Clouds, Aerosols and Radiation

A Meteorology and Satellite Driven Analysis of Effective Radiative Forcing from Aerosol-Cloud Interactions

W.S.J. Kroese

Delft University of Technology | Netherlands Institute for Space Research

TU Delft

SRON
Netherlands Institute for Space Research

Clouds, Aerosols and Radiation

A Meteorology and Satellite Driven Analysis of Effective Radiative Forcing from Aerosol-Cloud Interactions

by

W.S.J. Kroese

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday November 30, 2023 at 14:45.

Student number:	5434947	
Project duration:	March 1, 2023 – November 30, 2023	
Thesis committee:	Dr. sc. F. Glassmeier,	TU Delft, chair
	Dr. B. van Diedenhoven,	SRON
	P. Alinaghi Msc.,	TU Delft
	Dr. M. A. Schleiss,	TU Delft

Cover: NASA Worldview image of Eastern Australia during bushfires on the 19th of November 2019
Style: TU Delft Report Style, with modifications by Daan Zwaneveld

TU Delft
SRON
Netherlands Institute for Space Research

Preface

"If a sword had memory, it might be grateful to the forge fire, but never fond of it."

-Rand Al'Thor from
Winter's Heart by Robert Jordan
Book 9 of the Wheel of Time

Dear reader,

Before you lies the end product of my master thesis research. Over the past months I have worked on this thesis at SRON and TU Delft as a part of my degree in Geoscience and Remote Sensing. This period has been my conceptual forge fire, with many ups and downs. Looking back I would have done some things differently, but I suppose that was the point all along: to gain wisdom and understanding through perseverance and routine.

This thesis would not have been possible without the support of some wonderful people. I would like to thank my supervisors for all the help they offered me. Bastiaan, thank you for all the good advice and helping me throughout the entire process. You were always ready to offer help and support. Our meetings always left me with new ideas to try. Franziska, thank you for all the feedback and the conversations we had. They always gave me new inspiration and thanks to you I started thinking more about microphyiscs instead of the global view that satellites give. Pouriya, thanks for being so passionate about clouds. I really loved listening to you talking (and sometimes chiming in) during all the meetings we had. Marc, thank you for all your feedback and good comments. Thanks to you I realised that I should be more attentive of statistics, which I will be sure to take with me in the future. This thesis also would not have been possible without the unwavering support of many lovely, kind, outstanding, fabulous, funny people, whom I may call family and friends. Thank you for being you.

Throughout the process of my thesis I must have read 8 books of the Wheel of Time by Robert Jordan and in a weird way the series has become interwoven into this work. The only physical references you will find to the books are on this page, but for me the two stories, my thesis and Jordan's series, are entwined. It is fitting at the end of my thesis journey that I am also starting the last book of the series. During the writing process, I have listened *a lot* to two albums: the Complete Partitas and the French Suites by J.S. Bach, played by Yuan Sheng. Sheng might have earned a cup of coffee from me streaming his albums.

With love,
W.S.J. Kroese
Leiden, November 2023

Abstract

Uncertainty in the radiative forcing from anthropogenic activities since the Industrial Revolution is dominated by how clouds respond to aerosol. Climate projections are limited by this uncertainty. The cloud response to aerosol is influenced by the meteorological conditions of the atmosphere wherein the cloud is suspended. Understanding the covariation between meteorological state and cloud response to aerosol is a path forward to improve our understanding of aerosol effects on the climate. In this thesis we study aerosol-cloud interactions while controlling for meteorology using clustering techniques. This allows us to study the interactions per meteorological regime and gain deeper understanding of the effect of meteorology on aerosol-cloud interactions. Cloud-controlling factors are clustered using k-means clustering. Six meteorological clusters are found and satellite observations over ocean between -60° and 60° latitude are used to study cloud response to changes in aerosol concentrations per cluster. The choice of clustering does not create significant variability in the sensitivity and the radiative forcing of the cloud albedo effect, but the sensitivity of cloud liquid water path and cloud fraction adjustments do show variability between clusters. This indicates that controlling for meteorology is specifically important for the adjustments to the cloud albedo effect. Our results show that the effective radiative forcing from aerosol-cloud interactions (ERF_{aci}) over our study domain since 1850 is -1.0 W m^{-2} with a 90% confidence interval of $[-1.6, -0.48]$ W m^{-2}. There are variations in the forcing estimates depending on the number of clusters, but this signal is small compared to other sources of uncertainty. Our findings corroborate recent findings and present a novel method to control for meteorological covariation using cluster analysis.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
ACI	Aerosol-cloud interactions
AI	Aerosol index
AOD	Aerosol optical depth
CCN	Cloud condensation nuclei
CDNC	Cloud droplet number concentration
CER	Cloud drop effective radius
CF	Cloud fraction
COT	Cloud optical thickness
CTP	Cloud top phase
CTT	Cloud top temperature
ERF	Effective radiative forcing
GCM	Global climate model
INP	Ice nucleation particles
LWP	Liquid water path
NH	Northern hemisphere
NRMSRE	Normalised root-mean-squared error
PI	Pre-industrial
PD	Present-day
RH	Relative humidity
RFO	Relative frequency of occurrence
SH	Southern hemisphere

Symbols

Symbol	Definition	Typical Unit
C	Cloud fraction	1
EIS	Estimated inversion strength	[K]
ERF_{aci}	Effective radiative forcing from aerosol-cloud interactions	[W m^{-2}]
$F^{\downarrow}$	Downwelling shortwave flux	W m^{-2}
f_{liq}	Mean liquid cloud fraction	[1]
f_{ice}	Mean ice cloud fraction	[1]
IRF_{aci}	Instantaneous radiative forcing from aerosol-cloud interactions	[W m^{-2}]
$\mathcal{L}$	Liquid water path	[g m^{-2}]
LCL	Lifted condensation level	[m]
LTS	Lower tropospheric stability	[K]
N_d	Cloud droplet number concentrations	[cm^{-3}]
N_{ccn}	CCN column number concentration	[cm^{-2}]
p	Pressure	[hPa]
r_e	Effective cloud droplet radius	[μm]
$RF_{\mathcal{L}}$	Radiative forcing from liquid water path adjustments	[W m^{-2}]
RF_C	Radiative forcing from cloud fraction adjustments	[W m^{-2}]
T	Temperature	[K]
UV	Wind-speed	[m s^{-2}]
z	Height	[m]
α_{cld}, α_{clr}	Cloud (cld) and clear-sky (clr) albedo	[1]
$\beta_{\ln N_d - \ln N_{ccn}}$	Susceptibility of N_d to changes in N_{ccn}	1
$\beta_{C - \ln N_d}$	Susceptibility of CF to changes in N_d	1
$\beta_{\ln \mathcal{L} - \ln N_d}$	Susceptibility of LWP to changes in N_d	1
$\Gamma_m(T, p)$	Moist-adiabatic potential temperature gradient	[K]
θ	Potential temperature	[K]
λ	Latitude	[° N]
τ_c	Cloud optical thickness	1
ϕ	Longitude	[° E]
Ω	Vertical pressure velocity	[Pa s^{-1}]

List of Figures

1

Introduction

At any moment, approximately 68% to 73% of the Earth's surface is covered by clouds (Stubenrauch et al. 2013). Clouds affect the climate system by reflecting sunlight and trapping heat, altering the energy budget of the Earth. Despite decades of research dedicated to their study, the diverse and complex nature of clouds has left many uncertainties and open questions to be answered. One of those uncertainties stems from their interactions with atmospheric aerosols: small solid particles suspended in the atmosphere.

1.1. Earth's energy budget

The Earth's energy budget is the balance between incoming and outgoing radiation (Figure 1.1). The incoming solar radiation (shortwave flux, yellow arrows) is well known as the solar irradiation is close to constant. A part of the shortwave flux is reflected, primarily by clouds, but also by atmospheric particles and bright surface areas like snow and ice. The remainder of the flux is absorbed by the surface and atmosphere. The Earth's surface emits energy in the form of infrared radiation (longwave flux, purple arrows), and by heating the atmosphere through sensible and latent heat. Sensible heat is the conductive heat flux from the Earth's surface to the atmosphere, while latent heat is the energy transfer due to evaporation or transpiration of water (Farmer et al. 2013).

When the incoming and outgoing fluxes balance each other, the system will be in equilibrium and the global mean temperature will be constant. When the balance between incoming solar radiation and outgoing heat radiation is altered, either by human activities like the release of carbon dioxide or natural events like volcanic eruptions, a *radiative forcing* is introduced. The perturbed system will adjust to form a new equilibrium, which can result in long-term changes in temperature and climate patterns (Bellouin et al. 2020).

1.2. Aerosols

Aerosols are fine solid and/or liquid particles suspended within a gas. Formally, an aerosol includes both the solid/liquid particles and the gas within which these are suspended. The atmospheric aerosol (referred to as aerosol hereafter) is the suspension of fine solid particles in Earth's atmosphere (Prospero et al. 1983). Typical sizes range between a few nanometers and hundreds of microns. The distribution of aerosols shows many peaks which have been classified as modes. Aerosols can either be directly emitted into the atmosphere or they can form from precursor compounds through nucleation (Prospero et al. 1983). The distribution of aerosol is heterogeneous in both space and time, because of the heterogeneity of sources and aerosol removal. Aerosols are removed from the atmosphere on a timescale of weeks by deposition on the Earth's surface and by being washed out after mixing with water droplets (J. Williams et al. 2002; Raes et al. 2000).

Aerosols exert influence on the climate system through their ability to both absorb and scatter incoming shortwave radiation, changing the planetary albedo, which is known as *aerosol radiation interaction* (ARI; Zelinka et al. 2014). Different types of aerosols interact differently with their surroundings. For example, black carbon, a type of aerosol emitted by burning biomass and fossil fuels, heats the

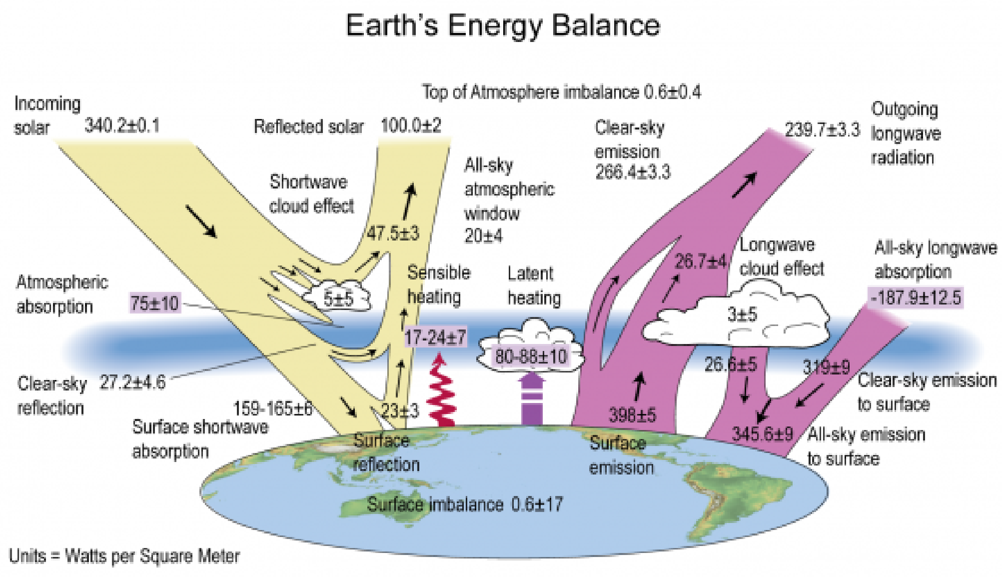


Figure 1.1: The energy balance of Earth. Yellow lines indicate shortwave radiation and purple lines indicate longwave radiation. Aerosols directly modify the clear-sky reflection and the shortwave/longwave cloud effect. (Figure is modified from Stephens et al. 2012)

atmosphere by absorbing radiation and by deposition on snow and ice which reduces the planetary albedo (Ramanathan et al. 2008). Sulphate aerosols cool the atmosphere by reflecting radiation (Charlson et al. 1991).

Additionally, aerosols can act as *cloud condensation nuclei* (CCN) and as *ice nucleating particles* (INP). A cloud is composed of liquid droplets and/or ice crystals. To form cloud droplets, there must be a surface for water vapour to condense upon, which are aerosols. Size and composition determine if an aerosol can act as CCN, though Dusek et al. (2006) show that it is primarily the size distribution that determines the CCN fraction. There are multiple pathways for ice crystals to form. Homogeneous freezing (freezing without INP) can occur when temperatures reach -38° C. Between 0 and -38° C super-cooled droplets can freeze through heterogeneous ice nucleation which is catalysed by INP. The study of INP, ice crystal formation, and the radiative effect of changes in concentrations are fascinating, but beyond the scope of this thesis, primarily because liquid clouds have the largest contribution to the energy budget (L'Ecuyer et al. 2019; Wood et al. 2012). See Burrows et al. (2022) for a review on the current state of knowledge on INP.

As clouds are dependent on aerosols, changing aerosol concentrations will affect clouds and in turn affect the Earth's energy budget. The focus for the remainder of this document will be on the aerosol-cloud interactions (ACI) in warm liquid clouds. We will not study ARI of aerosols, as ACI contributes more to the anthropogenic forcing on the climate (Forster et al. 2021).

1.3. Aerosol-Cloud Interactions

By acting as CCN, aerosols change the radiative, microphysical, and macrophysical properties of clouds. Anthropogenic emissions since the Industrial Revolution have led to an increase in aerosol concentrations (Forster et al. 2021). Increased CCN concentrations entering a cloud cause an increase in *cloud droplet number concentrations* (CDNC or N_d; Seinfeld et al. 2016). The mean radius of the droplets in the polluted cloud will be smaller if the liquid water content stays constant. This causes the cloud albedo to increase due to a higher reflective surface area which creates a negative radiative forcing (i.e. a cooling effect; Twomey 1977). This effect is known as the *cloud albedo effect*, or Twomey effect. Direct observations of the Twomey effect are ship tracks. As a freighter ship sails under a cloud field, the aerosol emitted by the ships' exhaust act as CCN which increases CDNC and brightens the clouds, as seen in Figure 1.2. These were first observed in 1966 (Conover 1966) and were the starting point for the study of aerosol-cloud interactions.

As N_d increases and r_e decreases, the cloud adjusts to the new conditions, resulting in possible changes in the liquid water path (LWP) and the cloud fraction (CF). The LWP is the vertically integrated liquid water content from base to the top of the cloud. These changes in LWP and CF may induce a significant radiative forcing. The decrease in r_e may cause precipitation to be suppressed, but also increases cloud top entrainment and evaporation (Albrecht 1989). By suppressing precipitation, cloud lifetime increases. These effects are known as adjustments to the cloud albedo effect.

The adjustments to the cloud albedo effect are poorly constrained in global climate models (GCM) and most of the uncertainty in estimates of the radiative forcing induced by anthropogenic activities stems from aerosol-cloud interactions. This uncertainty hampers quantification of the sensitivity of the climate system to increased greenhouse gas emission. Improving the constraints on aerosol-cloud interactions is important to improve climate projections. The current expert consensus is that aerosol-cloud interactions have induced a negative radiative forcing (i.e. a cooling effect) over the industrial period (1750, 1830 or 1850 to present day), but the magnitude is highly uncertain (Bellouin et al. 2020).

Two definitions of radiative forcings are used by the IPCC: the instantaneous radiative forcing ($\mathrm{IRF}_{\mathrm{aci}}$) and the effective radiative forcing ($\mathrm{ERF}_{\mathrm{aci}}$), where we use subscripts to distinguish between ACI and ARI (Forster et al. 2021). The $\mathrm{IRF}_{\mathrm{aci}}$ is the forcing at the top of the atmosphere (TOA) caused by the cloud albedo effect. The $\mathrm{ERF}_{\mathrm{aci}}$ is the $\mathrm{IRF}_{\mathrm{aci}}$ combined with the TOA radiative forcing from rapid adjustments. In essence, $\mathrm{ERF}_{\mathrm{aci}}$ offers a more comprehensive view of the long-term consequences of a given driver by considering how the climate system adapts and responds to the initial perturbation, but quantifying the radiative effect of adjustments is non-trivial.

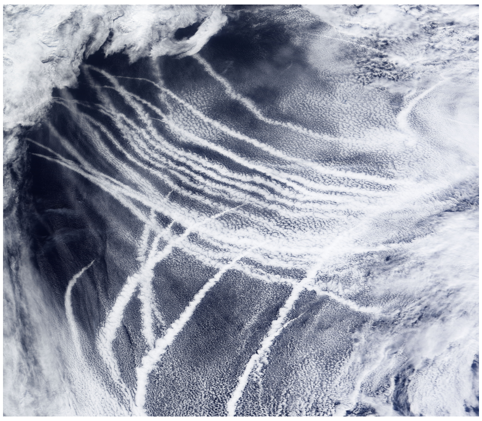


Figure 1.2: A MODIS satellite image showing multiple ship tracks. These bright clouds are a direct observation of the cloud albedo effect first observed by Twomey 1977. (Figure fom NASA/Goddard Space Flight Center Scientific Visualisation Studio)

One key issue is the dependence of both clouds and aerosols on meteorology (e.g. humidity or atmospheric stability). Meteorology is the key driver in cloud formation, so changes in meteorology will affect clouds and aerosol-cloud interactions. This means that correlations between aerosol and cloud parameters will be mediated by meteorological covariation (Stevens et al. 2009), which makes it difficult to determine the causal relationship between aerosol and cloud. Accounting for these correlations can be done by techniques like sampling data, using reanalysis data, or statistical methods (Gryspeerdt et al. 2016; Koren et al. 2010; Wall et al. 2022; M. W. Christensen et al. 2017). These methods aim to define specific cloud regimes based on meteorological data and sub-sampling data into these regimes. Another potential path forward is improving our knowledge of the correlations between cloud, aerosol, and meteorology. Feingold et al. (2016) show that the co-variability of meteorology and aerosol affects $\mathrm{ERF}_{\mathrm{aci}}$ using numerical simulations. If we improve our understanding of the meteorological condition where the aerosol has the largest impact, we could start quantifying the rate of occurrence of those conditions instead of directly constraining the effect (Stevens et al. 2009; Mülmenstädt et al. 2018). This method would focus more on constraining the different relationships, instead of sub-sampling observations.

Recently, M. W. Christensen et al. (2022) summarised most work on opportunistic experiments, i.e. using emission events like volcanoes, ship tracks, industry, natural fires, and temporal trends to constrain aerosol-cloud interactions. Using these experiments help our process understanding, but implications for the global scale are difficult to make. Glassmeier et al. (2021) showed that ship track studies overestimate cooling by LWP adjustments due to the temporal evolution of the ship track which points to the fundamental 'scale problem' in ACI research: different temporal and spatial scales for both processes and observations make quantification and comparisons difficult (Mülmenstädt et al. 2018). Cloud processes operate at the micro- and meso-scale while aerosol-cloud interactions affect global circulation that feedback to cloud processes. GCMs and satellite data can not resolve cloud-process scales (~10 m), and comparing models and observations without considering the scales can lead to significant bias (Schutgens et al. 2016).

