technology* vol. 29, no. 1, pp. 116–141.

Auer, A., Veal, D. and Marwitz, J. (1969). 'Observations of Ice Crystal and Ice Nuclei Concentrations in Stable Cap Clouds'. In: *Journal of the Atmospheric Sciences* vol. 26, p. 1342.

Barthlott, C., Zarboo, A., Matsunobu, T. and Keil, C. (16th Feb. 2022). 'Importance of aerosols and shape of the cloud droplet size distribution for convective clouds and precipitation'. In: *Atmospheric Chemistry and Physics* vol. 22, no. 3. Publisher: Copernicus GmbH, pp. 2153–2172.

Bergeron, T. (1928). 'Über die dreidimensional verknüpfende Wetteranalyse'. In: *Geophysical Publications* vol. 5, no. 3.

Berkhin, P. (2006). 'A Survey of Clustering Data Mining Techniques'. In: *Grouping Multidimensional Data: Recent Advances in Clustering*. Ed. by Kogan, J., Nicholas, C. and Teboulle, M. Berlin, Heidelberg: Springer, pp. 25–71.

Berry, M. W., Mohamed, A. and Yap, B. W. (2020). *Supervised and Unsupervised Learning for Data Science*. 1st ed. 2020. Unsupervised and Semi-Supervised Learning. Cham: Springer International Publishing : Imprint: Springer. 1 online resource (VIII, 187 p. 55 illu, 45 illu in color.)

Bigg, E. K. (1953). 'The Supercooling of Water'. In: *Proceedings of the Physical Society. Section B* vol. 66, no. 8. Publisher: IOP Publishing, pp. 688–694.

Bodas-Salcedo, A., Hill, P. G., Furtado, K., Williams, K. D., Field, P. R., Manners, J. C., Hyder, P. and Kato, S. (1st June 2016). 'Large Contribution

of Supercooled Liquid Clouds to the Solar Radiation Budget of the Southern Ocean'. In: *Journal of Climate* vol. 29, no. 11. Publisher: American Meteorological Society Section: Journal of Climate, pp. 4213–4228.

Boer, G. d., Eloranta, E. W. and Shupe, M. D. (1st Sept. 2009). 'Arctic Mixed-Phase Stratiform Cloud Properties from Multiple Years of Surface-Based Measurements at Two High-Latitude Locations'. In: *Journal of the Atmospheric Sciences* vol. 66, no. 9. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 2874–2887.

Burrows, S. M. et al. (2022). 'Ice-Nucleating Particles That Impact Clouds and Climate: Observational and Modeling Research Needs'. In: *Reviews of Geophysics* vol. 60, no. 2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021RG000745, e2021RG000745.

Campello, R. J. G. B., Kröger, P., Sander, J. and Zimek, A. (2020). 'Density-based clustering'. In: *WIREs Data Mining and Knowledge Discovery* vol. 10, no. 2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1343, e1343.

Campello, R. J. G. B., Moulavi, D. and Sander, J. (2013). 'Density-Based Clustering Based on Hierarchical Density Estimates'. In: *Advances in Knowledge Discovery and Data Mining*. Ed. by Pei, J., Tseng, V. S., Cao, L., Motoda, H. and Xu, G. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 160–172.

Carey, L. D., Niu, J., Yang, P., Kankiewicz, J. A., Larson, V. E. and Vonder Haar, T. H. (2008). 'The Vertical Profile of Liquid and Ice Water Content in Midlatitude Mixed-Phase Altocumulus Clouds'. In: *Journal of Applied Meteorology and Climatology* vol. 47, no. 9. Publisher: American Meteorological Society, pp. 2487–2495.

Chavez, S. P. and Barros, A. P. (2023). 'Aerosol indirect effects on orographic clouds and precipitation'. In: *Frontiers in Earth Science* vol. 11.

Clapeyron, É. (1834). 'Mémoire sur la puissance motrice de la chaleur'. In: *Journal de l'École polytechnique* vol. 14, pp. 153–190.

Clausius, R. (1850). 'Über die bewegende Kraft der Wärme'. In: *Annalen der Physik* vol. 79, pp. 368–397.

Collins, W. et al. (2004). 'Description of the NCAR Community Atmosphere Model (CAM 3.0)'. In.

Cooper, W. A. (1986). 'Ice Initiation in Natural Clouds'. In: *Precipitation Enhancement—A Scientific Challenge*. Ed. by Braham, R. R. et al. Meteorological Monographs. Boston, MA: American Meteorological Society, pp. 29–32.

David, R. O. et al. (20th Dec. 2019). 'Development of the DRoplet Ice Nuclei Counter Zurich (DRINCZ): validation and application to field-collected snow samples'. In: *Atmospheric Measurement Techniques* vol. 12, no. 12. Publisher: Copernicus GmbH, pp. 6865–6888.

DeMott, P. J., Prenni, A. J., Liu, X., Kreidenweis, S. M., Petters, M. D., Twohy, C. H., Richardson, M. S., Eidhammer, T. and Rogers, D. C. (22nd June 2010). 'Predicting global atmospheric ice nuclei distributions and their impacts on climate'. In: *Proceedings of the National Academy of Sciences* vol. 107, no. 25. Publisher: Proceedings of the National Academy of Sciences, pp. 11217–11222.