1.4. Satellite and model based studies

The study of aerosol-cloud interactions has yielded a wealth of research findings, spanning a wide range of methodologies. These include in-situ measurements (McComiskey et al. 2009; Shupe et al. 2016; Jia et al. 2019; D'Alessandro et al. 2023), modelling efforts spanning from micro (Y. C. Chen et al. 2011; Yamaguchi et al. 2019) to global scales (Gryspeerdt et al. 2020), laboratory experiments (Stratmann et al. 2004; Wagner et al. 2009), and satellite-based investigations (X. Ma et al. 2014; Hasekamp et al. 2019; Wall et al. 2023).

Satellite studies go back two decades when Nakajima et al. (2001) showed a correlation between satellite derived cloud optical thickness (COT) and aerosol concentrations. Since then, technology has improved significantly, but inferring $\mathrm{ERF_{aci}}$ is still difficult due to the issues mentioned above. Satellite studies historically estimated lower forcings from the cloud albedo effect compared to global climate models. This has been partly overcome by using more comprehensive proxies for CCN concentrations (Hasekamp et al. 2019), and using different methodologies to constrain the effect (I. L. McCoy et al. 2020; Wall et al. 2023). Estimates for the radiative forcing from satellite studies range from -0.4 to $-1.4\ \mathrm{W\ m^{-2}}$ for IRF_{aci} (Quaas et al. 2008; Lebsock et al. 2008; Bellouin et al. 2013; X. Ma et al. 2014; Gryspeerdt et al. 2017; I. L. McCoy et al. 2020; Wall et al. 2023). The IPCC reports a range of $-0.7 \pm 0.5\ \mathrm{W\ m^{-2}}$ for satellite based estimates with medium confidence (Forster et al. 2021). Climate model studies produce values in the range of -1 to $-1.8\ \mathrm{W\ m^{-2}}$ (Bellouin et al. 2020). In AR6, the IPCC does not report an $\mathrm{IRF_{aci}}$ range from model based studies. $\mathrm{ERF_{aci}}$ from model-based studies has a range of $-1 \pm 0.8\mathrm{W\ m^{-2}}$ with medium confidence, but many important micro- and mesoscale processes are not captured by GCMs (Forster et al. 2021).

Whereas most studies rely upon observations of aerosol optical depth, Hasekamp et al. (2019) introduced polarimetric retrievals of CCN concentrations which leads to a IRF_{aci} estimate more in line with modelling studies: -1.14 $\mathrm{W\ m^{-2}}$ with a 90% confidence interval of -0.84 to $-1.72\ \mathrm{W\ m^{-2}}$, which is in line with the result from Bellouin et al. (2020), though only if $\mathrm{ERF_{ari}}$ and the adjustments are small. The cloud albedo effect is considered for different geographical regions, as was first introduced by Quaas et al. (2008).

Recently, Wall et al. (2022) estimated ERF by adjusting for meteorology with cloud controlling factor analysis, introduced by Scott et al. (2020). Their estimate of $\mathrm{ERF_{aci}}$ is $-1.16\pm0.48\ \mathrm{W\ m^{-2}}$, which is in line with the result of Bellouin et al. (2020). Y. Chen et al. (2022) argued that the CF adjustments dominate the $\mathrm{ERF_{aci}}$, whereas previous studies found that the cloud albedo effect was the largest contributor.

To retrieve CDNC using satellite measurements, the adiabatic assumption is used. In thermodynamics, an adiabatic process is a process in which there is no exchange of heat or mass with the environment. In the atmosphere, when an air parcel rises adiabatically it will expand due to the decreasing pressure. This expansion will lead to adiabatic cooling. Satellites retrieve cloud parameters at cloud top and then the cloud base parameters are estimated assuming adiabatic (or subadiabatic) ascend (Grosvenor et al. 2018). The main assumptions are: CDNC is vertically constant in the cloud and the liquid water contents of clouds are a constant fraction of the expected value assuming adiabatic uplift. There are common sources of bias and uncertainty with satellite based studies, some of which due to the adiabatic assumption. We touch upon a few of these uncertainty sources.

- All studies rely on a proxy of aerosol concentration. Past studies often used aerrosol optical depth (AOD) which often leads to underestimation of the radiative forcing. This can partly be explained by the aerosols that can not act as CCN affect AOD, and by the humidity dependence of AOD (Stier 2016). Using aerosol index (AI) or CCN proxies based on polarimetric measurements, partly overcomes this issue, but both create large uncertainties at low aerosol concentrations (Hasekamp et al. 2019; Gryspeerdt et al. 2017; Stier 2016)
- To correlate aerosol and cloud quantities, retrievals are needed in the same atmospheric column, as one is interested in aerosol properties at cloud base. However, aerosol properties can not be retrieved in a cloudy pixel. The common method is to retrieve aerosol information next to clouds and use those retrievals as estimate of the aerosols below the clouds. It is unclear, however, if the aerosol information next to clouds is representative of the aerosols at cloud base (Gryspeerdt et al. 2015).
- Clouds can cast shadows and can scatter sunlight which contaminates neighbouring pixels. This leads to an error in the retrieval of CDNC (Grosvenor et al. 2018). These effects are generally referred to as '3D effects'.

- Cloud properties are retrieved at cloud top height. Then the adiabatic assumption is used to determine CDNC from cloud optical depth and cloud effective radius at cloud top. These assumptions can lead to significant biases, which can be partially overcome by careful filtering (Grosvenor et al. 2018)

The estimation of $\mathrm{ERF}_{\mathrm{aci}}$ is in reference to a pre-industrial (PI) atmosphere. Human activities have changed the aerosol load, and this has induced an $\mathrm{ERF}_{\mathrm{aci}}$. However, the lack of knowledge on PI aerosol concentrations and distributions creates large uncertainties in our best estimates (Carslaw et al. 2017). Satellite based studies generally rely on GCMs to estimate the change in CCN concentrations, but inferring PI aerosol emissions and sinks from present-day (PD) conditions is difficult. Most of the uncertainty come from the natural aerosol emissions since 1750 (Carslaw et al. 2013; D. T. McCoy et al. 2015). Understanding the atmosphere in pristine conditions is therefore an important step forward.

In the literature there are different definitions of 'pristine conditions', based either on number concentrations, prevalence of carbon monoxide or black carbon, the location on the globe, or combinations of multiple parameters. Recent studies attempt to help constrain this by observing the conditions in the Southern Ocean (SO). Hamilton et al. (2014) used models to find regions where finding pristine near-PI conditions is most likely. There are also field campaigns (e.g. Schmale et al. 2019) aimed at improving our understanding of the conditions in the SO and the aerosol-cloud interactions there. One potential problem is that remote oceans are not necessarily a good proxy for the PI aerosol concentrations over the entire globe, as remote oceans are less affected by events like volcanic eruptions or wildfires. Therefore model-based studies also investigate process differences between PI and PD atmospheres (e.g. Hamilton et al. 2014; Nishant et al. 2017).

1.5. The Afternoon Constellation

One unique group of satellites is the Afternoon Constellation (or A-Train). This is a formation of Earth observation satellites following each other closely on the same orbit. This allows all the satellites to observe similar features on the Earths surface as the time difference between satellites is in the order of 100 seconds. They have an equator crossing time of around 13:30 local time. As of today, the A-Train consists of 3 satellites: OCO-2, GCOM-W2, and Aura. Satellites that used to fly in the formation are: PARASOL, CloudSat, Aqua, and CALIPSO. All satellites have their own scientific goals, but the synergy between them allows for more comprehensive studies.

One such synergies is Aqua and PARASOL. The Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on Aqua can provide cloud information, like N_d, while the POLarization and Directionality of the Earth's Reflectances (POLDER) instrument which flew on PARASOL, can provide proxies for CCN concentrations using polarimetric retrievals (Hasekamp et al. 2011; Stap et al. 2015). Since the end of the PARASOL mission, there has been a gap of polarimetric measurements from space. The APS instrument on the Glory satellite was supposed to the next aerosol polarimeter, but it was destroyed due to a launch failure (Mishchenko et al. 2007). PACE, which is planned to launch on the 30th of January 2024, carries two spectral polarimeters (SPEXone and HARP2) which will greatly improve our capabilities to retrieve aerosol information, including CCN concentrations, as well as cloud properties (Gorman et al. 2019).

1.6. Research question

From this introduction we pose three research questions that we aim to answer in this thesis.

- How do clusters of cloud-controlling factors correlate with cloud parameters?
- Can we reduce uncertainty in the $\mathrm{RF}_{\mathrm{aci}}$ estimate from Hasekamp et al. 2019 when controlling for meteorology using clustering techniques?
- How do the aerosol-cloud interactions and cloud adjustments differ per cluster?

2
Data

In this chapter the data and data processing steps are presented. All datasets are openly accessible. Multiple datasets for the year 2006 are used and combined for this study. All data are gridded to a 1° latitude by 1° longitude grid. At the end of the chapter there is a flowchart to give an overview of the different datasets and processing steps (Figure 2.3).

2.1. Satellite data

Data from three satellite instruments are used: MODIS and CERES, both flying on Aqua (EOS PM-1), and POLDER which flew on the PARASOL satellite from 2004 to 2013.

The MODIS Collection-6 data is used to analyse cloud parameters, specifically the level 3 daily 1 degree gridded data (MOD08_D3; Platnick et al. 2017a; Platnick et al. 2017b). The cloud effective radius (r_e) and cloud optical thickness (τ_c) are used to retrieve cloud droplet number concentrations (CDNC or N_d). This is done using the adiabatic approximation (Quaas et al. 2006):

$$N_d = \gamma f(T) \tau_c^{\frac{1}{2}} r_e^{-\frac{5}{2}}, \tag{2.1}$$

where $\gamma = 1.37 \cdot 10^{-5} m^{-\frac{1}{2}}$, τ_c is unitless, and r_e is in micron. CDNC depends on the condensation rate, which in turn depends on temperature and pressure for a parcel ascending under moist-adiabatic conditions. We use a temperature dependent condensation rate $f(T) = 0.0192T - 4.293$, as the condensation rate depends more strongly on the temperature than the pressure (Grosvenor et al. 2018). In the formula, the temperature (T) at cloud base is estimated by the maximum cloud top temperature in a MODIS grid box (Gryspeerdt et al. 2016). To retrieve meaningful values of r_e and τ_c, we limit the data to points with $r_e > 4$ μm, $\tau_c > 4$, solar zenith angle < 65°, and sensor zenith angle < 41° (see section 6 of Grosvenor et al. 2018).

The CERES SSF1deg daily product (Doelling et al. 2013) is used to analyse the radiative effect. This data provides shortwave radiation and albedo estimates, aggregated to daily averages on a 1° by 1° grid. CERES also provides cloud parameters based on MODIS measurements, but as the processing differs from the processing of the MODIS team, we do not use the cloud products from the CERES team.

We use POLDER-3 aerosol retrievals from the SRON RemoTAP aerosol retrieval algorithm (Hasekamp et al. 2011; Stap et al. 2015). The RemoTAP algorithm retrieves the aerosol loading N_a, effective radius r_{eff}, and effective variance v_{eff} for the fine and coarse mode, assuming log-normal bi-modal size distribution. In addition, complex refractive indices and the fraction of spherical particles is inferred. N_{ccn} is computed from the fine and coarse mode for particles with $r_{eff} > 0.15$ μm, because size primarily determines if a particle can act as CCN (Hasekamp et al. 2019; Dusek et al. 2006). Only scenes where the spherical aerosol amount is higher than 90% are considered. This is done to remove scenes where there are higher concentrations of hydrophobic mineral dust. The spherical fraction is also a proxy for particle hygroscopicity, as hygroscopic particles will absorb water and form haze particles which are more spherical than non-hygroscopic particles. Hasekamp et al. (2019) performed many validation steps and provide an error estimate of $0.20 \cdot N_{ccn} + 4 \cdot 10^6$ cm^{-2}. Typical values for N_{ccn} are $10^6 - 10^8$ cm^{-2}. Retrievals are done using pixels of 18×18 km^2 and gridded to a 1° to 1° grid. Figure 2.1 shows the

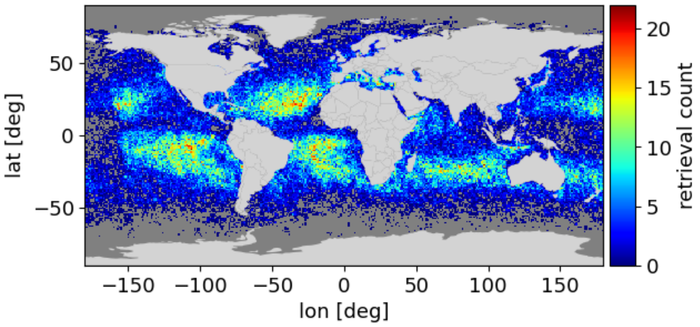


Figure 2.1: Global map showing the amount of retrievals of both N_{ccn} and N_d per grid box. The grey colour represents grid boxes where there are no data points.

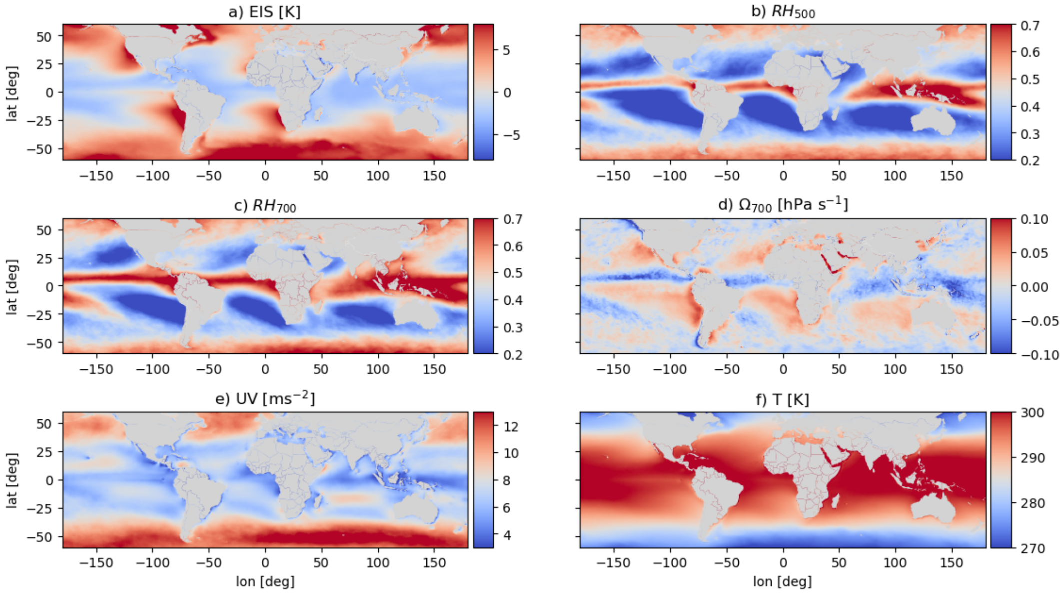


Figure 2.2: Yearly means of the MERRA-2 cloud-controlling factors.

number of retrievals per grid box on a global map. The distribution is spread over the entire domain, but there are certain areas where there are little to no retrievals. This could be caused by high amounts of cloud cover, high amounts of ice clouds, or low amounts of cloud cover. N_{ccn} is only retrieved for cloud-free pixels, but as this dataset is made to relate N_{ccn} to N_d, grid boxes with no clouds are discarded, even though we would be able to retrieve N_{ccn} in these grid boxes.

2.2. MERRA-2 reanalysis data

To analyse meteorological conditions, data from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2; Gelaro et al. 2017) is used. Reanalysis data is model data which is heavily constrained by observations and it provides the most comprehensive picture of meteorology. The data has a spatial resolution of 0.5 by 0.625 degrees latitude and longitude, on 42 different pressure levels. The temporal resolution is 3 hourly.

A subset of parameters are chosen that affect warm clouds, known as cloud-controlling factors (Klein et al. 2018). These parameters are estimated inversion strength (EIS), vertical pressure velocity at 700 hPa (Ω_{700}), relative humidity at 500 en 700 hPa (RH_{700} and RH_{500}), surface temperature, and surface wind-speed. The yearly mean values per grid box of these parameters are shown in Figure 2.2. EIS is a measure of stability based on potential temperature, as defined by Wood et al. (2006). The formal definition of EIS is:

$$EIS = LTS - \Gamma_m^{850}(z_{700} - LCL), \tag{2.2}$$

where LTS is the lower tropospheric stability, $\Gamma_m(T,p)$ is the moist-adiabatic potential temperature gradient, z_{700} is the height at 700 hPa. LTS is defined as the potential temperature difference between 700 hPa and the surface (Klein et al. 1993):

$$LTS = \theta_{700} - \theta_0. \tag{2.3}$$

The LCL is computed using the exact expression derived by (Romps 2017) assuming constant heat capacities. The equations and constants, together with the calculation of the moist-adiabatic potential temperature gradient, can be found in Appendix A.

2.2.1. Interpolation

The MERRA-2 data is interpolated along 4 dimensions: latitude (λ), longitude (ϕ), pressure level, and time. The latitude longitude grid of the MERRA-2 data is linearly interpolated to 1° by 1°, so it coincides with the satellite data.

As mentioned above, the equator crossing time of the satellites of the A-Train is around 13:30 local time. MERRA-2 is outputted 3-hourly starting at 0:00 UTC. To retrieve the meteorology at the time of satellite retrieval, a time interpolation is done. As we do not have orbit information, we apply a simple interpolation scheme and quantify the resulting error. To convert a local time at a longitude ϕ to UTC, we use the following equation:

$$t_{UTC}(\phi) = t_{local} - \frac{\phi}{15} = 13.5 - \frac{\phi}{15}\ , \tag{2.4}$$

where t_{UTC} and t_{local} are the time in UTC and the corresponding local time at longitude ϕ respectively, in hours.

Between -60° and 60° latitude on a Equirectangular map projection, the orbital track of the satellites is assumed to be linear. The slope of this line is computed by fitting a line to satellite retrievals along multiple flight tracks. We are interested in the equator crossing, because we know that the time at equator crossing is 13:30 local time. The coordinates of a grid box and the equator crossing along the flight path are linearly related:

$$\phi_0 = \phi_{retr} + \tan(\alpha)\lambda_{retr},$$

where (λ_{retr},ϕ_{retr}) are the coordinates of a grid box, ϕ_0 is the longitude of the equator crossing point along the flight track, and α is the fitting parameter.