Dong, X. and Mace, G. G. (2003). 'Arctic stratus cloud properties and radiative forcing derived from ground-based data collected at Barrow, Alaska'. In: *J. Climate* vol. 16, pp. 445–461.

Dufour, J.-M. and Farhat, A. (2002). 'Exact Nonparametric Two-Sample Homogeneity Tests'. In: *Goodness-of-Fit Tests and Model Validity*. Ed. by Huber-Carol, C., Balakrishnan, N., Nikulin, M. S. and Mesbah, M. Statistics for Industry and Technology. Boston, MA: Birkhäuser, pp. 435–448.

Ester, M., Kriegel, H.-P., Sander, J. and Xu, X. (2nd Aug. 1996). 'A density-based algorithm for discovering clusters in large spatial databases with noise'. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. KDD'96. Portland, Oregon: AAAI Press, pp. 226–231.

Field, P. R. et al. (1st Jan. 2017). 'Secondary Ice Production: Current State of the Science and Recommendations for the Future'. In: *Meteorological Monographs* vol. 58, no. 1. Publisher: American Meteorological Society Section: Meteorological Monographs, pp. 7.1–7.20.

Findeisen, W. (1938). 'Kolloid-meteorologische Vorgänge bei Neiderschlagsbildung'. In: *Meteorologische Zeitschrift* vol. 55, pp. 121–133.

Fu, S., Deng, X., Shupe, M. D. and Xue, H. (1st Nov. 2019). 'A modelling study of the continuous ice formation in an autumnal Arctic mixed-phase cloud case'. In: *Atmospheric Research* vol. 228, pp. 77–85.

Fugal, J. P., Shaw, R. A., Saw, E. W. and Sergeyev, A. V. (10th Nov. 2004). 'Airborne digital holographic system for cloud particle measurements'. In: *Applied Optics* vol. 43, no. 32. Publisher: Optica Publishing Group, pp. 5987–5995.

Georgakaki, P., Sotiropoulou, G., Vignon, É., Billault-Roux, A.-C., Berne, A. and Nenes, A. (11th Feb. 2022). 'Secondary ice production processes in wintertime alpine mixed-phase clouds'. In: *Atmospheric Chemistry and Physics* vol. 22, no. 3. Publisher: Copernicus GmbH, pp. 1965–1988.

Georgakaki, P. et al. (20th July 2021). 'On the drivers of droplet variability in alpine mixed-phase clouds'. In: *Atmospheric Chemistry and Physics* vol. 21, no. 14. Publisher: Copernicus GmbH, pp. 10993–11012.

Gibbons, J. D. (1992). *Nonparametric statistical inference*. In collab. with Chakraborti, S. 3rd ed., rev. and expanded. Vol. 131. Statistics : textbooks and monographs. New York: Marcel Dekker. xix, 544.

Gierens, R., Kneifel, S., Shupe, M. D., Ebell, K., Maturilli, M. and Löhnert, U. (24th Mar. 2020). 'Low-level mixed-phase clouds in a complex Arctic environment'. In: *Atmospheric Chemistry and Physics* vol. 20, no. 6. Publisher: Copernicus GmbH, pp. 3459–3481.

Hallett, J. and Mossop, S. C. (May 1974). 'Production of secondary ice particles during the riming process'. In: *Nature* vol. 249, no. 5452. Number: 5452 Publisher: Nature Publishing Group, pp. 26–28.

Hellmuth, F. and Hofer, S. (2019). *Weathermaps From MEPS latest runs*. Github.

Henneberg, O., Henneberger, J. and Lohmann, U. (1st Nov. 2017). 'Formation and Development of Orographic Mixed-Phase Clouds'. In: *Journal of the Atmospheric Sciences* vol. 74, no. 11. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 3703–3724.

Heymsfield, A. J. and Miloshevich, L. M. (1st Aug. 1993). 'Homogeneous Ice Nucleation and Supercooled Liquid Water in Orographic Wave Clouds'. In:

Journal of the Atmospheric Sciences vol. 50, no. 15. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 2335–2353.

Hodges, J. L. (1958). 'The significance probability of the smirnov two-sample test'. In: *Arkiv för matematik* vol. 3, no. 5, pp. 469–486.

Hong, S.-Y., Noh, Y. and Dudhia, J. (1st Sept. 2006). 'A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes'. In: *Monthly Weather Review* vol. 134, no. 9. Publisher: American Meteorological Society Section: Monthly Weather Review, pp. 2318–2341.

Houze, R. A. (1st Jan. 2014). 'Chapter 3 - Cloud Microphysics'. In: *International Geophysics*. Ed. by Houze, R. A. Vol. 104. Cloud Dynamics. Academic Press, pp. 47–76.

James, R. L., Phillips, V. T. J. and Connolly, P. J. (21st Dec. 2021). 'Secondary ice production during the break-up of freezing water drops on impact with ice particles'. In: *Atmospheric Chemistry and Physics* vol. 21, no. 24. Publisher: Copernicus GmbH, pp. 18519–18530.