As the satellites take approximately 34 minutes to travel from -60° and 60° latitude, we can calculate the retrieval time of every grid box by computing the intercept of the line with the equator, computing the time in UTC for the longitude at the equator crossing, and adding a latitude offset factor. Translating this to equation form gives:

$$t_{retrieval} = t_{UTC}(\phi_0) + \frac{\lambda_{retr}}{60}\frac{17}{60} = 13.5 - \frac{\phi_0}{15} + \frac{17\lambda_{retr}}{3600}, \tag{2.5}$$

where the factor 17/60 represents the offset due to the latitude of the retrieval (λ_{retr}). If a grid box has zero latitude (i.e. $\lambda_{retr} = 0°$), the retrieval time will be equal to the local time in UTC as the term with latitude vanishes: $t_{retrieval} = t_{UTC}(\phi_0) = t_{UTC}(\phi_{retr})$. In the case of $\lambda_{retr} = -60°$, the latitude factor becomes $\frac{\lambda_{retr}}{60}\frac{17}{60} = -\frac{17}{60}$ hr, i.e. a time offset of 17 minutes.

The time interpolation introduces an error, because points retrieved on the same orbital pass are given different retrieval times. The instrument makes measurements along a swath. Points that are located in the nadir direction relative to the satellite's position will have the correct local time, while estimated times for points on the edge of the swath will either be late or early. To quantify this error we calculate the maximum longitude difference this creates, and we compute the bootstrap error in the interpolated MERRA-2 parameters.

The MODIS swath width is $2.33 \cdot 10^3$ km. We round this to $2.5 \cdot 10^3$ km to compensate for the angle of the orbit relative to a meridian. We can express the swath width in degrees of longitude by using:

$$\phi = \left(\frac{2\pi R_E}{360}\cos(\lambda)\right)^{-1} \cdot d = (1.1132 \cdot 10^5 \cos(\lambda))^{-1} \cdot d \tag{2.6}$$

where d is the swath width in meters, λ is the latitude, and the prefactor $\frac{2\pi R_E}{360}$ is the circumference of the Earth divided by 360. This formula is approximate because it does not consider Earth's polar flattening, but between -60° and 60° latitude this effect will be small. From Equation 2.6 we can compute the swath width in degrees longitude. On the equator (i.e. $\lambda = 0$) we find: $\phi = 21°$. Half a swath width of offset implies $\frac{21/2}{15} = 0.7\ hr$ of maximum time error.

To estimate the error we compute random 'true' samples from the MERRA-2 parameters, linearly interpolating in time and picking points around 13:30 local time. We then create 'wrong' samples by randomly adding an offset between -0.7 to 0.7 hr. We then compute the normalised root-mean-squared error (NRMSE):

$$\text{NRMSRE} = \frac{\sqrt{\frac{1}{n}\sum_i (x_i - \hat{x}_i)^2}}{x_{max} - x_{min}} * 100\% \tag{2.7}$$

The NRMSE is low (<1%) for all parameters except Ω_{700} where the NRMSE is 1.5%. This can be explained by the variation of Ω being high between two time steps.

The MERRA-2 data is outputted on 42 different pressure levels. For the highest pressure levels, the step size is 25 hPa. So the highest two levels are 1000 hPa and 975 hPa (the highest value in pressure units is equivalent to the lowest value in height units as the two are inversely correlated). If the surface pressure is between these values, 983 hPa for example, the 1000 hPa level will have a no-data value. This is because this point is 'below the ground'. To compute the surface pressure and the surface wind-speed, the pressure levels need to be extrapolated to the surface pressure. This is done by linearly extrapolating the two pressure levels above the ground pressure.

To quantify the error of the pressure interpolation we randomly remove pressure points and extrapolate the two pressure points below to the removed pressure level and again compute the NRMSE. We do this for UV, T and EIS. We find low NRMSE for the parameters, with the highest NRMSE of 1.4% for T.

2.3. Climate model data

Climate model data is used to estimate the relative change in CCN concentrations from pre-industrial to present-day: $\Delta \ln N_{ccn}$ and to estimate the ratio between the radiative forcing above ocean and above land. For estimating $\Delta \ln N_{ccn}$, five global aerosol climate models (CAM5.3, CAM5.3-CLUBB, CAM5.3-CLUBB-MG2, HadGEM3 and SPRINTARS) are used. The PI and PD model outputs are computed from a pair of nudged simulations that are the same except for the aerosol emission. The PI reference year is 1850. Compared to the work of Hasekamp et al. (2019), we do not use ECHAM6-HAM because the CCN output of this model was not available, but we use HadGEM3 which was not used in the Hasekamp et al. (2019) study.

All these models are part of the AEROCOM intercomparison project, which aims to understand differences in aerosol climate models using multiple lines of evidence (Ghan et al. 2016). The column CCN at 0.3% supersaturation are used as an CCN estimate and the difference between PI and PD is computed according to Gryspeerdt et al. (2017) and Hasekamp et al. (2019). The grid size ($\lambda \times \phi$) of the models are either 1.125°×1.125°, 1.875°×2.5°, or 1.25°×1.875°. All models are linearly interpolated to the native 1° by 1° grid used in this study. The models have a temporal resolution of 3 hours. To compute one $\Delta \ln N_{ccn}$ estimate per grid box, we average every PD grid box over 2006 and every PI gridbox over 1850. The PI model runs are limited by our limited knowledge of the PI aerosol state of the atmosphere.

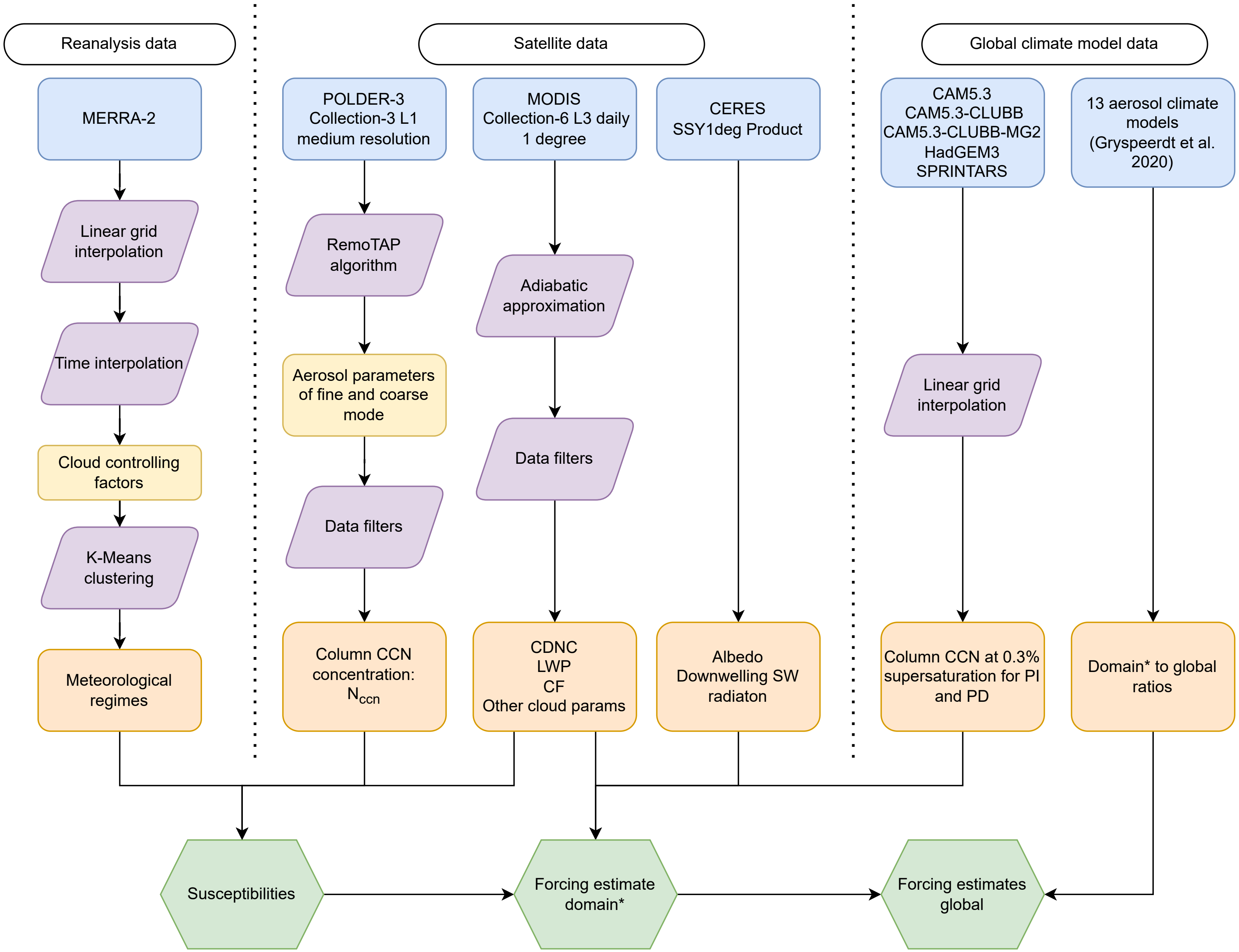


Figure 2.3: Flowchart showing all data products used, processing steps, and results. All data is for the year 2006, daily, on a 1 by 1 degree grid. *The study domain encompasses ocean grid boxes on a 1° by 1° grid, between -60 and 60 degrees latitude.

3

Clustering and cloud data

3.1. Clustering

Cluster analysis is the division of data into groups, also called *clusters*, such that elements of the same group are more alike than elements in another group. What is meant by 'more alike' depends on the criteria used to define the groups. Many quantitative similarity metrics have been defined to measure similarity (Saxena et al. 2017). The study of clustering within atmospheric science has recently attracted attention. In air quality studies, different clustering techniques have been used extensively (see Govender et al. 2020 for a review). Gryspeerdt et al. (2012) used k-means clustering on cloud properties, based on work by K. D. Williams et al. (2009), to analyse aerosol-cloud interactions. They found significant differences in ACI strength between regimes. Clustering on meteorological data (humidity, temperature, etc.) has not been done often. Evans et al. (2012) used clustering on meteorological data around Darwin, Australia to define weather states. More recently, Abraham et al. (2022) used k-means clustering on relative humidity (RH) profiles to study the relationships between RH, the sea-surface temperature and the outgoing LW radiation.

3.1.1. Clustering metrics and cluster occurrence patterns

Here, the MERRA-2 reanalysis data is clustered using the k-means clustering algorithm, to create global meteorological regimes. The only parameter that needs to be defined is the number of clusters k. The algorithm initialises k centroids and then optimises the inertia (see list below) until some threshold is reached to find the best clustering. To find the optimum number clusters the algorithm is run multiple times for many different values of k and different initialisations. The algorithm could converge to a local minimum, so doing multiple runs for different initialisations prevents this. For every run a host of metrics are calculated. As there is no ground truth, evaluation can only be done using the model itself. The metrics used are:

- **Inertia:** is the within-cluster sum-of-squares. In other words, it is the sum of the absolute value of the difference between the points and the centre of a cluster. In the limit of $k \rightarrow N$, where N is the amount of data points, the inertia will go to 0. In a plot of inertia and k, one looks for the 'elbow point' or point where the slope significantly tapers of to find the optimum value of k.
- **Calinski-Harabasz Index:** or the Variance Ratio Criterio, is the ratio of the sum of between-clusters dispersion and within-cluster dispersion for all clusters. Dispersion is the sum of distances squared. Higher values indicate better defined clusters, though the index prefers normally distributed clusters.
- **Silhouette score:** is calculated from the mean distance of a point from all other points in its cluster, and the mean distance of the point and all the points in the next closest cluster. Higher values indicate better separated clusters.
- **Davies-Bouldin score:** compares the within-cluster distance to the between-cluster distance. Values closer to zero indicate better separated clusters.
- **SSE Ratio:** is the ratio of the inertia for clustering k and k-1, so it measures how the decrease in inertia slows down.

These metrics are calculated for $k = [2, 3, \ldots, 14]$. Every run is done 10 000 times for different random initialisations and the metrics are averaged. Figure 3.1 shows the mean of metrics as a function of clusters. The standard deviation of the metrics are small, as the different clustering runs should converge to the optimal clustering. We see local maximums in the silhouette score at k=6 and k=12. There are local minima in the DB score at k=3, k=6, and k=12. The SSE ratio goes above 90% at k=5 and above 95% at k=8, and from the shape we can see that the decrease in inertia slows down after k=8. The CH index has a maximum at k=3 and no significant other features. The choice of k is somewhat arbitrary, but we will continue with k=6, and compare every computation to different values of k to get a measure of how the analysis is affected by the choice of k.

Figure 3.2 shows the relative frequency of occurrence (RFO) for the clustering with k=6. The RFO is computed by summing the occurrences per cluster per grid box, and dividing by the total number per grid box (which is 365 as we have daily data for 1 year). There are white pixels over the ocean caused by islands (e.g. Hawaii or the Azores) in that specific grid box. The clusters show substantial spatial variation. C0 and C1 are near the equator, while C2 is primarily on the equator. C3 and C4 are mid-latitude clusters, while C5 is spread out over the globe, but the cluster is sparse.

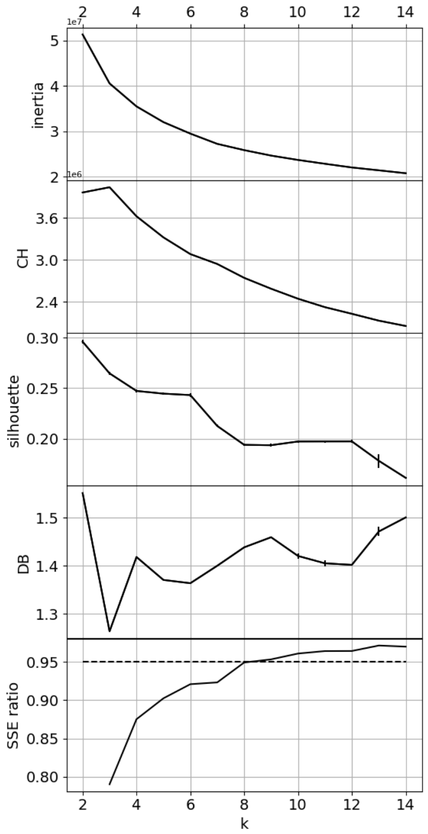


Figure 3.1: Cluster metrics for different values of k. The metrics are computed for 10 000 runs and then averaged. The standard deviations of the metrics are small.

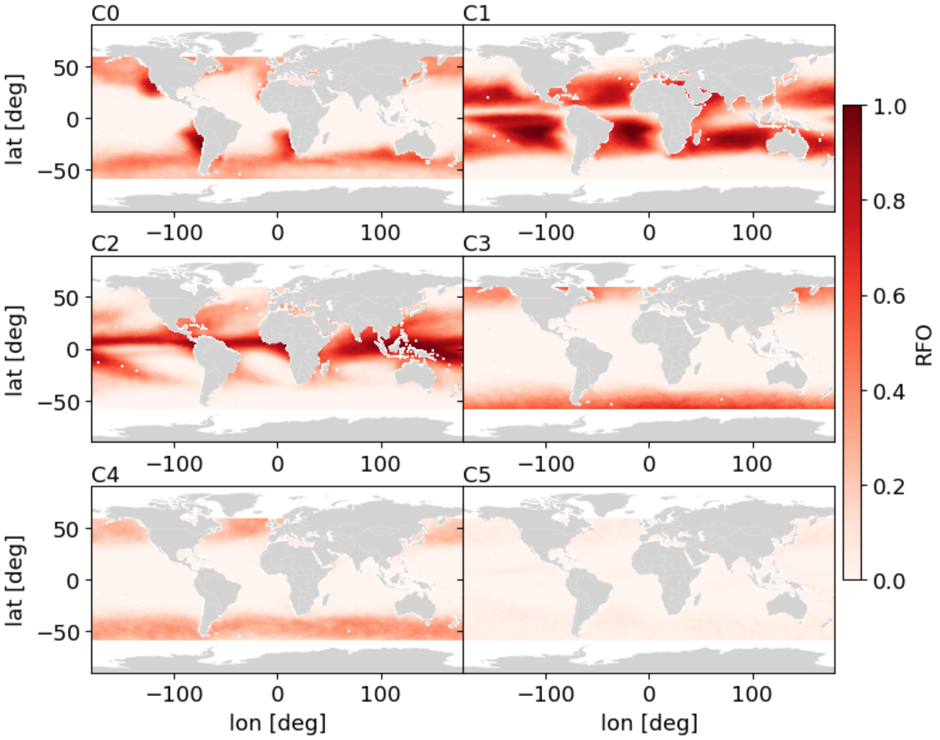


Figure 3.2: Relative frequency of occurrence of the clusters for k=6.

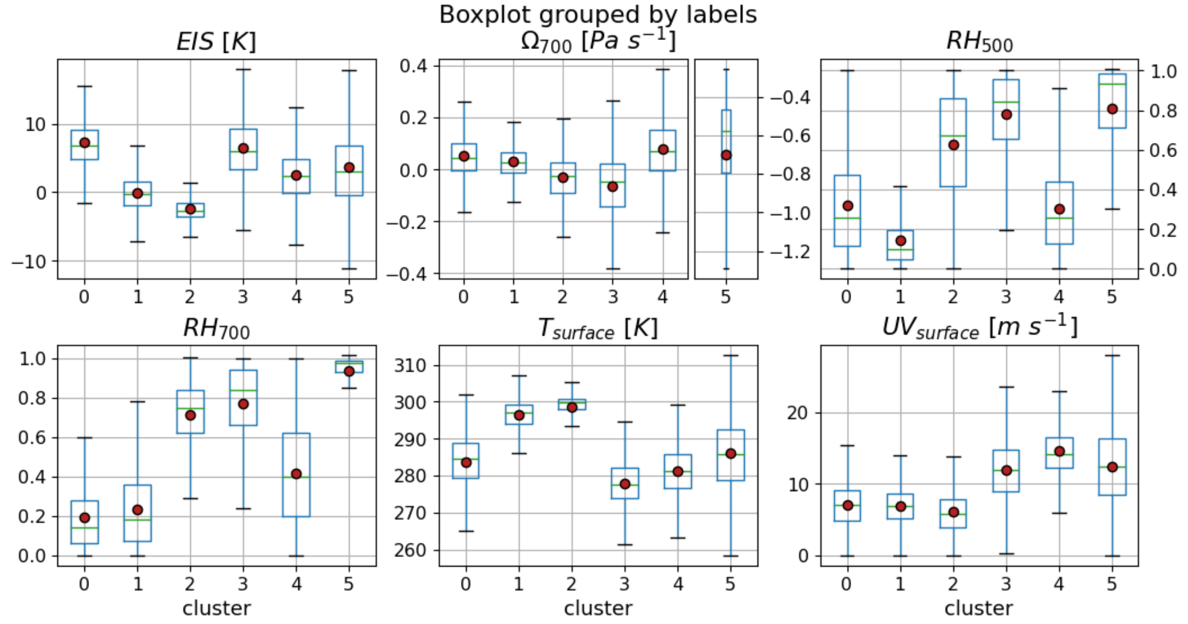


Figure 3.3: Here the meteorological parameters are shown per cluster. The boxes represent the 25% (Q1) and 75% (Q3) quantiles. The green represents the 50% quantile and the red dot the mean. The whiskers show the range of the data, excluding points that lie more than $1.5 \cdot (Q3 - Q1)$ from the edges of the box.

The regions where C0 has high RFO coincide with regions where stratocumulus clouds are prevalent driven by SST increase (Wood 2012). These stratocumulus clouds move over the open ocean following the trade winds and form trade wind cumulus clouds, which could be partly captured by C1. C2 has high concentration near the equator, specifically around the Indonesian islands, where cirro- and alto-class clouds are prevalent. In the mid-latitude clusters (C3 and C4) we can expect high wind-speeds and both stratiform and alto-class cloud formation (Kuma et al. 2023).

In Figure 3.3, a boxplot is shown of the cluster parameters. Note that for Ω_{700} , cluster 5 is plotted on a separate y-axis to make it more readable as this cluster has very negative values of Ω_{700} . The y-axis of RH_{500} is plotted on the righ-hand side. Cluster 0 has high EIS with a mean of around 8 Kelvin. Ω_{700} is mostly positive with Q2 being close to zero, indicating subsidence, but the magnitude is small. RH is low for the 700 hPa level but somewhat higher for the 500 hPa level. This could indicate moist free tropospheric conditions which could dampen entrainment drying. The surface temperature is centered around 285 Kelvin but the spread is large as this cluster covers both the near-equatorial region and the mid-latitudes. The surface wind speed is low.

Cluster 1 has lower EIS compared to cluster 0, indicating lower stability and therefore lower cloud cover, which also indicates a stratocumulus-to-cumulus transition between C0 and C1. Ω_{700} is small, but most of the data has positive values. RH at both levels is low, but at 500 hPa the RH is lower, indicating a dry free troposphere. The temperature is high as the cluster is centered around the equator. The wind speed is low.

Cluster 2 has the lowest EIS values. Ω_{700} is on average negative, but the mean is close to 0. RH is high at both pressure levels, but the mean of RH_{700} lies higher compared to the 500 hPa level. The surface temperature is high, as this cluster is primarily in the tropics. The wind speed is also low.

Cluster 3 has high values of EIS, with a mean comparable to C0, but the spread in EIS is larger in both directions. Ω_{700} is negative, with the 75% quantile lying almost on 0. RH is high at both pressure levels. Surface temperature is low as this is a mid-latitude cluster. the wind speed is high with a mean of around 12 m/s.

Cluster 4 is almost an opposite to C3. The EIS is lower, while still close to the mean of the total dataset. Ω_{700} is positive, with the 25% quantile lying on the 0 line. RH is low at both pressure levels. T and UV are similar, but UV is slightly larger compared to C3, and there is less spread. The wind speed is high. Cluster 5 is a sparse cluster, primarily characterised by negative Ω_{700} values and very high RH (especially at 700 hPa). EIS, T and UV cover the entire range of possible values.

There is difference between the clusters in the cluster parameters, but there is also significant overlap. As the meteorological parameters are spread out continuously, overlap between the clusters is large. The spatial distribution of the clusters (Figure 3.2) shows patterns which coincide with regimes found in literature (Abraham et al. 2022; Tselioudis et al. 2013). Clustering cloud-controlling factors thus seems feasible. When we choose different values for k, the clustering changes, but general patterns remain. In Appendix B, the RFO maps and boxplot for k=3 and k=12 are shown.

3.2. Cloud data

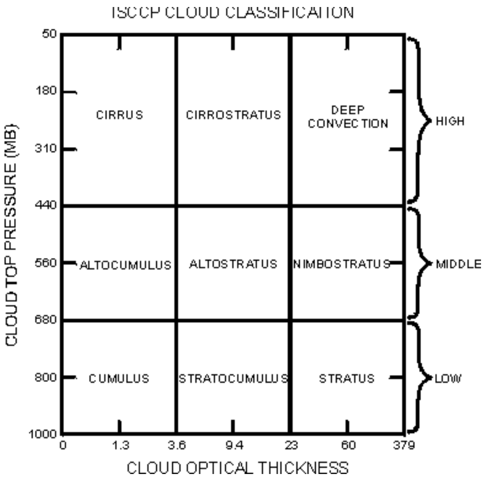


Figure 3.4: Cloud classification using a joint histogram of cloud optical thickness and cloud top pressure. (Figure adopted from Rossow et al. 1991)

To see what clouds are present in the clusters, we analyse MODIS L3 cloud data. Cloud optical thickness (COT, τ_c), cloud top pressure (CTP), liquid water path for both the liquid and ice phase, and cloud retrieval fraction for both the liquid and ice phase are retrieved, as well as joint histograms of cloud phase versus temperature, and CTP versus COT.

Note that for analysing aerosol-cloud interactions, we work with N_{ccn} and N_d retrievals that can only be done for liquid clouds. We will still consider ice clouds in this section, but later we limit the data to grid boxes without ice clouds.

The MODIS data has 2 cloud fraction products: the cloud mask CF, and the CF retrieved from optical properties. The latter has the advantage of providing a CF for the liquid, ice, and undetermined cloud phase. This is done by an advanced phase retrieval algorithm based on the shortwave radiation measured by MODIS (Marchant et al. 2016). This allows one to have a CF for both liquid and ice in one grid box.

CTP and COT can be used to classify cloud types. CTP tells you about the height (low, alto, or cirrus), while COT gives a sense of vertical development (stratus or cumulus). This dates back to work done in 1991 by the International Satellite Cloud Climatology Project (ISCCP) (Rossow et al. 1991). Figure 3.4 shows the joint histogram of CTP and COT, together with the limits for the different cloud classifications. COT gives a sense of vertical development, as it depends on cloud depth, but it also depends on LWP and CDNC. Two clouds with the same cloud depth can have different values of COT. For this analysis, we are only interested in the global classification introduced by the ISCCP, so we will not consider this further.

The clouds optical phase versus temperature joint histogram utilises the cloud phase label to show the cloud top temperature for each phase bin. Figure 3.5 shows the joint histogram per cluster. Each joint histogram is normalised so all the bins sum to one. This makes it difficult to compare absolute values between clusters, but we are more interested in the absolute distribution of cloud types per cluster. The 'mixed+und.' bin are clouds classified as clouds containing both liquid and ice or clouds that could not be classified.

C0 and C1 both primarily have liquid clouds. C0 has colder liquid clouds which can be explained by the cluster covering both the low and mid-latitudes, while C1 lies only in the low-latitudes. C0 has more mixed/undetermined clouds, which could be because the mid-latitude clouds will have more ice content. This goes in line with the high concentration of mixed/undetermined clouds in the mid-latitude clusters (C3 and C4). C2 has a combination of low and high ice clouds, but also warm liquid clouds. C3 has a mix of liquid, ice, and mixed clouds. C4 has mostly liquid clouds, while C5 has mostly ice clouds. From this we can predict that C3 and C5 will have few N_d retrievals as these can only be done for warm liquid clouds. In both C1 and C2 there are some very cold (200 Kelvin) liquid clouds which is either a retrieval error or an issue caused by averaging retrievals over the $1° \times 1°$ degree grid boxes, as liquid water droplets can not exist at these temperatures.

In Figure 3.6 the CTP versus COT joint histogram per cluster is shown. All the histograms are normalised so the bins sum to 1. Stratocumulus and cirrostratus are prevalent in all clusters. C1 has more cumulus and cirrus clouds compared to C0, which is dominated by stratocumulus and cirrostratus. C2 mainly has cirrus and cirrostratus, with some deep convective clouds. C3 and C4 both have some nimbostratus and altostratus clouds (with $\tau_c > 60$). C3 primarily has cirrostratus, while C4 has more

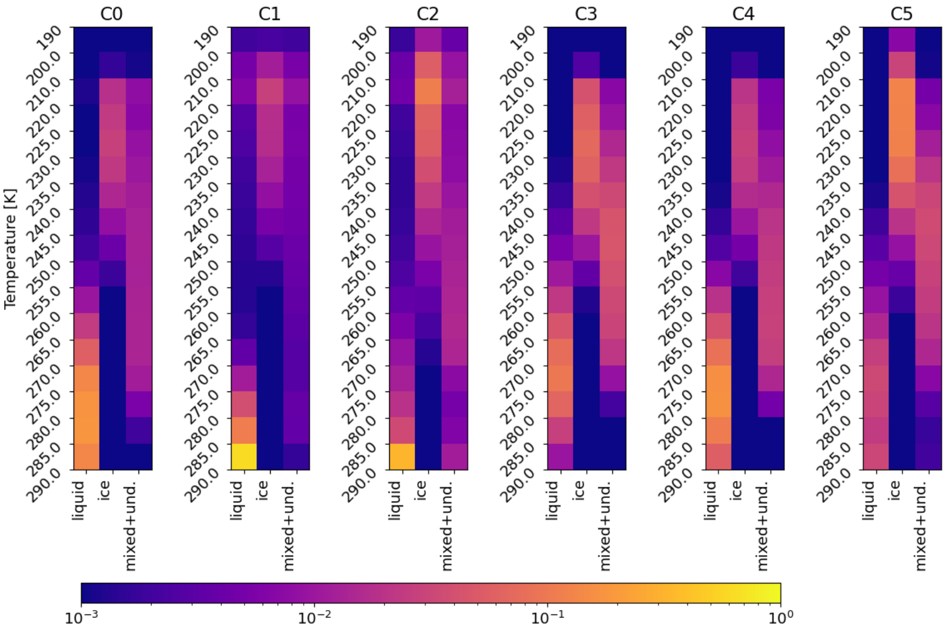


Figure 3.5: Joint histograms of the cloud phase and cloud top temperature for each cluster. The cloud phase is retrieved using optical properties. The cloud top temperature is also a measure of cloud top height as temperature decreases with height in the atmosphere.

stratocumulus. C5 is characterised by cirrostratus and deep convective clouds. It is interesting to note that deep convective clouds are only visible in C2, C3, and C5 (with only a very small signal in C4). These are also the clusters that have primarily negative Ω_{700} , indicating updraft conditions.

Figure 3.7 shows kernel density estimates of the cloud top temperature (CTT), cloud fraction of both liquid and ice clouds (CRF_{liq} and CRF_{ice}), and liquid water path of liquid. The shows kernel density estimates are made by smoothing the histogram and normalising the area under the curve to unity. We see that there are two clusters (C1 and C2) that show an almost identical pattern with primarily high cloud top temperatures but they also have peak at very low CTT values. C0 and C4 also show an almost identical pattern with primarily CTT values around 270 Kelvin. C3 shows a similar pattern to C0/C4, but has higher density at colder values of CTT. This is also visible in the joint histograms: C3 has more ice clouds than C0/C4. C5 is spread out over most values with both warm and cold cloud tops. The values of CTT are influenced by the temperature variations across the globe.

The cloud fraction of liquid clouds shows peaks at both 0 and 1. The separation we saw for CTT is similar here. C1 and C2 have the lowest liquid cloud fraction. C3 and C4 have the most, and these clusters are also the mid-latitude clusters where there are clouds often. Note that these plots do not take the sizes of the clusters in account. C1 could still have more grid boxes with liquid clouds compared to C3 because the cluster simply has more points. The ice cloud fraction shows the highest values for C1 and C2, but the differences are small. The liquid water path shows a peak for all clusters which drops off. C1 and C2 drop off the fastest while the other clusters, especially C0 and C3, reach higher values of LWP, indicating more liquid water content for the clouds in those clusters.

From this we can conclude that clustering cloud-controlling factors translates to a reasonable separation of MODIS cloud parameters. Stratocumulus and cirrostratus are captured in all clusters, but each cluster has specific characteristics not captured by other clsuters. From the meteorological and cloud features of the clusters, we define novel names for the clusters to refer to them. Up until this point we already referred to low-latitude (C0, C1, and C2) and mid-latitude clusters (C3 and C4), but the names will allow us to more easily recognise what the characteristics of the clusters are.

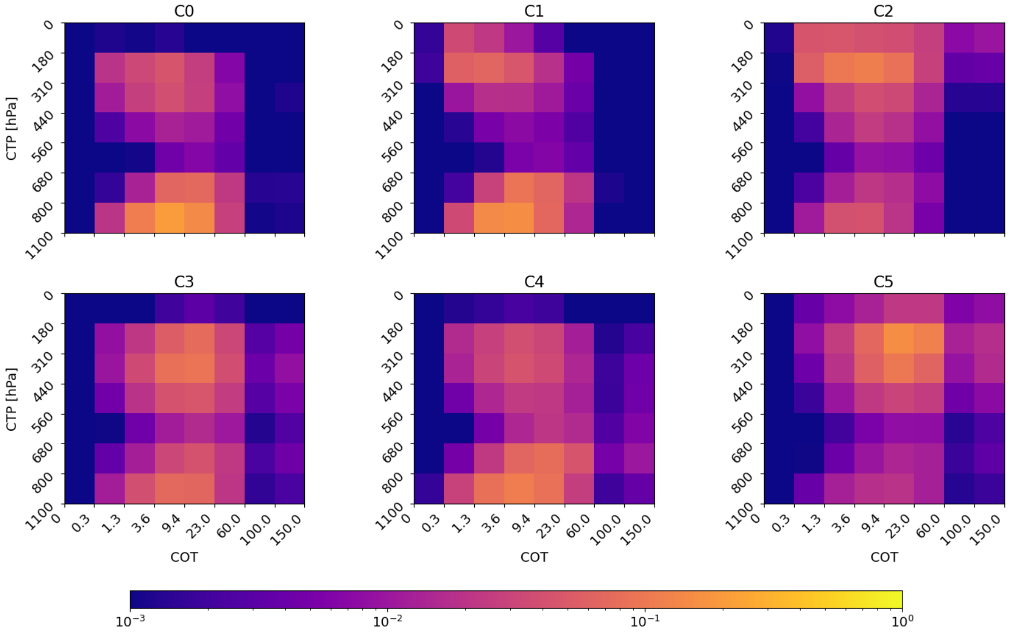


Figure 3.6: Joint histograms of the cloud optical thickness and cloud top pressure for each cluster.

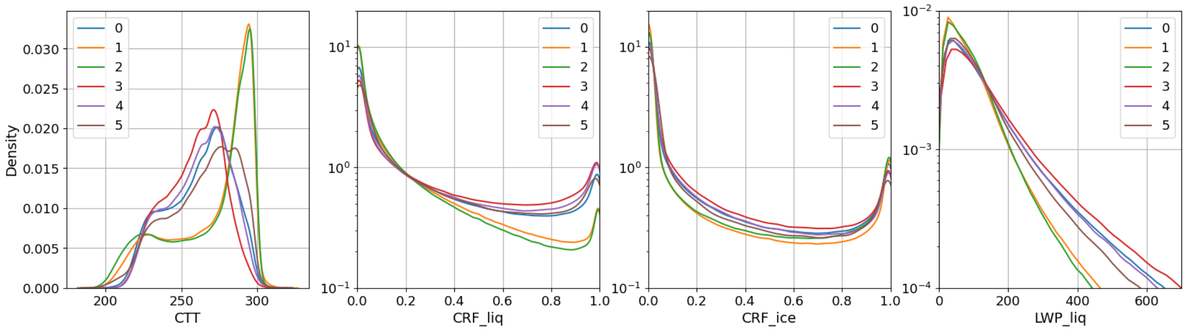


Figure 3.7: Kernel density estimates of MODIS cloud parameters per cluster.

Cluster	Name	Characteristics
0	Dry and stable	High EIS, low RH, positive Ω_{700}
1	Dry and low stability	Low RH, low EIS, positive Ω_{700}
2	Equatorial low stability	High RH, negative Ω_{700} , low EIS
3	Moist mid-latitude	High RH, negative Ω_{700} , high wind-speed
4	Dry mid-latitude	Low RH, positive Ω_{700} , high wind-speed
5	Omega outliers	Very negative Ω_{700} , very high RH

Table 3.1: Names and characteristics per cluster

4

Radiative Forcing of the cloud albedo effect

4.1. Susceptibility of CDNC to CCN perturbation

The susceptibility of the cloud droplet number concentration to CCN concentration is the linear relation between CDNC and N_{ccn} in log-log space. It is a measure of how strong N_d is influenced by a change in N_{ccn} . It is defined as:

$$\beta_{\ln N_d - \ln N_{ccn}} = \frac{\partial \ln N_d}{\partial \ln N_{ccn}}. \tag{4.1}$$

Other aerosol measures than N_{ccn} can be used (AOD or AI as mentioned in Chapter 1), but N_{ccn} will most accurately capture the actual variation in CCN concentrations. The clusters defined in Chapter 3 are matched to the satellite retrievals from the dataset by Hasekamp et al. (2019). Only grid boxes where both a CDNC and N_{ccn} estimate are retrieved are used. The data is split into equal sized N_{ccn} bins and the susceptibility is computed using a linear least squares regression on the median values of the bins. This is done for points with $N_{ccn} > 10^7$ cm^{-2}, because of absolute retrieval errors dominating at lower concentrations.

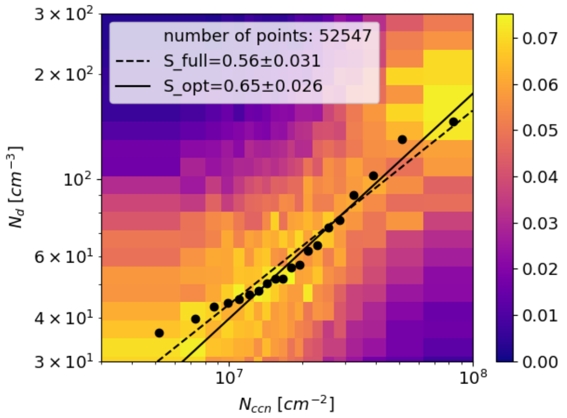


Figure 4.1: Joint histogram of CDNC and N_{ccn} . The black dots represent the median value per CCN bin. The linear fit through the points represents $\beta_{\ln N_d - \ln N_{ccn}}$. The solid line is the linear fit only considering points with $N_{ccn} > 10^7$.

Figure 4.1 shows the susceptibility $\beta_{\ln N_d - \ln N_{ccn}}$ for the entire dataset. The colours show the normalised histogram of N_d for each N_{ccn} bin. The bins have different widths because every bin contains the same amount of data points. The solid line shows the regression through the median points for points with $N_{ccn} > 10^7$ cm^{-2} (S_{opt}). To compute this fit, the binning is done again with the subset of the data. The dashed line shows the regression through the median values of the bins with all data (S_{full}). S_{full} is lower compared to S_{full}, due to the taper effect at low concentrations.

It is thought that the taper effect at low N_{ccn} values, which is partly visible in Figure 4.1 is caused by retrieval errors. P. L. Ma et al. (2018) showed that areas with low aerosol loading are not well characterised by satellites, while this regime is important for aerosol-cloud interactions in cloud models. This is caused by the absolute error of the retrieval, as the absolute error most significantly impacts low values of N_{ccn} .