Jiang, H., Cotton, W. R., Pinto, J. O., Curry, J. A. and Weissbluth, M. J. (1st July 2000). 'Cloud Resolving Simulations of Mixed-Phase Arctic Stratus Observed during BASE: Sensitivity to Concentration of Ice Crystals and Large-Scale Heat and Moisture Advection'. In: *Journal of the Atmospheric Sciences* vol. 57, no. 13. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 2105–2117.

Kanji, Z. A., Ladino, L. A., Wex, H., Boose, Y., Burkert-Kohn, M., Cziczo, D. J. and Krämer, M. (1st Jan. 2017). 'Overview of Ice Nucleating Particles'. In: *Meteorological Monographs* vol. 58, no. 1. Publisher: American Meteorological Society Section: Meteorological Monographs, pp. 1.1–1.33.

Keinert, A., Spannagel, D., Leisner, T. and Kiselev, A. (10th Aug. 2020). 'Secondary Ice Production upon Freezing of Freely Falling Drizzle Droplets'. In: *Journal of the Atmospheric Sciences* vol. 77, no. 8. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 2959–2967.

Kiselev, A., Bachmann, F., Pedevilla, P., Cox, S. J., Michaelides, A., Gerthsen, D. and Leisner, T. (27th Jan. 2017). 'Active sites in heterogeneous ice nucleation—the example of K-rich feldspars'. In: *Science* vol. 355, no. 6323. Publisher: American Association for the Advancement of Science, pp. 367–371.

Kolmogoroff, A. (1941). 'Confidence Limits for an Unknown Distribution Function'. In: *The Annals of mathematical statistics* vol. 12, no. 4. Publisher: Institute of Mathematical Statistics, pp. 461–463.

Komurcu, M., Storelvmo, T., Tan, I., Lohmann, U., Yun, Y., Penner, J. E., Wang, Y., Liu, X. and Takemura, T. (2014). 'Intercomparison of the cloud water phase among global climate models'. In: *Journal of Geophysical Research: Atmospheres* vol. 119, no. 6. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2013JD021119, pp. 3372–3400.

Korolev, A. (1st Sept. 2007). 'Limitations of the Wegener–Bergeron–Findeisen Mechanism in the Evolution of Mixed-Phase Clouds'. In: *Journal of the Atmospheric Sciences* vol. 64, no. 9. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 3372–3375.

Korolev, A., Isaac, G. A., Cober, S. G., Strapp, J. W. and Hallett, J. (2003). 'Microphysical characterization of mixed-phase clouds'. In: *Quarterly Journal of the Royal Meteorological Society* vol. 129, no. 587. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1256/qj.01.204, pp. 39–65.

Korolev, A. and Leisner, T. (19th Oct. 2020). 'Review of experimental studies of secondary ice production'. In: *Atmospheric Chemistry and Physics* vol. 20, no. 20. Publisher: Copernicus GmbH, pp. 11767–11797.

Korolev, A. and Mazin, I. P. (1st Dec. 2003). 'Supersaturation of Water Vapor in Clouds'. In: *Journal of the Atmospheric Sciences* vol. 60, no. 24. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 2957–2974.

Korolev, A. and Milbrandt, J. (2022a). 'How Are Mixed-Phase Clouds Mixed?' In: *Geophysical Research Letters* vol. 49, no. 18. e2022GL099578 2022GL099578, e2022GL099578. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022GL099578.

— (2022b). 'How Are Mixed-Phase Clouds Mixed?' In: *Geophysical Research Letters* vol. 49, no. 18, e2022GL099578.

Kumjian, M. R., Rutledge, S. A., Rasmussen, R. M., Kennedy, P. C. and Dixon, M. (1st June 2014). 'High-Resolution Polarimetric Radar Observations of Snow-Generating Cells'. In: *Journal of Applied Meteorology and Climatology* vol. 53, no. 6. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, pp. 1636–1658.

Köhler, H. (1921). 'Zur Kondensation des Wasserdampfes in der Atmosphäre'. In: *Geophysical Publications* vol. 2, pp. 3–15.

Ladino, L. A., Korolev, A., Heckman, I., Wolde, M., Fridlind, A. M. and Ackerman, A. S. (2017). 'On the role of ice-nucleating aerosol in the formation of ice particles in tropical mesoscale convective systems'. In: *Geophysical Research Letters* vol. 44, no. 3. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL072455, pp. 1574–1582.

Lamb, D. and Verlinde, J. (2011). *Physics and chemistry of clouds*. Cambridge: University Press. XIV, 584.

Lau, K. M. and Wu, H. T. (2003). 'Warm rain processes over tropical oceans and climate implications'. In: *Geophysical Research Letters* vol. 30, no. 24. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2003GL018567.

Lauber, A., Henneberger, J., Mignani, C., Ramelli, F., Pasquier, J. T., Wieder, J., Hervo, M. and Lohmann, U. (15th Mar. 2021). 'Continuous secondary-ice production initiated by updrafts through the melting layer in mountainous regions'. In: *Atmospheric Chemistry and Physics* vol. 21, no. 5. Publisher: Copernicus GmbH, pp. 3855–3870.