Hasekamp et al. (2019) computed the susceptibility for different geographical ocean regions defined by Quaas et al. (2008). This is done to create regimes where there is similar meteorology. We reproduce the susceptibility and implied radiative forcing using these geographical regions, and compare them to

Acronym	Full Name
NPO	North Pacific Ocean
NAO	North Atlantic Ocean
TPO	Tropical Pacific Ocean
TAO	Tropical Atlantic Ocean
TIO	Tropical Indian Ocean
SPO	South Pacific Ocean
SAO	South Atlantic Ocean
SIO	South Indian Ocean

Table 4.1: Ocean regimes from Quaas et al. (2009)

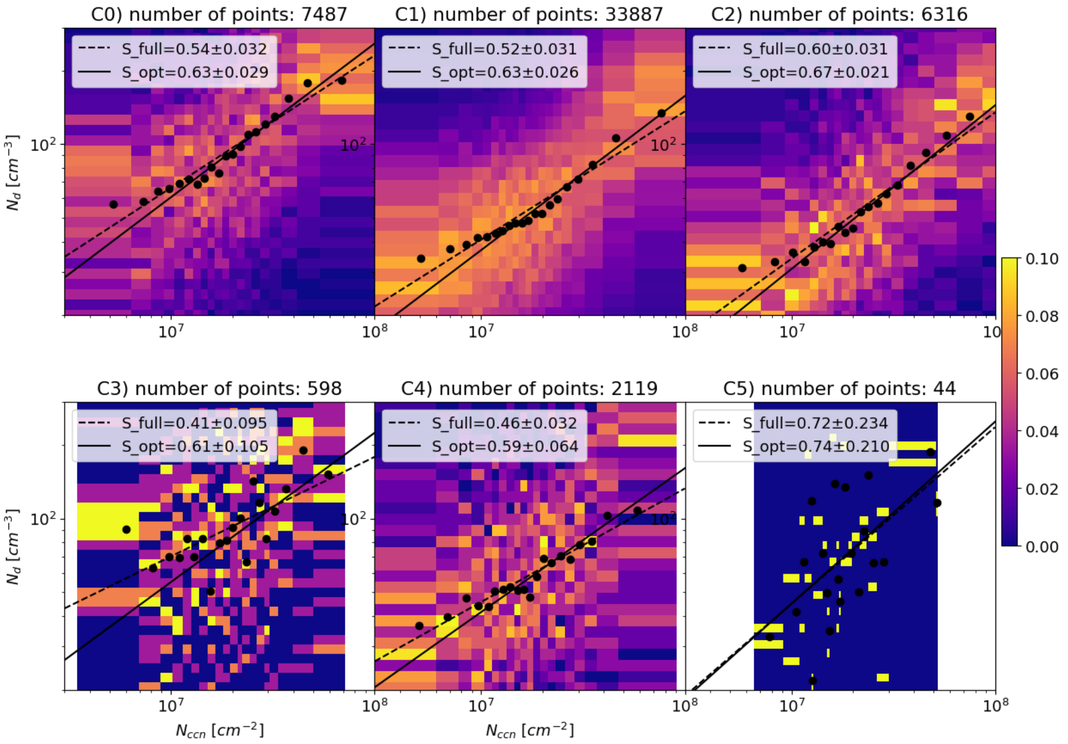


Figure 4.2: The susceptibilities of CDNC to N_{ccn} for the different clusters. The solid line shows the linear least squares fit to the bin median values for $N_{ccn} > 10^7$, as values below this threshold cause an underestimation of the slope.

the susceptibilities and implied forcing from the meteorological regimes. Table 4.1 show the acronyms for these geographical regions.

Figure 4.2 shows the susceptibility per cluster. The title shows the number of grid boxes where there are both a N_{ccn} and N_d retrieval. The computed susceptibilities differ slightly. C0 and C1 have the same susceptibility, which is slightly lower than the global value (0.65). C2 has a higher susceptibility, but the errors of the susceptibilities overlap. The biggest meteorological difference between C2 and C0/C1, is higher RH at both levels, and more negative Ω_{700} . The higher RH could help prevent evaporation at cloud top, as the air above the cloud is moist. The more negative Ω_{700} implies more grid boxes with updrafts, which enables more CCN to be activated, as higher supersaturations can be reached.

The mid-latitude clusters (C3 and C4) have lower values for the susceptibility, but the error is large because of the low number of retrievals. Computing a susceptibility for C5 is not possible due to the low number of data points.

4.1.1. Hemispheric difference

Figure 4.3 shows the susceptibility $\beta_{\ln N_d - \ln N_{ccn}}$ for both the meteorological clusters and the geographical regimes for different values of the lower CCN limit used to compute the susceptibility ($N_{ccn,lim}$). The blue bars correspond to the values computed using $N_{ccn,lim}$ defined by Hasekamp et al. (2019). The

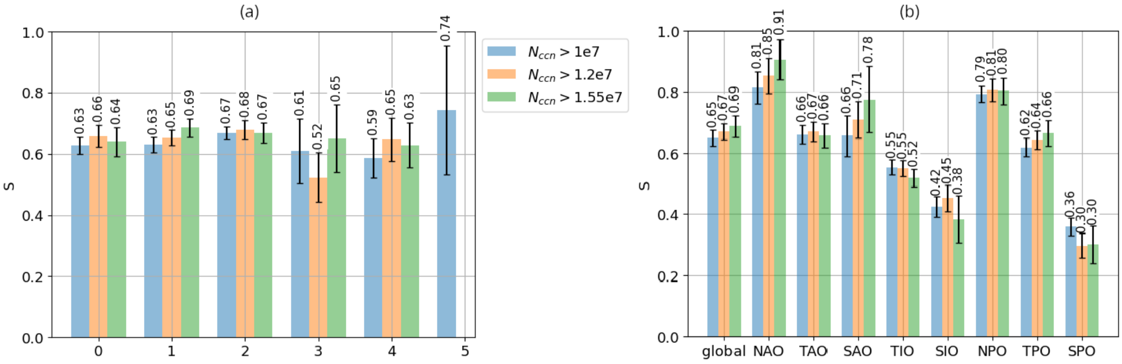


Figure 4.3: The susceptibilities for the direct aerosol effect are shown for a) the meteorological clusters and b) the geographical regimes defined in Table 4.1. The numbers above show the value of the susceptibility and the errorbars indicate the least-squares fit error.

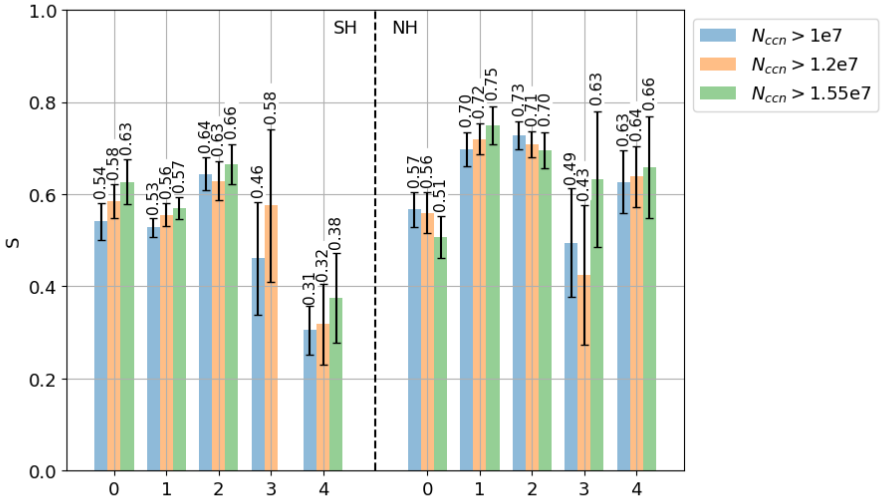


Figure 4.4: The susceptibilities for the Northern (positive) and Southern (negative) hemispheric components of the meteorological clusters are shown. The susceptibilities for the northern components of each cluster are more positive compared to the southern components.

values that we compute for the geographical regimes (blue bars in panel (b) of Figure 4.3) are identical to the values reported by Hasekamp et al. (2019), except that the value for the SAO regime is higher in our findings. This is not expected to influence the radiative forcing estimate.

It is interesting to note that in Figure 4.3 the meteorological regimes show less variation compared to the geographical regimes. In the geographical regimes we see larger variation, especially between the southern regimes (SIO, SPO) and the northern regimes (NAO, NPO). This is unexpected from a process based view. The Northern Hemisphere (NH) has much higher aerosol loading compared to the Southern Hemisphere (SH), as there is much more anthropogenic activity in the NH. This means that clouds in the NH are on average more likely to be aerosol saturated: adding more aerosols will not affect N_d. This means that an aerosol perturbation has more effect on clouds in the SH, as these clouds are less likely to be aerosol saturated (Carslaw et al. 2013).

To assess the hemispheric difference, we split our clusters, presented in Chapter 3, into the NH component and the SH component and compute the susceptibility. This is shown in Figure 4.4 for different values of $N_{ccn,lim}$. The bars on the left represent the SH clusters, and the bars on the right represent the NH clusters. When looking at the blue bars, we see that the NH clusters have higher susceptibilities compared to the southern clusters. When taking the error bars into account, the northern component are higher for some of the clusters (C1, C2, and C4), and lower for C0. This goes against our expectation, but it is in line with the geographical regimes in Figure 4.3.

Looking back at the computation of the susceptibilities, a choice was made by Hasekamp et al. (2019)

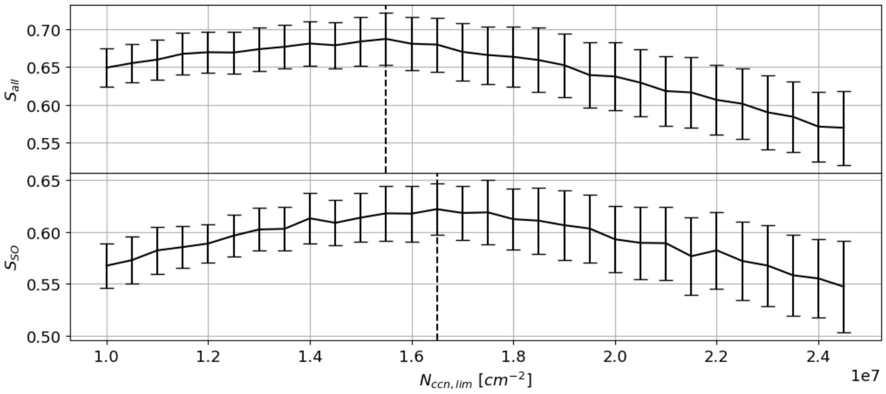


Figure 4.5: This figure shows A) the global and B) Southern Hemispheric susceptibility as a function of the lower data limit for CCN: $N_{ccn,lim}$. The error bars indicate the least-squares fit error. The dotted line indicates where the suceptibility is highest.

to compute S using only data with $N_{ccn} > 10^7$ cm^{-2}. This is done because absolute errors dominate at low CCN values. It is possible that for the southern retrievals this limit is too low. Because there are less high values of CCN, the taper effect at low CCN dominates and the susceptibility is underestimated. To assess this, we compute the susceptibility for a range of limits for both the global dataset and only SH data. This is shown in Figure 4.5. The x-axis shows the limit on N_{ccn} , and the y-axis shows the computed value of S with the linear least-squares fit error. The dashed line indicates the point where S is highest. The plot above shows the result for the entire dataset (S_{all}) and the plot above for only the SH points (S_{SO}). We see that S increases as we increase the limit, and at some point starts to decrease. The differences are small compared to the error bars. It is therefore difficult to make any definite statements. It does show that the taper effect at low N_{ccn} value due to retrieval error causes an underestimation of the susceptibility. This effect could affect SH points more. The magnitude of the error bars increase as $N_{ccn,lim}$ increases, because more data gets discarded.

The orange bars in Figure 4.3 shows the susceptibilities seen before if we recompute them with $N_{ccn} > 1.2 \cdot 10^7$ cm^{-2} and the green bars for $N_{ccn} > 1.55 \cdot 10^7$. For cluster 5 there are no other estimates as this cluster has almost no data points to compute a fit with. We see regimes where S increases, and regimes where S decreases. There are some where (e.g. cluster 3) where there is first a decrease and then an increase. The increase is strongest in NAO, as the susceptibility for the NAO goes as high as 0.91. Interestingly, 2 of the southern regimes (SIO and SPO), have lower values for S if $N_{ccn} > 1.55 \cdot 10^7$ cm^{-2}. The SAO does increase with the new limits. $1.2 \cdot 10^7$ cm^{-2} was the only limit where S for SIO increases. This increase in the meteorological clusters is smaller, because the clusters have a combination of SH and NH points. The slight increase does indicate that some of the noise due to retrieval error can be lost if we choose a higher limit on N_{ccn} . Choosing a higher limit does reduce the amount of data to compute a fit with. The error bars are large, so to test if the differences are significant, we apply a two-sample t-test for equal means (Snedecor et al. 1989). This tests the null hypothesis that two sample means are the same. Assuming equal variances, the test statistics is given by:

$$t = \frac{\mu_1 - \mu_2}{s_p\sqrt{1/n_1 + 1/n_2}}, \tag{4.2}$$

where the pooled variance is defined as:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \tag{4.3}$$

The parameter n is the sample size, μ is the sample mean, and s^2 is the sample variance (or standard deviation squared). The full definition of the two-sample t-test for equal means for both equal and unequal variances can be found in Appendix A. The critical that we test against is 1.686 following from a t-distribution with 38 degrees of freedom at a critical level of 0.05. The sample size is 20, as a fit is made of 20 data points computed from 20 bins with equal amount of data per bin. If the sample size is smaller, the uncertainty in the bins will be higher. It is not clear how this should translate to the two-sided t-test. The uncertainty will affect the variances, so lower sample size will result in lower test statistics.

We apply the test to the bars in Figure 4.3a, comparing the values of the different $N_{ccn,lim}$ slopes (test orange against blue, green against orange, and green against blue). Cluster 1 is the only cluster for which all three tests show that the slopes are different, with p-values of 2.99, 3.86, and 6.55. For cluster 3 and cluster 4, we find a that the estimate for $N_{ccn,lim} = 1.2e7\ cm^{-2}$ is significantly different from the other 2.

For the hemispheric difference (Figure 4.4), we find that for the SH component of cluster 0 and 1, and the NH component of cluster 1, all show a significant increase. For the NH component of cluster 0, the green bar is significantly lower compared to the other two. For the other clusters, the susceptibilities are not statistically different. For Figure 4.5, we find that the slope at maximum mean is different compared to the value at $N_{ccn,lim} = 1e7$ cm^{-2}. The results did not change when we assume unequal variances.

This shows that, globally, the susceptibilities are underestimated due to retrieval errors, but per cluster, only a few clusters show this effect. The cluster that does show an increase (C1) is also the largest cluster, meaning that the fit is best resolved. Hasekamp et al. (2019) also found this when modelling the effect of retrieval error. This effect is taken into account when computing the forcing estimate. The fact that satellite retrievals can not accurately capture the effects at low N_{ccn} values is problematic as this is where the cloud albedo effect has the biggest impact (Carslaw et al. 2013). Reducing retrieval error with better algorithms and measuring devices could help alleviate this.

4.2. Estimating the radiative forcing

To compute the radiative forcing of the cloud albedo effect (RF_{aci}), we use the same methodology as Hasekamp et al. (2019), based on work by Gryspeerdt et al. (2017). As mentioned above we use five different aerosol global climate models (GCM) to estimate $\Delta \ln N_{ccn} = (\ln N_{ccn}^{PD} - \ln N_{ccn}^{PI})$: the log difference between present-day (PD) and pre-industrial (PI) CCN concentrations. The different models give us a range of possible forcings. The radiative forcing can be calculated using the following equation:

$$IRF_{aci} = -\frac{1}{3}F^{\downarrow} f_{liq}\alpha_{cld}(1-\alpha_{cld})\frac{\Delta N_d}{N_d}, \tag{4.4}$$

following Gryspeerdt et al. (2017) and Hasekamp et al. (2019). We adopt the terminology from the IPCC where IRF_{aci} stands for the instantaneous radiative forcing of aerosol-cloud interaction, i.e. the forcing of the cloud albedo effect. Here $F^{\downarrow}$ is the daily mean incoming solar radiation flux taken from CERES, f_{liq} is the yearly average liquid cloud cover per grid box taken from MODIS, α_{cld} is the cloud albedo taken from CERES, and $\frac{\Delta N_d}{N_d}$ is the change in cloud droplet concentrations. We relate $\Delta N_d / N_d$ to $\Delta \ln N_{ccn}$ using our susceptibilities. For this we use the following two relations:

$$\Delta \ln x \approx \frac{d \ln x}{dx}\Delta x = \frac{1}{x}\Delta x, \tag{4.5}$$

$$\Delta \ln N_d = \beta_{\ln N_d - \ln N_{ccn}} \Delta \ln N_{ccn}. \tag{4.6}$$

This approximation holds for small values of Δx. Inserting into Equation 4.4 gives us the following equation:

$$IRF_{aci} = -\frac{1}{3}F^{\downarrow} f_{liq}\alpha_{cld}(1-\alpha_{cld})\beta_{\ln N_d - \ln N_{ccn}} \Delta \ln N_{ccn}. \tag{4.7}$$

IRF_{aci} is computed daily for every grid box, using the specific value of $\beta_{\ln N_d - \ln N_{ccn}}$ for that day at that grid box. To create a map of the radiative forcing, the mean of every grid box is calculated. To compute the global average, every grid box needs to be weighted by its surface area, as grid boxes near the equator are significantly larger than grid boxes near the poles. As our domain is between -60° and 60° latitude over ocean, we can only compute a domain average forcing based on the satellite data.

$$IRF_{aci}^{domain} = \sum_i \frac{IRF_{aci}^{i} \cdot A(\phi_i)}{A_{total}}. \tag{4.8}$$

IRF_{aci}^{i} is the mean radiative forcing per gridbox i, $A(\phi_i)$ is the area for gridbox i in m^2, and A_{total} is the total area of the gridboxes in m^2. The area per gridbox only depends on latitude, as gridboxes with the same longitude have the same area.

Figure 4.6 shows maps of the radiative forcing estimate for A) the geographical regimes and B) the meteorological regimes. C) shows the difference between the two. We see a strong signal in the northern

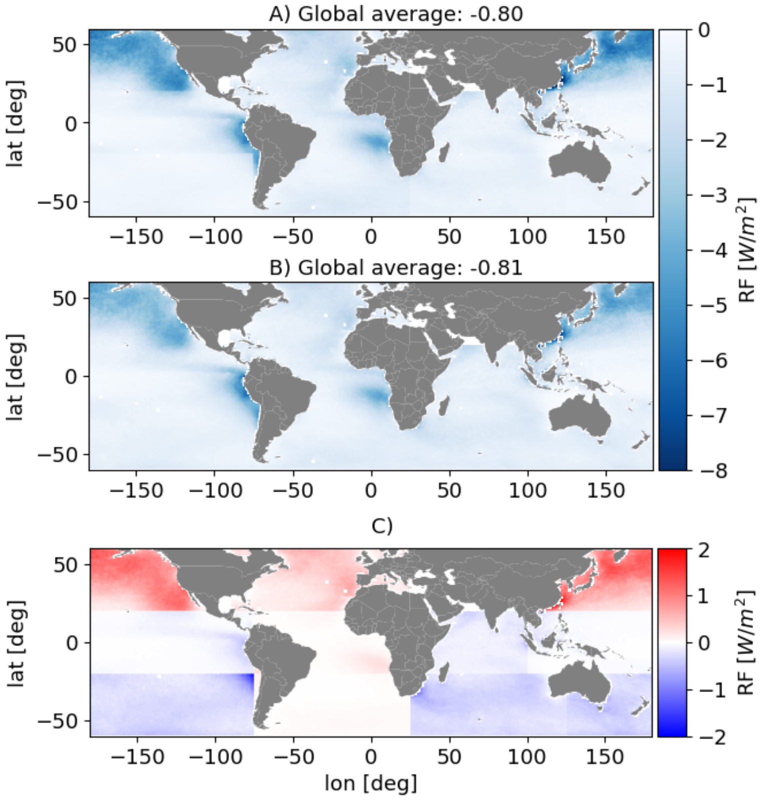


Figure 4.6: Maps showing the radiative forcing of the cloud albedo effect for A) the geographical regimes and B) the meteorological regimes. C) shows the difference between the two maps.

pacific due to the high concentration of anthropogenic emissions. The difference map shows a box like pattern because these correspond to the geographical regions where the scusceptibility is computed. A specific geographical regions has 1 value for $\beta_{\ln N_d - \ln N_{ccn}} \Delta \ln N_{ccn}$, while that value can change for the meteorological clusters. This creates a more smooth forcing estimate for the meteorological regimes because when we average over the year, $\beta_{\ln N_d - \ln N_{ccn}} \Delta \ln N_{ccn}$ varies between different values. The global average forcing is almost identical between the two methodologies: -0.80W m^{-2}for the geographical regimes and -0.81W m^{-2}for the meteorological regimes.

The difference map shows that using the geographical regimes leads to higher forcings in the northern hemisphere, and a lower forcing in the southern hemisphere. This is expected from Figure 4.3, as the susceptibilities in the northern and southern hemisphere are higher and lower respectively compared to the meteorological regimes.

To translate the forcing over our domain to a global estimate, scaling factors from 13 different climate models are used. These give a range of values between 1.12 and 2.24 with a mean of 1.5. The uncertainty in the N_{ccn} retrievals leads to an underestimation of $\beta_{\ln N_d - \ln N_{ccn}}$. This follows from the simulator results of Hasekamp et al. (2019) and from our analysis above. This underestimation leads to an uncertainty in the RF, which is modelled as a normal distribution with mean 1.05 and standard deviation 0.05. Combining the different model output, the scaling factors, and the susceptibility error, a range of possible forcings are retrieved. The histograms are shown in Figure 4.7. The left figure shows the forcing when using meteorological clusters and the right with the geographical regimes. We take the median value as the best estimate of IRF_{aci} and the 5-95% quantiles give a range of uncertainty : -1.21 W m^{-2}between -0.87 and -1.84W m^{-2}for the meteorological clusters. The estimates of the clusters and the geographical regimes are near identical. This is caused by the large uncertainty in the domain scaling factor. To alleviate this we also estimate the forcing over the domain.

Figure 4.8 shows the forcing estimates without the scaling factor. The range of values are thus made up from the uncertainty in ΔN_{ccn} and the uncertainty in the susceptibility due to retrieval error. The higher susceptibilities of the geographical regimes mean that the forcing is higher with a median of

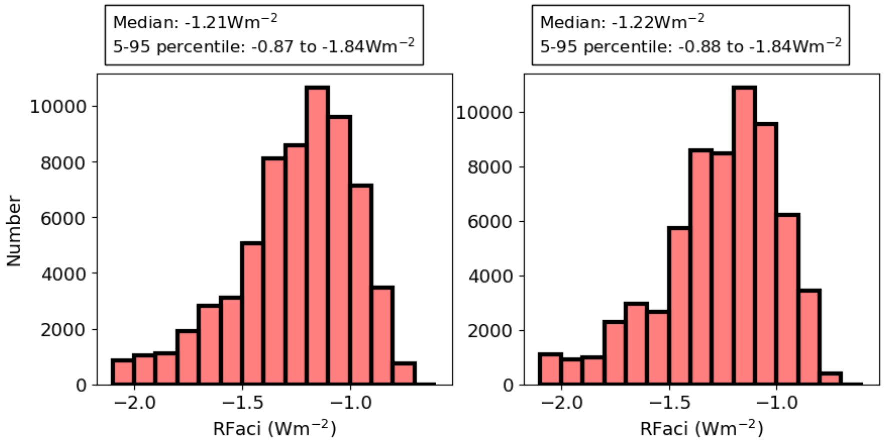


Figure 4.7: Histograms of the estimate of IRF_{aci} . The histograms are made using the different domain to global ratios, the retrieval error, and the different model estimates of ΔN_{ccn}. Left is for the meteorological clusters and right is for the geographical regimes.

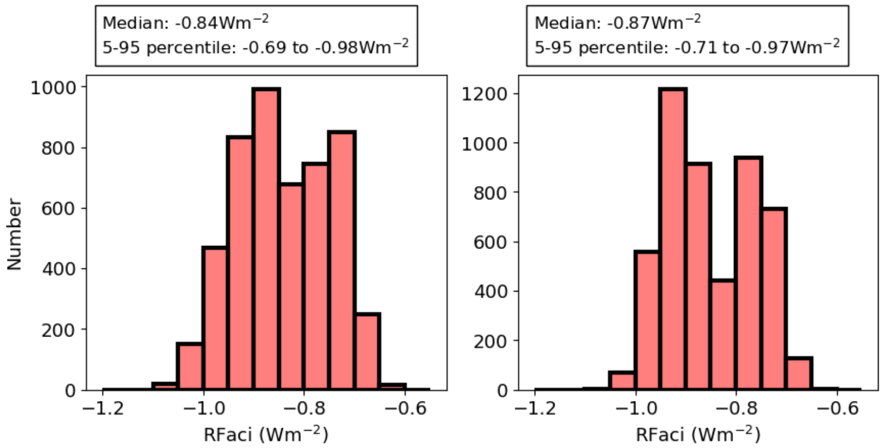


Figure 4.8: Histograms of the estimate of IRF_{aci} . The histograms are made using the the retrieval error, and the different model estimates of ΔN_{ccn}. Left is for the meteorological clusters and right is for the geographical regimes.

-0.87 compared to -0.84 for the clusters. The confidence interval is very similar for both.

To see how sensitive IRF_{aci} is to the choice of clusters, we recompute the global average for different values of k. Figure 4.9 shows the estimates of IRF_{aci} per cluster. The dots represent the global means of the different model estimates of ΔN_{ccn} , and the red dots are the mean. The spread of the black dots appears to be constant, but it does change slightly depending on where the model computes the highest values of ΔN_{ccn} .

The variations in the mean are small. The forcing seems to increase with increasing k, but the uncertainty in RF is dominated by the different model estimates. The lowest forcing is at $k = 3$, and the second lowest is at $k = 6$. These were also the 'best' clusters based on the clustering metrics. This could be a coincidence. For $k = 12$ we also had a local minimum or maximum in the clustering metrics, but as the number of clusters increases, the uncertainty in the susceptibility also increases.

The forcing estimates for the different models has a maximum spread of 0.23W m^{-2}while the variations in the mean due to changing k is 0.054W m^{-2}. From this we can conclude that the computation of IRF_{aci} is not very sensitive to the number of clusters. This is primarily due to the large spread in the ΔN_{ccn} estimates. If the quality of the retrievals improves and the amount of data increases, we might see stronger variability.

From this analysis, studying the radiative forcing from the cloud albedo effect using meteorological clusters instead of geographical regimes does not have significant advantages. There are differences in the computed susceptibilities which leads to large regional differences in the computed forcing, but for

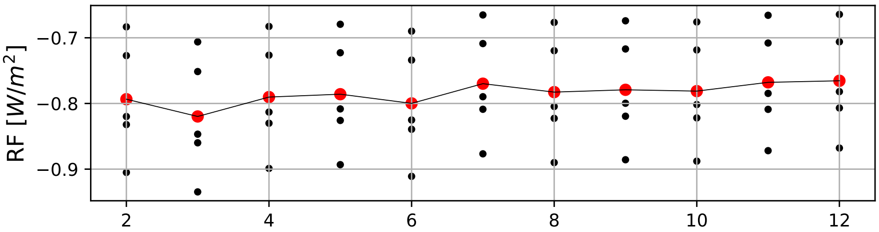


Figure 4.9: The IRF_{aci} as function of cluster number k. The black dots are the estimates for different ΔN_{ccn} estimates, and the red dots are the mean value. The line connects the means.

the global average this has no effect. Improving our understanding of what this relationship means and to bring microscale and satellite scales closer together, meteorological clustering could help. Specifically by continuing to study which meteorological parameters affect the susceptibility the most.

In Appendix B, the susceptibilities for $k = 3$, $k = 8$, and $k = 12$ are plotted. The values of the susceptibilities do not change significantly. For $k = 12$ there are clusters with high (0.77) and low (0.41) susceptibilities that are comparable to the geographical regimes. However, the uncertainty is large as the data is split up in small chunks.

5

Cloud Adjustments to the cloud albedo effect and the implied forcing

For the cloud adjustments we analyse the relation of the CF and LWP with N_d. We can then relate the total framework by linking relationships: N_{ccn} affects N_d, which in turn affects LWP and CF. We can use the relations to estimate the effective radiative forcing of aerosol-cloud interactions ERF_{aci}. The reason that we are not directly relating N_{ccn} to LWP and CF is the limitation in the sampling of N_{ccn} mentioned in Chapter 2.

To assess the adjustments to the cloud albedo effect, N_d is retrieved using the joint histogram of τ_c (cloud optical thickness) and r_e (cloud drop effective radius) (Gryspeerdt et al. 2016). This allows us to filter based on the level 2 information. The joint histogram gives the amount of L2 pixels with a specific τ_c and r_e. We then filter out low values ($\tau_c > 4$ and $r_e > 4$ μm), and compute the new mean τ_c and r_e for that grid box by computing the mean. We also filter out grid boxes with high solar zenith angle and high sensor zenith angle, as these lead to overestimation of r_e. As we do not have L2 data or joint histograms of the angles, we limit the mean plus the standard deviation of the grid box, in other words: $SolZA_{mean} + SolZA_{std} < 65°$ and $SenZA_{mean} + SenZA_{std} < 41°$, where SolZA is the solar zenith angle and SenZA is the sensor zenith angle. Then we recompute N_d for the filtered data, as described in Chapter 2.

5.1. Liquid water path adjustments

An important measure to assess the adjustment to liquid water path is the susceptibility of LWP to N_d. Similarly to the cloud albedo effect, this susceptibility is defined as:

$$\beta_{\ln\mathcal{L}-\ln N_d} = \frac{\partial \ln LWP}{\partial \ln N_d}. \tag{5.1}$$

Typical ranges are -0.36 to -0.011 (Bellouin et al. 2020; Toll et al. 2017; Gryspeerdt et al. 2019). The strength of the relationship depends on the relative magnitude of two processes: precipitation suppression and entrainment enhancement. A decrease in droplet size (i.e. increase in N_d) due to an aerosol perturbation reduces precipitation rates, which leads to an increase in LWP. Entrainment is enhanced by increased cloud top turbulence and cloud top cooling due to evaporation which leads to a decrease in LWP. Meteorology affects the strength of these processes and thus affects the relationship strength of LWP and aerosol. This necessitates a measure of meteorological variation, which in our study is the meteorological clustering approach.

For the LWP adjustments, an additional filtering step for N_d is done based on cloud retrieval fraction. In broken cloud formations, cloud inhomogeneity can lead to a bias in the retrieved r_e. This could lead to an underestimation in N_d, but some studies have found an increase of N_d with cloud inhomogeneity (Grosvenor and Wood, 2014). To limit bias in CDNC, we only use N_d retrievals where the cloud retrieval fraction if above 65%.

Figure 5.1 shows the joint histogram of LWP and CDNC. The N_d bins are normalised to unity so that they represent conditional probabilities. The dots represent the median LWP in a N_d bin. The solid lines represent to the linear regressions of binned data left and right of the maximum. The value of the slopes are called S_l and S_r respectively, standing for left and right. The left slope represents the precipitation regime and the right slope represents the entrainment regime. The white dotted line represents the linear fit through all points. Similar to other studies (Michibata et al., 2016; Gryspeerdt et al., 2019), the global susceptibility is negative.

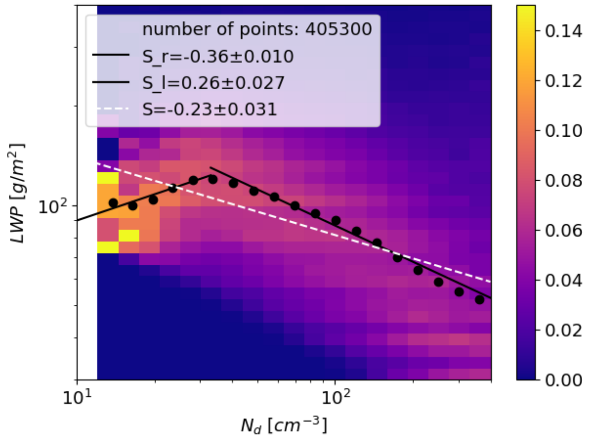


Figure 5.1: Joint histogram of LWP and CDNC. The black dots represent the median value per N_d bin. The dotted white line represents through the points represents the susceptibility $\beta_{\ln \mathcal{L} - \ln N_d}$.

Figure 5.2 shows the joint histograms per cluster. Note that the x-axis ranges from 10 to 400 cm^{-3}. The overall shape of the plots are similar, but there are some differences between clusters. Looking at C0, we see a slope of 0.24 in the precipitation regime and a strong slope of -0.39 in the entrainment regime. C0 is characterised by low free tropospheric humidity, which could enhance cloud top entrainment evaporation. Comparing to C1 we see a slightly higher precipitation regime slope (0.29) and a similar entrainment regime slope (-0.37). The main difference between the two clusters is lower EIS values for C0, indicating less stability and less stratiform cloud formations. The clusters have similar values for RH. It appears that C1 is more peaked around the maximum, while C0 is more spread out. C1 could suffer more from retrieval bias in N_d due to cloud inhomogeneity which could affect the precipitation regime slope.

For C2 we see a very strong precipitation regime slope of 0.51, indicating strong precipitation suppression, but the fit is uncertain due to the low number of data points there. The entrainment slope is lower (0.26) compared to C0 and C1. C2 has quite high free tropospheric humidity, which could dampen cloud top entrainment evaporation. At very high N_d we see an increase in LWP, which is possibly a retrieval bias in r_e or τ_c. We see this same effect for C3 and C4. This effect might cause an underestimation of the susceptibility.

For C3 and C4, we see weaker responses with $S_l = 0.20$ and $S_r = -0.21$ for C3 and $S_l = 0.29$ and $S_r = -0.17$ for C4. C3 has high free tropospheric humidity, and C4 has moderate free tropospheric humidity with a mean RH_{700} of 0.40. What sets these clusters apart is higher surface wind-speed. Gerber et al. (2013) have shown that cloud top evaporation decreases with increasing wind shear for stratocumulus clouds. We do control for surface wind speed, but not for wind shear at cloud top. The joint histogram of C5 is not well resolved due to the small size of the cluster.

From this we can conclude that free tropospheric humidity appears to be important for the entrainment regime, where clusters with low humidity have a steeper slope at high N_d values. This holds well for C0, C1, C2, and C3, but C4 does not have this steep slope, even though the RH is somewhat low. There could be other processes that control the LWP. EIS does not seem to control the slope much, as C0 and C1 are very similar, but the EIS for these two clusters is different. The highest value of S_l is the value of C2. C1 and C4 also have high slopes (0.29), but these lie close to the value of C0 (0.24). C2 is also the cluster with the lowest values for EIS. C1 and C4 also have lower values for EIS. It appears that precipitation suppression is more dominant in low EIS clusters, where there could be more cumulus formation, as EIS is correlated with stratiform cloud cover. The signal, however, is not very strong. The other parameters (Ω_{700} , surface temperature) do not appear to affect this relationship much. For example. C3 and C4 have very different values for Ω_{700} , but the slopes are very similar between the two.

5.2. Cloud fraction adjustments

Aerosols can affect cloud fraction by reducing the radius of cloud droplets. By suppressing precipitation, cloud lifetime could increase or the transition from closed- to open-celled stratocumulus could be changed. Many studies have found correlations between CF and aerosol proxies, but meteorological

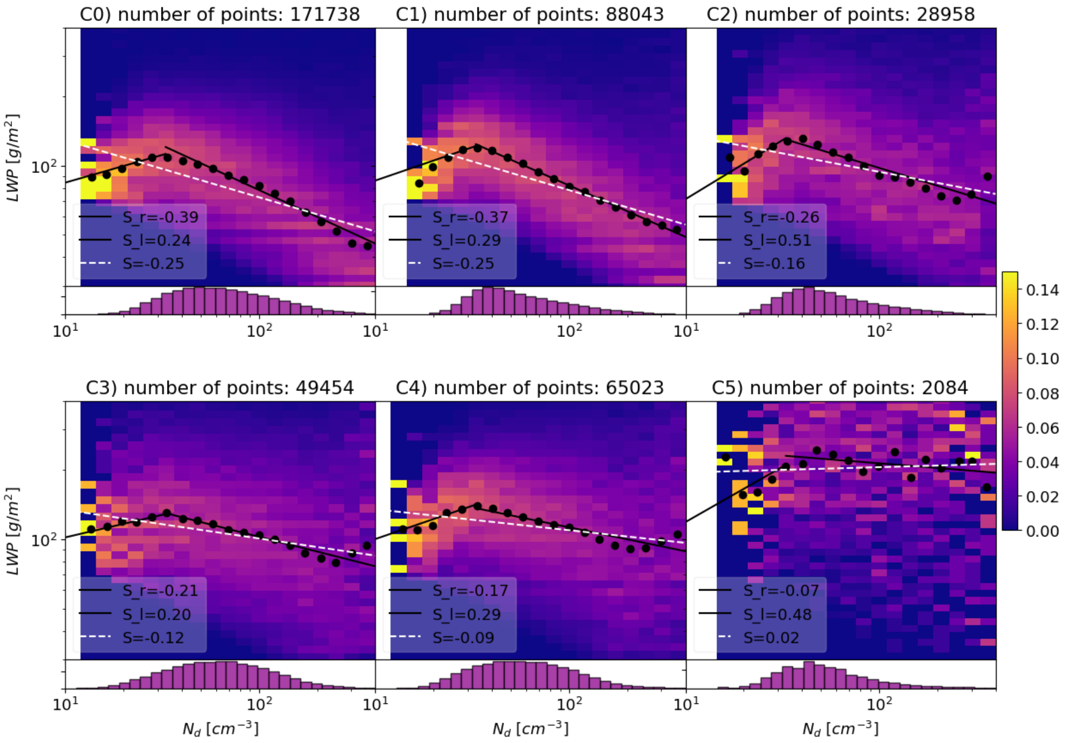


Figure 5.2: Joint histograms of LWP and CDNC per cluster.

covariation and retrieval errors affect the relationships. The susceptibility of CF to N_d is defined as:

$$\beta_{C-\ln N_d} = \frac{\partial C}{\partial \ln N_d}. \tag{5.2}$$

Note that this susceptibility is not in log-log space, but in linear-log space. The distribution of CF is more logical in linear space as it varies between 0 and 1 with high concentration close to 0 and 1. Typical values for $\beta_{C-\ln N_d}$ are 0.0 to 0.15 based on LES and GCM simulations (Ghan et al., 2016; Zelinka et al., 2014; Seifert et al., 2015; Xue & Feingold, 2006). This relationship is also subject to covariation with meteorology, as meteorological variations can affect both CF and N_d. Our meteorological clusters could help overcome this covariation.