Lauber, A., Kiselev, A., Pander, T., Handmann, P. and Leisner, T. (1st Aug. 2018). 'Secondary Ice Formation during Freezing of Levitated Droplets'. In: *Journal of the Atmospheric Sciences* vol. 75, no. 8. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 2815–2826.

Leck, C. and Svensson, E. (6th Mar. 2015). 'Importance of aerosol composition and mixing state for cloud droplet activation over the Arctic pack ice in summer'. In: *Atmospheric Chemistry and Physics* vol. 15, no. 5. Publisher: Copernicus GmbH, pp. 2545–2568.

Liu, C. et al. (2020). 'Anthropogenic Effects on Cloud Condensation Nuclei Distribution and Rain Initiation in East Asia'. In: *Geophysical Research Letters* vol. 47, no. 2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL086184, e2019GL086184.

Lohmann, U., Henneberger, J., Henneberg, O., Fugal, J. P., Bühl, J. and Kanji, Z. A. (2016). 'Persistence of orographic mixed-phase clouds'. In: *Geophysical Research Letters* vol. 43, no. 19. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL071036, pp. 10, 512–10, 519.

Lohmann, U., Lüönd, F. and Mahrt, F. (2016). *An Introduction to clouds: from the microscale to climate*. Cambridge: University Press. XXVI, 391.

Marsland, S. (2014). *Machine learning: an algorithmic perspective*. Second edition. Chapman & Hall/CRC machine learning & pattern recognition series. Boca Raton, FL: Chapman, Hall/CRC, an imprint of Taylor and Francis.

Maturilli, M. and Kayser, M. (1st Oct. 2017). 'Arctic warming, moisture increase and circulation changes observed in the Ny-Ålesund homogenized radiosonde record'. In: *Theoretical and Applied Climatology* vol. 130, no. 1, pp. 1–17.

McCoy, D. T., Hartmann, D. L., Zelinka, M. D., Ceppi, P. and Grosvenor, D. P. (2015). 'Mixed-phase cloud physics and Southern Ocean cloud feedback in climate models'. In: *Journal of Geophysical Research: Atmospheres* vol. 120, no. 18. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015JD023603, pp. 9539–9554.

McInnes, L., Healy, J. and Astels, S. (21st Mar. 2017). 'hdbscan: Hierarchical density based clustering'. In: *The Journal of Open Source Software* vol. 2, no. 11, p. 205.

Mignani, C., Creamean, J. M., Zimmermann, L., Alewell, C. and Conen, F. (23rd Jan. 2019). 'New type of evidence for secondary ice formation at around 15 °C in mixed-phase clouds'. In: *Atmospheric Chemistry and Physics* vol. 19, no. 2. Publisher: Copernicus GmbH, pp. 877–886.

Morrison, H., Curry, J. A. and Khvorostyanov, V. I. (1st June 2005). 'A New Double-Moment Microphysics Parameterization for Application in Cloud and Climate Models. Part I: Description'. In: *Journal of the Atmospheric Sciences* vol. 62, no. 6. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 1665–1677.

Morrison, H., Thompson, G. and Tatarskii, V. (1st Mar. 2009). 'Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes'. In: *Monthly Weather Review* vol. 137, no. 3. Publisher: American Meteorological Society Section: Monthly Weather Review, pp. 991–1007.

Morrison, H., Boer, G. de, Feingold, G., Harrington, J., Shupe, M. D. and Sulia, K. (Jan. 2012). 'Resilience of persistent Arctic mixed-phase clouds'. In: *Nature Geoscience* vol. 5, no. 1. Number: 1 Publisher: Nature Publishing Group, pp. 11–17.

Morrison, H. and Grabowski, W. W. (1st May 2008). 'A Novel Approach for Representing Ice Microphysics in Models: Description and Tests Using a Kinematic Framework'. In: *Journal of the Atmospheric Sciences* vol. 65,

no. 5. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 1528–1548.

Morrison, H. et al. (2020). 'Confronting the Challenge of Modeling Cloud and Precipitation Microphysics'. In: *Journal of Advances in Modeling Earth Systems* vol. 12, no. 8. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001689, e2019MS001689.

Mossop, S. C. (1978). 'The influence of drop size distribution on the production of secondary ice particles during graupel growth'. In: *Quarterly Journal of the Royal Meteorological Society* vol. 104, no. 440. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49710444007, pp. 323–330.

Motos, G. et al. (15th May 2023). 'Aerosol and dynamical contributions to cloud droplet formation in Arctic low-level clouds'. In: *EGUsphere*. Publisher: Copernicus GmbH, pp. 1–25.

Murray, B. J., O'Sullivan, D., Atkinson, J. D. and Webb, M. E. (2012). 'Ice nucleation by particles immersed in supercooled cloud droplets'. In: *Chemical Society Reviews* vol. 41, no. 19, p. 6519.

Mülmenstädt, J., Sourdeval, O., Delanoë, J. and Quaas, J. (2015). 'Frequency of occurrence of rain from liquid-, mixed-, and ice-phase clouds derived from A-Train satellite retrievals'. In: *Geophysical research letters* vol. 42, no. 15. Edition: Mülmenstädt, J., O. Sourdeval, J. Delanoë, and J. Quaas (2015), Frequency of occurrence of rain from liquid-, mixed-, and ice-phase clouds derived from A-Train satellite retrievals, Geophys. Res. Lett., 42, 6502-6509, doi:10.1002/2015GL064604. Place: WASHINGTON Publisher: Blackwell Publishing Ltd, pp. 6502–6509.

Nath, S. and Boreyko, J. B. (23rd Aug. 2016). 'On Localized Vapor Pressure Gradients Governing Condensation and Frost Phenomena'. In: *Langmuir* vol. 32, no. 33, pp. 8350–8365.

Naud, C. M., Booth, J. F. and Genio, A. D. D. (1st Mar. 2014). 'Evaluation of ERA-Interim and MERRA Cloudiness in the Southern Ocean'. In: *Journal of Climate* vol. 27, no. 5. Publisher: American Meteorological Society Section: Journal of Climate, pp. 2109–2124.

Noether, G. E. (1963). 'Note on the kolmogorov statistic in the discrete case'. In: *Metrika* vol. 7, no. 1, pp. 115–116.

Pasquier, J. et al. (1st Dec. 2020). 'Understanding Microphysical Processes in Arctic Mixed-Phase Clouds from Balloon-Borne Holographic Instrument'. In: vol. 2020. Conference Name: AGU Fall Meeting Abstracts ADS Bibcode: 2020AGUFMA011.0007P, A011–0007.

Pasquier, J. T., David, R. O. et al. (14th Nov. 2022). 'The Ny-Ålesund Aerosol Cloud Experiment (NASCENT): Overview and First Results'. In: *Bulletin of the American Meteorological Society* vol. 103, no. 11. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, E2533–E2558.

Pasquier, J. T., Henneberger, J., Ramelli, F., Lauber, A., David, R. O., Wieder, J., Carlsen, T., Gierens, R., Maturilli, M. and Lohmann, U. (12th Dec. 2022). 'Conditions favorable for secondary ice production in Arctic mixed-phase clouds'. In: *Atmospheric Chemistry and Physics* vol. 22, no. 23. Publisher: Copernicus GmbH, pp. 15579–15601.

Pastukhov, A., Styrnal, M. and Carbon, C.-C. (6th Aug. 2020). *History-dependent changes of Gamma distribution in multistable perception.* Pages: 2020.08.06.239285 Section: New Results.

Phillips, V. T. J., Patade, S., Gutierrez, J. and Bansemer, A. (1st Sept. 2018). 'Secondary Ice Production by Fragmentation of Freezing Drops: Formulation and Theory'. In: *Journal of the Atmospheric Sciences* vol. 75, no. 9. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 3031–3070.

Phillips, V. T. J., Yano, J.-I. and Khain, A. (1st June 2017). 'Ice Multiplication by Breakup in Ice–Ice Collisions. Part I: Theoretical Formulation'. In: *Journal of the Atmospheric Sciences* vol. 74, no. 6. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 1705–1719.

Pruppacher, H. and Klett, J. (1997). 'Homogeneous Nucleation'. In: *Microphysics of Clouds and Precipitation.* Ed. by Pruppacher, H. and Klett, J. 2nd rev. and enl. ed. with an introduction to cloud chemistry and cloud electricity. Vol. 18. Atmospheric and Oceanographic Sciences Library. Dordrecht: Kluwer.

Pummer, B. G. et al. (21st Apr. 2015). 'Ice nucleation by water-soluble macromolecules'. In: *Atmospheric Chemistry and Physics* vol. 15, no. 8. Publisher: Copernicus GmbH, pp. 4077–4091.

Ramelli, F., Beck, A., Henneberger, J. and Lohmann, U. (27th Feb. 2020). 'Using a holographic imager on a tethered balloon system for microphysical observations of boundary layer clouds'. In: *Atmospheric Measurement Techniques* vol. 13, no. 2. Publisher: Copernicus GmbH, pp. 925–939.

Raoult, F.-M. (1889). 'Recherches expérimentales sur les tensions de vapeur des dissolutions'. In: *Journal de Physique Théorique et Appliquée* vol. 8, no. 1. Publisher: Société Franccaise de Physique, pp. 5–20.

Saleeby, S. M. and Cotton, W. R. (1st Feb. 2008). 'A Binned Approach to Cloud-Droplet Riming Implemented in a Bulk Microphysics Model'. In: *Journal of Applied Meteorology and Climatology* vol. 47, no. 2. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, pp. 694–703.

Seifert, A. and Beheng, K. D. (1st Feb. 2006). 'A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 2: Maritime vs. continental deep convective storms'. In: *Meteorology and Atmospheric Physics* vol. 92, no. 1, pp. 67–82.

Serreze, M. C., Barrett, A. P. and Cassano, J. J. (2011). 'Circulation and surface controls on the lower tropospheric air temperature field of the Arctic'. In: *Journal of Geophysical Research: Atmospheres* vol. 116 (D7). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2010JD015127.