For the CF adjustments we can not apply a filtering on CF to reduce N_d retrieval bias due to cloud inhomogeneity, as we want to sample the entire distribution of CF. As mentioned above, the retrieval error can create a bias in N_d and either lead to an over or underestimation of $\beta_{C-\ln N_d}$. We will have to take this bias into account when computing the radiative forcing. In Figure 5.3 the joint histogram of CF and N_d. These are made similarly to the LWP-N_d histograms. Here, CF is the total optical cloud retrieval fraction for grid boxes with zero ice clouds. The N_d bins are normalised to one, the black dots are the means of the bin, and the line is the linear least-squares fit through the data points. We see high concentrations near CF = 0 and CF = 1. The mean values of the bin show a linear relationship that tapers of at high N_d.

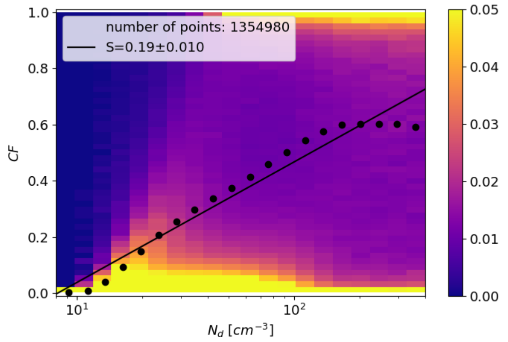


Figure 5.3: Joint histogram of CDNC and CF. The black dots represent the mean value per N_d bin. The linear fit through the points represents $\beta_{C-\ln N_d}$.

Figure 5.4 shows the joint histograms per clus-

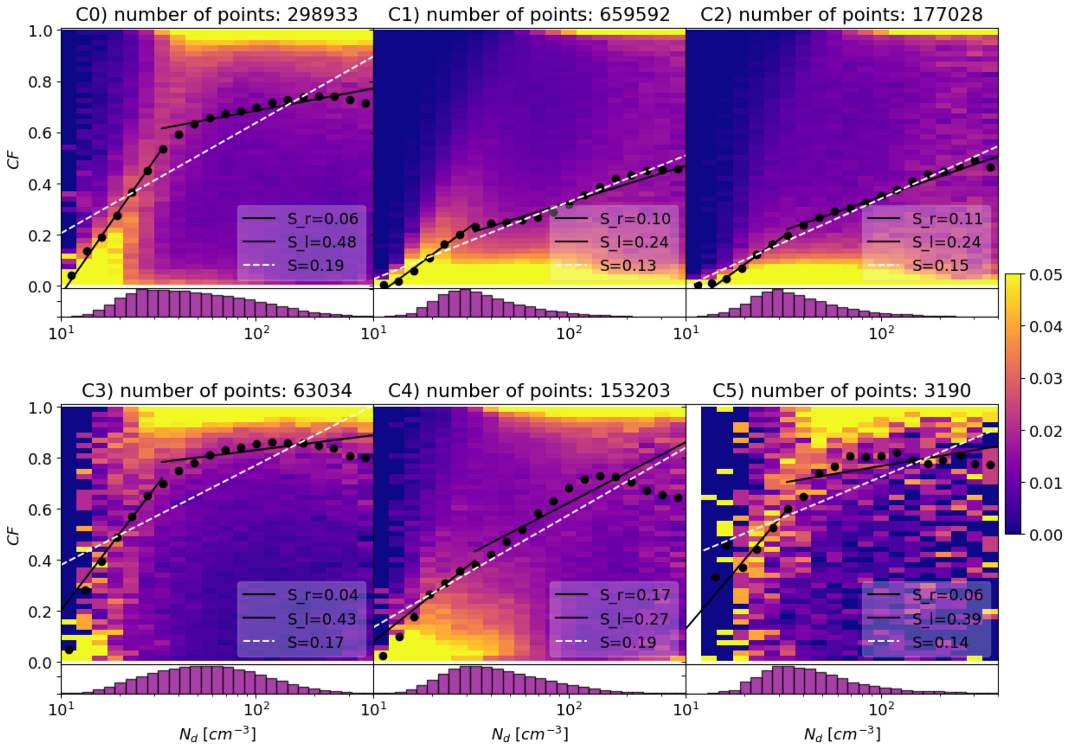


Figure 5.4: Joint histograms of cloud fraction and CDNC per cluster. The dots represent the mean in each CDNC bin. The bins are normalised to unity so that each CDNC bin gives a probability density of observing a specific CF. The histogram shows the normalised distribution of the retrieved CDNC.

ter. Each cluster shows a the two areas of high concentration, but there are significant differences. Here we have added a linear regression through all data points (white, dotted line) and through the 2 regimes shown in the LWP adjustments: precipitation and entrainment (solid, black lines). For C0 we see a strong increase in CF for low N_d concentrations with a slope of 0.48, after which the increases tapers off and at the end even decreases. We also see this decrease in CF for high N_d for C2, C3, and C4. As we have limited retrievals at high N_d concentrations (shown in the histograms), this reduction in CF could be a sampling issue, but it could also be a physical effect related to entrainment evaporation. The reduction in CF at high N_d concentrations could help explain the increase in LWP seen at high N_d concentrations in Figure 5.2. As CF decreases, cloud inhomogeneity increases which could lead to an overestimation of r_e. As LWP is linearly correlated with LWP, this would lead to an overestimation of LWP.

For C1 we see high concentrations of low CF for almost the entire N_d range and an area of high CF for $N_d > 100$ cm^{-3}. This indicates more broken cloud cumulus formations, while C0 would have many scenes with stratocumulus clouds, which is inline with previous chapters. There is a region for low N_d where C1 appears to follow the steep slope of C0, but the mean values of the bins are dominated by the high concentration at low CF. C2 looks very similar to C1 with almost identical slopes. The main similarity between these clusters (C1 and C2) is low EIS and high surface temperature. All other parameters are different. This could indicate that EIS and temperature are the most important meteorological confounders for CF adjustments.

Cluster 3 shows a similar pattern as C0, with even less points with low CF. The slopes are also very similar. Both C0 and C3 have high values for EIS. The joint histogram is less resolved due to low sample size and the CDNC distribution shows a peak for higher N_d values.

For C4 we see a combination of the two different patterns. The high concentration at low CF is more spread out compared to C0/C3, but not as extended as C1/C2. The high CF area is more spread out compared to C1/C2 but not as much as C0/C3. The reduction in CF for high N_d is most pronounced here. The joint histogram of C5 is poorly resolved, but shows a similar pattern to C0/C3.

For the C0 and C3, the low values of S_r indicate that these regimes are not very sensitive for changes in N_d for $N_d > 50$ cm^{-3}, while they are very sensitive at lower values of N_d. Cluster 3 does suffer from poor data coverage, so increases the amount of data could affect the relationship.

5.3. Effective Radiative Forcing

To compute the effective radiative forcing we use the framework presented by Bellouin et al. (2020). The radiative forcing of LWP adjustments is computed using:

$$RF_{\mathcal{L}} = -\frac{5}{6} F^{\downarrow} f_{liq} \alpha_{cld} (1 - \alpha_{cld}) \beta_{\ln \mathcal{L} - \ln N_d} \beta_{\ln N_d - \ln N_{ccn}} \Delta \ln N_{ccn}. \tag{5.3}$$

The formula is similar to the formula of the RF of the cloud albedo effect (Equation 4.4), except that there is a double sensitivity ($N_{ccn} \rightarrow N_d \rightarrow LWP$), and a factor 5/6 instead of 1/3, which comes from how LWP affects cloud albedo (Bellouin et al. 2020). The susceptibility $\beta_{\ln \mathcal{L} - \ln N_d}$ is the linear regression of the entire dataset (white dotted line in Figure 5.2). As we only have information about relative differences of N_{ccn} , we can not take the two different slopes into account.

The radiative effect of the CF adjustments is less straightforward than the LWP and cloud albedo forcing. The LWP and cloud albedo effect alter the cloud properties, while the CF adjustments affect the cloud fraction. This means that the CF adjustments can operate on any area where there are no overlying ice clouds. This means that instead of a liquid cloud fraction (f_{liq}), we need $(1 - f_{ice})$ as an initial cloud fraction (Gryspeerdt et al. 2016; Bellouin et al. 2020). This leads to the following equation:

$$RF_{C} = (1 - f_{ice}) \cdot f_{liq} \cdot (\alpha_{cld} - \alpha_{cs}) \beta_{C - \ln N_d} \beta_{\ln N_d - \ln N_{ccn}} \Delta \ln N_{ccn}. \tag{5.4}$$

We again have a double sensitivity, but now with CF ($N_{ccn} \rightarrow N_d \rightarrow CF$). The term $(\alpha_{cld} - \alpha_{cs})$ represents the difference between the overcast (α_{cld}) and clear-sky (α_{cs}) top of the atmosphere albedo from CERES for grid boxes with no ice clouds. The susceptibility $\beta_{C - \ln N_d}$ is the linear regression of the entire dataset (white dotted line in Figure 5.4). The effective radiative forcing can be calculated using:

$$ERF_{aci} = IRF_{aci} + RF_{C} + RF_{\mathcal{L}}, \tag{5.5}$$

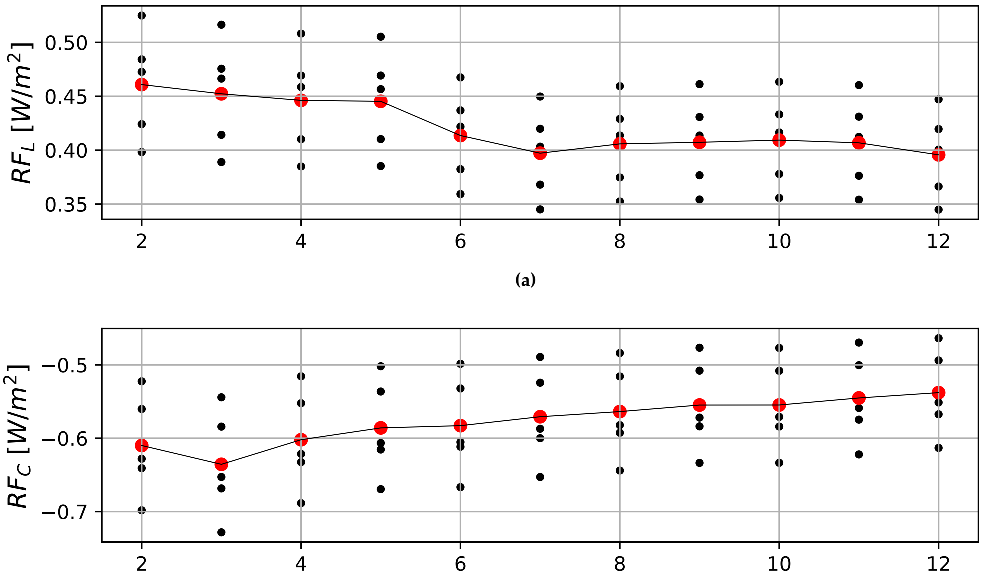
according to Bellouin et al. (2020). To compute a domain estimate, we again compute the area weighted average using Equation 4.8, but instead of IRF_{aci}, we use RF_C and $RF_{\mathcal{L}}$.

In Figure 5.5, the forcing estimates for the LWP and CF adjustments are plotted as a function of cluster amount k. The LWP adjustment does not show large variations. The mean of $RF_{\mathcal{L}}$ varies between 0.39 and 0.46 W m^{-2}. The forcing shows a decrease between $k = 5$ en $k = 7$. The values of RF_C shows variations in the mean between -0.64 and -0.54 W m^{-2}. The mean of the RF_C shows a monotonic increase after $k = 3$, but the range of estimates overlap for all k. These figures show that the uncertainty of ΔN_{ccn} dominates over variations due to choice of k. Optimising the methodology and studying the response of clouds to aerosol perturbations remains important as we aim to constrain ERF_{aci} using multiple lines of evidence. Understanding how cloud parameters respond to aerosol perturbations is one of the key problems.

Figure 5.6 shows maps of ERF_{aci} and its different components. Figure 5.6A is the same as the map shown in Chapter 4, but the colourbar is different. The global average forcing from the LWP adjustments are positive, with different model estimates varying between 0.36 and 0.47 W m^{-2}. The global average forcing from the different ΔN_{ccn} estimates for the CF adjustments are negative and vary between -0.50 and -0.67 W m^{-2}.

The spatial patterns are similar, as this is primarily influenced by the ΔN_{ccn} distribution from the models. The LWP adjustments (Figure 5.6B) show positive forcings. The effect is strongest for the stratocumulus decks on the west coast of Africa and the Americas. The CF adjustments (Figure 5.6C) is more defined in the Northern Pacific, specifically near China, and around the equator. This is also visible in the sum of the two adjustment terms (Figure 5.6D), where the forcing near the west coast of Africa and the Americas are positive. This indicates that the LWP adjustments are specifically important in these regions.

To asses the effect of retrieval error, we use a simulator to model the effect of errors on the derived susceptibilities. For RF_C we assume a relationship of $C = S \cdot np.log(N_d) - b$, where S=0.19 and b=-0.4. This assumes a perfect relationship and allows us to test the effect of errors. When we apply a 50% error on N_d, and recompute the susceptibility, we find a lower value of 0.13, indicating a underestimation



(a)

(b)

Figure 5.5: Figures of a) $RF_{\mathcal{L}}$ and b) RF_C as a function of clusters. The black dots indicate the different model estimates, the red dot shows the mean of the estimates, and the solid line connects the means.

of 30% due to the error in N_d. If we apply a 50% error on C, we find S equal to 0.19, indicating that random errors on CF do not affect the susceptibility, even though a bias could still affect the relationship. The bias due to cloud inhomogeneity can lead to a dependence of N_d on CF, which we model using the following equation:

$$N_d = N_d \cdot (1 + F(1 - C)) \tag{5.6}$$

where C is the cloud fraction and F is the strength of the bias. This formula offsets the underestimation of N_d for low values of CF. When we apply this bias with $F = [0.3, 0.5, 1.0, 2.0]$ and compute the susceptibility, we find higher susceptibilities between 0.20 and 0.23, indicating a underestimation of 5%-20%. However, when we first apply a 50% error on N_d and then apply the bias for the different values of F, we find lower susceptibilities showing an overestimation of the slope between 7% and 47%. This is caused by points with low values C reducing the means of the high N_d bins, as they are shifted to higher N_d values. Adding the 50% error in C, together with the other two error sources, does not affect the susceptibility significantly. The non-linearity of the relationship is not taken into account. When we apply the same tests to the actual measured C (instead of our ideal relationship) we find similar deviations. It is not clear which relationship is more physically plausible. To make a conservative estimate, we assume that the susceptibility is either underestimated or overestimated by 50%.

For the susceptibility of $\mathcal{L}$ to N_d, we assume a perfect relationship of $\mathcal{L} = cN_d^S$, with $S = -0.23$ and $c = 300$, and then study the effect of errors. When we apply a 50% random error on $\mathcal{L}$, we do not find a significant variation in the computed slope. If we apply a 50% random error on N_d we find a slope of -0.13 indicating an underestimation of the slope of 43%. This does not take the non-linearity of the relationship into account. When we apply the same 50% error on N_d for the measured values of $\mathcal{L}$, we compute a slope of -0.18, indicating an underestimation of 21%. As we applied additional filtering when computing the susceptibilities of the LWP adjustments, the effect of N_d bias due to cloud inhomogeneity should play a smaller role. However, we still test the effect by applying Equation 5.6 where we use the retrieved values of C per data point. Computing the slope for $F = [0.3, 0.5, 1.0, 2.0]$ shows small deviations from the 'true' value. For $F = 2$, we find a a slope of -0.19, indicating an overestimation of the slope by 17%

Additionally, as the distribution of N_d peaks between 30 and 80 cm^{-3}, the relationships could be affected by the low sampling at high and low N_d concentrations. For the LWP adjustments, resolving the low N_d concentrations could reduce the global susceptibility if the precipitation regime is stronger than implied by these findings. The effect of LWP increase at high N_d concentrations could be due to retrieval errors which would imply an underestimation of the LWP adjustment strength. From this we

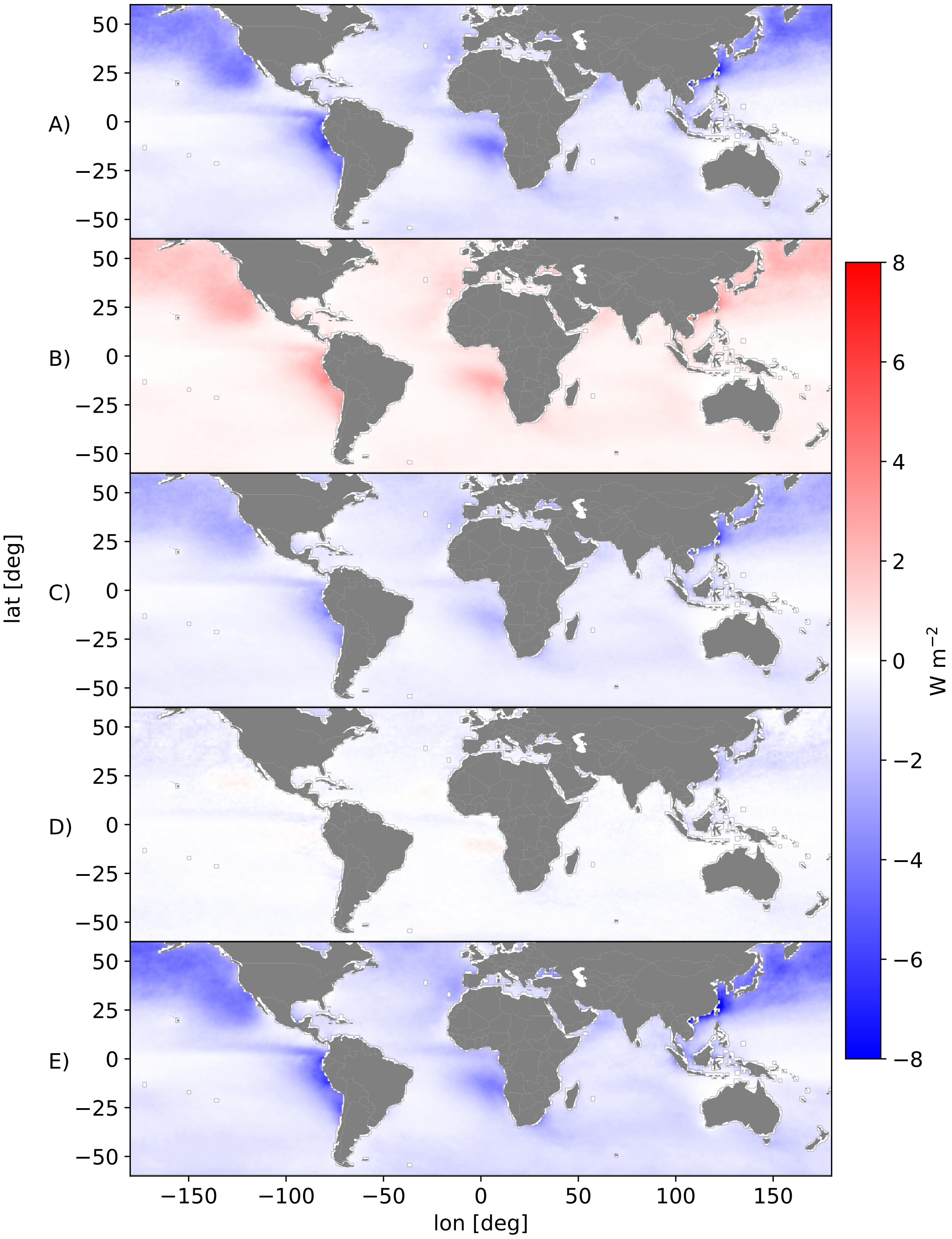


Figure 5.6: Maps showing A) the RF of the cloud albedo effect, B) the RF of the LWP adjustments, C) the RF of the CF adjustments, D) the RF of the adjustments: CF+LWP, and E) the effective radiative forcing ERF.