Shupe, M. D., Daniel, J. S., Boer, G. de, Eloranta, E. W., Kollias, P., Long, C. N., Luke, E. P., Turner, D. D. and Verlinde, J. (2008). 'A focus on mixed-phase clouds'. eng. In: *Bulletin of the American Meteorological Society* vol. 89, no. 10, pp. 1549–1562.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W. and Powers, J. G. (2019). *A Description of the Advanced Research WRF Version 4.*

Smirnov, N. (1939). 'On the estimation of the discrepancy between empirical curves of distribu- tion for two independent samples'. In: *Bulletin of Mathematics University of Moscow* vol. 2, no. 2.

Sotiropoulou, G., Sullivan, S., Savre, J., Lloyd, G., Lachlan-Cope, T., Ekman, A. M. L. and Nenes, A. (4th Feb. 2020). 'The impact of secondary ice production on Arctic stratocumulus'. In: *Atmospheric Chemistry and Physics* vol. 20, no. 3. Publisher: Copernicus GmbH, pp. 1301–1316.

Sotiropoulou, G., Vignon, E., Young, G., Morrison, H., O'Shea, S., Lachlan-Cope, T., Berne, A. and Nenes, A. (19th Jan. 2021). 'Secondary ice production in summer clouds over the Antarctic coast: An underappreciated process in atmospheric models'. In: *Atmospheric Chemistry and Physics* vol. 21, pp. 755–771.

Storelvmo, T., Kristjánsson, J. E., Lohmann, U., Iversen, T., Kirkevåg, A. and Seland, Ø. (Oct. 2008). 'Modeling of the Wegener–Bergeron–Findeisen process—implications for aerosol indirect effects'. In: *Environmental Research Letters* vol. 3, no. 4, p. 045001.

Storelvmo, T. and Tan, I. (21st July 2015). 'The Wegener-Bergeron-Findeisen process – Its discovery and vital importance for weather and climate'. In: *Meteorologische Zeitschrift* vol. 24, no. 4, pp. 455–461.

Sun, Z. and Shine, K. P. (1994). 'Studies of the radiative properties of ice and mixed-phase clouds'. In: *Quarterly Journal of the Royal Meteorological Society* vol. 120, no. 515. Cited by: 159, pp. 111–137.

Tan, I., Storelvmo, T. and Zelinka, M. D. (2016). 'Observational constraints on mixed-phase clouds imply higher climate sensitivity'. In: *Science (American Association for the Advancement of Science)* vol. 352, no. 6282. Place: WASHINGTON Publisher: American Association for the Advancement of Science, pp. 224–227.

Thomson, W. (1872). '4. On the Equilibrium of Vapour at a Curved Surface of Liquid'. In: *Proceedings of the Royal Society of Edinburgh* vol. 7. Publisher: Royal Society of Edinburgh Scotland Foundation, pp. 63–68.

Touloupas, G., Lauber, A., Henneberger, J., Beck, A. and Lucchi, A. (8th May 2020). 'A convolutional neural network for classifying cloud particles recorded by imaging probes'. In: *Atmospheric Measurement Techniques* vol. 13, no. 5. Publisher: Copernicus GmbH, pp. 2219–2239.

Twomey, S. (1974). 'Pollution and the planetary albedo'. In: *Atmospheric Environment (1967)* vol. 8, no. 12, pp. 1251–1256.

Vali, G., DeMott, P. J., Möhler, O. and Whale, T. F. (16th Sept. 2015). 'Technical Note: A proposal for ice nucleation terminology'. In: *Atmospheric Chemistry and Physics* vol. 15, no. 18. Publisher: Copernicus GmbH, pp. 10263–10270.

Vali, G. (1st Apr. 1971). 'Quantitative Evaluation of Experimental Results an the Heterogeneous Freezing Nucleation of Supercooled Liquids'. In: *Journal of the Atmospheric Sciences* vol. 28, no. 3. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 402–409.

Vardiman, L. (1st Nov. 1978). 'The Generation of Secondary Ice Particles in Clouds by Crystal–Crystal Collision'. In: *Journal of the Atmospheric Sciences* vol. 35, no. 11. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 2168–2180.

Vergara-Temprado, J., Miltenberger, A. K., Furtado, K., Grosvenor, D. P., Shipway, B. J., Hill, A. A., Wilkinson, J. M., Field, P. R., Murray, B. J. and Carslaw, K. S. (13th Mar. 2018). 'Strong control of Southern Ocean cloud reflectivity by ice-nucleating particles'. In: *Proceedings of the National*

Academy of Sciences vol. 115, no. 11. Publisher: Proceedings of the National Academy of Sciences, pp. 2687–2692.

Wallace, J. M. and Hobbs, P. V. (1st Jan. 2006). '6 - Cloud Microphysics'. In: *Atmospheric Science (Second Edition)*. Ed. by Wallace, J. M. and Hobbs, P. V. San Diego: Academic Press, pp. 209–269.

Wang, J. et al. (25th Feb. 2022). 'Aerosol and Cloud Experiments in the Eastern North Atlantic (ACE-ENA)'. In: *Bulletin of the American Meteorological Society* vol. 103, no. 2. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, E619–E641.

Wegener, A. (1911). *Thermodynamik der atmosphäre*. J.A. Barth.

Wieder, J., Mignani, C., Schär, M., Roth, L., Sprenger, M., Henneberger, J., Lohmann, U., Brunner, C. and Kanji, Z. A. (9th Mar. 2022). 'Unveiling atmospheric transport and mixing mechanisms of ice-nucleating particles over the Alps'. In: *Atmospheric Chemistry and Physics* vol. 22, no. 5. Publisher: Copernicus GmbH, pp. 3111–3130.

Yano, J.-I., Phillips, V. T. J. and Kanawade, V. (2016). 'Explosive ice multiplication by mechanical break-up in ice–ice collisions: a dynamical system-based study'. In: *Quarterly Journal of the Royal Meteorological Society* vol. 142, no. 695. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2687, pp. 867–879.

Young, I. T. (1st July 1977). 'Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources.' In: *Journal of Histochemistry & Cytochemistry* vol. 25, no. 7. Publisher: Journal of Histochemistry & Cytochemistry, pp. 935–941.

Appendices

APPENDIX A

KS-Clustering Validation with Synthetic Data

KSC was evaluated using a synthetically produced dataset containing three different gamma distributions with known shape and scale parameters. The synthetic data consisted of 50 samples randomly drawn from each of the gamma distributions defined by the shapes 10, 20, 30 and scales 0.6, 0.5, 0.4 μm, respectively. Additionally, some noise was included represented by another 50 samples drawn from separate gamma functions of randomly chosen values for shapes and scales, within the range of 5 to 60 and 0.3 to 0.9 μm, respectively. The test data thus consists of a total of 200 'holograms' which are randomly shuffled around to change the order of appearance.

Figures A.1 and A.2 show the results of running KSC on the synthetic dataset using the *NOFIT* and *FIT* methods, respectively, as described in Section 4.2.1. Each is initialised with 10 ensemble members, and HDBSCAN with `min_cluster_size`=10 and `min_samples`=10 is applied for clustering. Notably, KSC successfully classified all of the non-noise distribution samples and attributed them to each of their respective parent classes for both *NOFIT* and *FIT*. Most of the noise samples were correctly labelled as noise in both cases, although a total of 14 and 5 samples for *NOFIT* and *FIT*, respectively, were incorrectly included in some of the other classes. However, it is worth noting that the noise samples are in fact randomly drawn from a shape and scale range, which includes the location of the actual classes. Therefore, the noise can by chance get values similar to the non-noise distributions and so legitimately belong to that class.

Moreover, it is evident from Figures A.1d and A.2d that KSC tends to classify similar holograms along certain isolines of scales and shapes of some relation. The incorrectly classified noise samples illustrate this dependence clearly as these are all located on extensions to the isolines for each of the respective classes. As shown for instance by Pastukhov et al., 2020, these isolines in the shape-scale-diagram represent states of constant mean of the gamma distribution - the mean can stay constant following an increase in the scale parameter provided the shape is decreased accordingly. This is an indication of the sensitivity of KSC towards the distribution location. A shift in the mean between two sample distributions causes a large increase in the KS-statistic between them, and thus a greater separation in the KS-space. From the synthetic dataset it appears

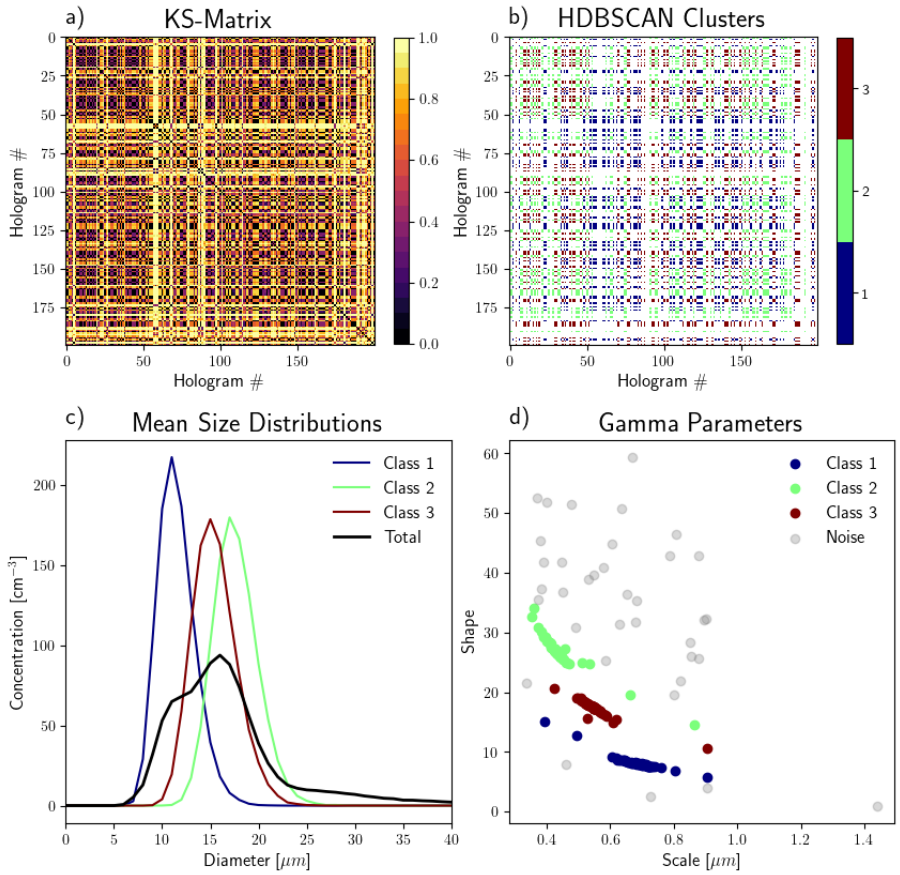

Figure A.1: Results of the KS-clustering of the synthetic data, using the *NOFIT* method and HDBSCAN with `min_sample_size`=20 and `min_samples`=20. The KS-matrix (panel a) shows the mean KS-score (colors) between all holograms as indicated by the axes. The cluster matrix (panel b) shows the class (colors) that each pair of holograms are predicted to belong to. The mean size distributions (panel c) and the fitted shape and scale parameters (panel d) are computed for each of the identified classes individually. Note that the class numbers are arbitrary and determined only by the random initiation of HDBSCAN.

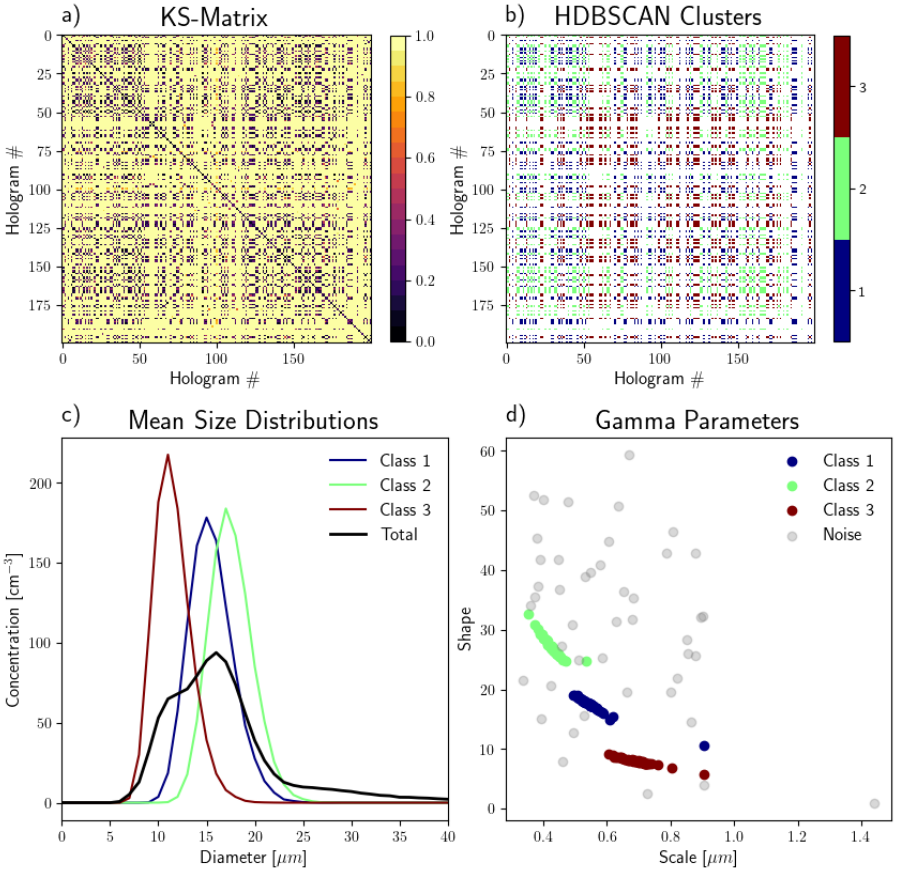

Figure A.2: Same as Figure A.1 but using *FIT*.

that *NOFIT* is perhaps more sensitive to the mean compared to *FIT* as the isolines of constant means are extended further away from the classes in the shape-scale-diagram in Figure A.1d.

APPENDIX B

Code Availability Statement

All analysis in this thesis was carried out using the programming language Python. All computer code developed during this thesis is publicly available on GitHub through the following link:

https://github.com/sldamman/Master.git

This includes:

- The spatial and temporal algorithms, developed for analysis of cloud-top phase heterogeneity in output from the WRF, but is easily generalizable to include also output from other models of reanalysis products, and likely also to observational data for instance from satellite.
- The KS-clustering algorithm, inspired by Allwayin et al., 2022 and adapted here to work also with discretized holographic data of higher spatial resolution.
- Statistical analysis methods, applied primarily in Section 6, as well as all visualization and plotting procedures.

The computer code and algorithms published on the GitHub repository may be maintained and updated following progress and improvements in the future.