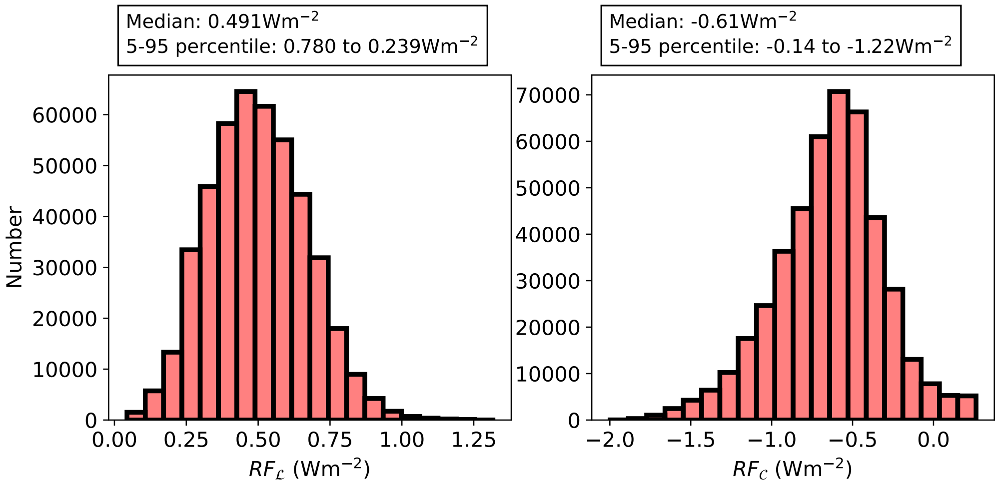


Figure 5.7: Histograms of the estimates of $RF_{\mathcal{L}}$ and RF_C.

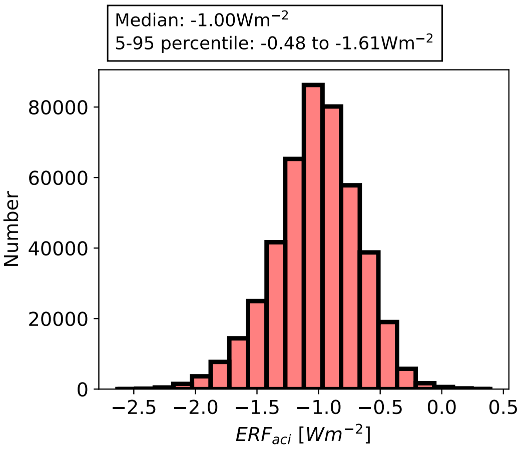


Figure 5.8: Histogram of possible estimates of ERF_{aci}.

estimate a possible underestimation of 50% and a possible overestimation of 25% for the susceptibility of the LWP adjustments.

The uncertainty in the susceptibility of the CF adjustments and the LWP adjustments are translated to forcing by assuming a linear dependence between susceptibility and forcing. For the CF adjustment we model the uncertainty as a normal distribution with mean 0 and standard deviation 0.5. For the LWP adjustments we model the uncertainty as a normal distribution with mean 1.125 and standard deviation 0.375. Combining these uncertainties with the uncertainty for $\beta_{\ln N_{ccn} - \ln N_d}$ presented in Chapter 4, and with the different model estimates of ΔN_{ccn} gives us a range of values shown in Figure 5.7. The median is taken as our best estimate and the 5 and 95 percentiles give the 90% confidence interval. This gives us the following estimates: $RF_{\mathcal{L}} = 0.49,\ 90\%\ CI\ [0.24,\ 0.78]$, $RF_C = -0.61,\ 90\%\ CI\ [-0.14,\ -1.22]$.

By randomly sampling the distributions of IRF_{aci}, RF_C, and $RF_{\mathcal{L}}$ we create a histogram of possible ERF_{aci} estimates. Figure 5.8 shows the histogram and gives us a best estimate of -1.0 W m^{-2} with a 90% confidence interval of $[-1.6, -0.48]$.

Figure 5.9 shows our domain estimates of ERF_{aci} and its components together with the result from Wall et al. (2023) and Bellouin et al. (2020) (hereafter W23 and B20). W23 use a shortwave radiative flux anomaly framework together with cloud-controlling factor analysis to estimate ERF_{aci} and its components. They computed two estimates, one with partly-cloudy MODIS pixels included. In Figure 5.9, we only include the cloudy-pixel estimate from W23. B20 combined all research up to that point in a 'most-likely' range for all the components, which is why the error bars are so large. The squares represent the best estimates and the error bars represent the confidence intervals. For the IRF_{aci}

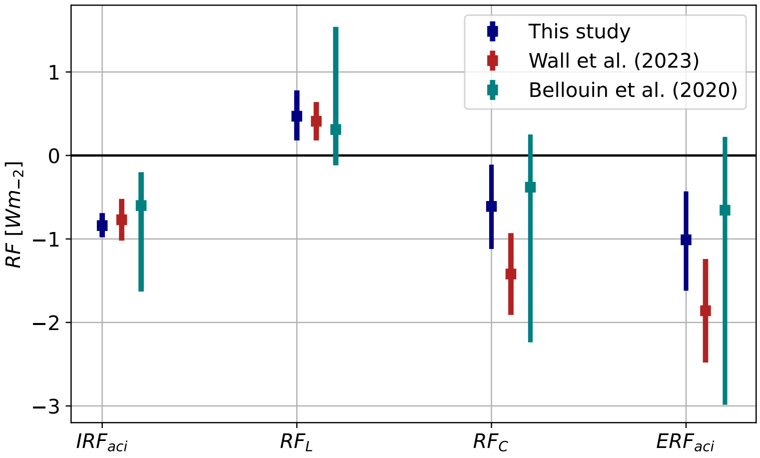


Figure 5.9: ERF_{aci} and its components estimated by this work, Wall et al. (2023), and Bellouin et al. (2020).

the estimates between our study and W23 are comparable, while our estimate is a little more negative and has smaller error bars. The 'best estimate' from B20 is more positive. For $RF_{\mathcal{L}}$, our estimate is more positive and our error bars are larger. For RF_C, W23 estimate a much more negative value compared to us. The error bars do overlap. The estimates of ERF_{aci} from our study and from W23 differ by 0.85W m^{-2}. Both our work and W23 indicate that the best estimate from B20 is too positive. Our value for RF_C is more negative compared to work by Gryspeerdt et al. (2016).

Y. Chen et al. (2022) found that the ERF_{aci} could be dominated by the CF effect, by studying the increase in cloud cover after a volcanic eruption. This is inline with the findings from W23 as they also estimate a very negative value for RF_C. If these estimates are accurate, than our methodology of estimating RF_C does not capture the magnitude of the effect, as our estimate is not as negative. It would also have important consequences for the climate sensitivity, as a very negative ERF_{aci} implies a strong climate sensitivity.

6

Conclusions, Limitations and Recommendations

6.1. Conclusions

In this thesis, the radiative forcing of aerosol-cloud interactions is investigated based on satellite data and meteorological clusters using cloud-controlling factors. Chapter 3 presented a description of the clustering and resulting cluster characteristics. Chapter 4 and Chapter 5 showed the sensitivities of cloud parameters to CCN perturbations per cluster and the implied global forcing. We have computed an effective radiative forcing from aerosol-cloud interactions ERF_{aci} over our domain of -1.0 W m^{-2} 90% CI $[-1.6, -0.48]$ since 1850. Our primary aim was to answer three research questions.

How do clusters of cloud-controlling factors correlate with cloud parameters?

By clustering cloud-controlling factors we defined unique regimes in which we studied clouds. The separation in cloud parameters per cluster is reasonable, but not very profound. The clustering separates regime where stratocumulus is dominant, regimes where trade wind cumulus are prevelant, and a regime where cirrostratus and deep convective clouds are present. There are characteristics in the distributions of cloud top temperature and cloud retrieval fraction per cluster. Specifically, there are two different regimes, corresponding to the stratocumulus and trade wind cumulus regimes.

Can we reduce uncertainty in the RF$_{\text{aci}}$ estimate from Hasekamp et al. 2019 when controlling for meteorology using clustering techniques?

The uncertainty in the RF$_{\text{aci}}$ estimate using meteorological clustering or geographical regimes is the same. The uncertainty is dominated by our lack of knowledge about the pre-industrial atmospheric aerosol state, and by retrieval errors. Variation in susceptibilities do not affect the forcing estimates much. The susceptibilities computed using clustering could be 'more true' compared to using the geographical regimes, but that is not possible to prove based on this work.

How do the aerosol-cloud interactions and cloud adjustments differ per cluster?

The susceptibilities of N_d to N_{ccn} per cluster are very similar. We found that the strength of relationship depends on the sampling across the globe, with differences between the southern and northern hemispheres. We found that retrieval errors at low CCN concentrations lead to underestimation of the susceptibility and that this could be specifically important for the southern hemisphere retrievals.

The liquid water path adjustments show differences between clusters. The strongest effect of entrainment evaporation is found in the clusters with low relative humidity at 700 hPa, which is in line with findings by Gryspeerdt et al. (2019). Wind speed also appears to be an important factor in the entrainment evaporation effect. Interpretation of the differences in precipitation suppression effect is hampered by the low data density at low CDNC values.

For the cloud fraction adjustments, the clusters seem to fall into two categories: primarily high values for CF , or primarily low values for CF. This leads to higher and lower susceptibilities respectively.

The dry mid-latitude cluster appears to be a transitional state between the two states. The relationship is most heavily influenced by EIS, because EIS is highly correlated with CF. Variations in RH between clusters does not seem to impact the relationship.

6.2. Limitations

The CDNC retrieval is influenced by bias in the satellite-retrieve effective radius for inhomogeneous clouds. Broken cloud formations can lead to an overestimation of the effective radius retrieved using 2.1-micron reflectances. As N_d is inversely proportional to $r_e^{5/2}$, this could lead to an underestimation of N_d. However, Grosvenor et al. (2014) found that N_d increases for broken clouds compared to closed cell clouds. The effect of this bias was tested using a simple simulator, but more investigation can be done on the effect of this error.

Another limitation is that the analysis of the adjustments is done with a different dataset than the analysis of the cloud albedo effect. This is unavoidable as the cloud albedo effect requires us to use both N_{ccn} and N_d retrievals in the same grid box. This limits the sample to grid boxes where the cloud fraction is around 0.2-0.6. Gryspeerdt et al. (2016) overcome this issue by interpolating AOD values neighbouring grid boxes with high CF. This could be an interesting route for future studies with N_{ccn} concentrations. Wall et al. (2023) used reanalysis data to estimate the CCN concentrations, which allows estimation of CCN levels in high CF grid boxes.

One of the cornerstones of this analysis is that we control for meteorological covariaton using clustering of reanalysis data. However, the confounding effect of meteorology could still affect the relationships within a cluster because of variations within a cluster, imperfections in the clustered data, or an imperfect clustering approach. This means that there could be spurious correlations that create biases in the susceptibilities we computed. It is unknown how much these possible correlations affect the result of this work, but they increase the uncertainty in the computed forcings.

The two adjustment relationships (LWP to N_d and CF to N_d) are both highly non-linear. Using a linear regression allows us to use susceptibilities to estimate the forcing since pre-industrial times, using a difference of CCN concentrations: $\Delta \ln N_{ccn}$. However, due to the non-linearity of the relationships, this approach could lead to inaccuracies. Using more elaborate relationships could prove to better capture the relationships, but we should consider what relationships we would expect from aggregated satellite data based on process scale understanding.

The clustering of cloud-controlling factors leads to a distinct separation of the meteorological data. Limitations of the method are that some of the parameters are not distributed as a Gaussian, while the k-means clustering algorithm assumes Gaussian distributed parameters. This could cause clusters to contain data points that would fit better in a different clusters.

6.3. Recommendations

- Possible future work can be done to improve the meteorological clustering by using the DBSCAN or OPTICS algorithms, as these do not assume normality (Saxena et al. 2017).
- Another factor that could possibly improve the clustering is how to use the vertical pressure velocity parameter from the reanalysis data: Ω. Differences in updraft conditions should strongly affect aerosol-cloud interactions. The Ω parameter in the reanalysis data only measures large-scale vertical motion over a grid box and shows large variations between adjacent time steps and between adjacent grid boxes. It is uncertain if this reanalysis parameter actually helps the clustering. For a future study, this could be investigated.
- It would be useful to analyse which cloud-controlling factor affects the separation in cloud parameters the most and whether other important factors are missing. For example, in our analysis, we did not include temperature advection across the ocean surface, which could be an additional important cloud controlling factor (Klein et al. 2018). Adding this parameter could potentially improve the clustering for the purpose of studying aerosol-cloud interactions.
- It would be beneficial to include more data in a future analysis. The N_{ccn} retrievals are limited by the fact that POLDER-3 moved away from the A-Train in December 2009, but including data up to that point would help improve the analysis.
- Redoing this analysis on a different reanalysis dataset could help test the robustness of this methodology. It would be interesting to see potential differences in the clustering.

References

Abraham, C. and C. Goldblatt (2022). "A satellite climatology of relative humidity profiles and outgoing thermal radiation over Earth's oceans". In: *Journal of the Atmospheric Sciences* 79 (6), pp. 2243–2265. ISSN: 15200469. DOI: 10.1175/JAS-D-21-0270.1.

Albrecht, B. A. (1989). "Aerosols, Cloud Microphysics, and Fractional Cloudiness". In: *Science* 245.4923, pp. 1227–1230. DOI: 10.1126/science.245.4923.1227. eprint: https://www.science.org/doi/pdf/10.1126/science.245.4923.1227. URL: https://www.science.org/doi/abs/10.1126/science.245.4923.1227.

Bellouin, N. et al. (2013). "Estimates of aerosol radiative forcing from the MACC re-analysis". In: *Atmospheric Chemistry and Physics* 13 (4), pp. 2045–2062. ISSN: 16807316. DOI: 10.5194/acp-13-2045-2013.

Bellouin, N. et al. (Mar. 2020). "Bounding Global Aerosol Radiative Forcing of Climate Change". In: *Reviews of Geophysics* 58 (1). ISSN: 19449208. DOI: 10.1029/2019RG000660.

Burrows, S. M. et al. (June 2022). *Ice-Nucleating Particles That Impact Clouds and Climate: Observational and Modeling Research Needs*. DOI: 10.1029/2021RG000745.

Carslaw, K. S. et al. (2013). "Large contribution of natural aerosols to uncertainty in indirect forcing". In: *Nature* 503.7474, pp. 67–71. ISSN: 1476-4687. DOI: 10.1038/nature12674. URL: https://doi.org/10.1038/nature12674.

Carslaw, K. S. et al. (2017). "Aerosols in the Pre-industrial Atmosphere". In: *Current Climate Change Reports* 3.1, pp. 1–15. ISSN: 2198-6061. DOI: 10.1007/s40641-017-0061-2. URL: https://doi.org/10.1007/s40641-017-0061-2.

Charlson, R. et al. (1991). "Perturbation of the northern hemisphere radiative balance by backscattering from anthropogenic sulfate aerosols". In: *Tellus A: Dynamic Meteorology and Oceanography* 43.4. DOI: 10.3402/tellusa.v43i4.11944. URL: https://doi.org/10.3402/tellusa.v43i4.11944.

Chen, Y. C. et al. (2011). "A comprehensive numerical study of aerosol-cloud-precipitation interactions in marine stratocumulus". In: *Atmospheric Chemistry and Physics* 11.18, pp. 9749–9769. DOI: 10.5194/acp-11-9749-2011. URL: https://acp.copernicus.org/articles/11/9749/2011/.

Chen, Y. et al. (Aug. 2022). "Machine learning reveals climate forcing from aerosols is dominated by increased cloud cover". In: *Nature Geoscience* 15 (8), pp. 609–614. ISSN: 17520908. DOI: 10.1038/s41561-022-00991-6.

Christensen, M. W. et al. (2017). "Unveiling aerosol-cloud interactions - Part 1: Cloud contamination in satellite products enhances the aerosol indirect forcing estimate". In: *Atmospheric Chemistry and Physics* 17.21. DOI: 10.5194/acp-17-13151-2017. URL: https://acp.copernicus.org/articles/17/13151/2017/.

Christensen, M. W. et al. (2022). "Opportunistic experiments to constrain aerosol effective radiative forcing". In: *Atmospheric Chemistry and Physics* 22 (1), pp. 641–674. ISSN: 16807324. DOI: 10.5194/acp-22-641-2022.

Conover, J. H. (1966). "Anomalous Cloud Lines". In: *Journal of Atmospheric Sciences* 23.6, pp. 778–785. DOI: https://doi.org/10.1175/1520-0469(1966)023<0778:ACL>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/atsc/23/6/1520-0469_1966_023_0778_acl_2_0_co_2.xml.

D'Alessandro, J. J. et al. (2023). "An Evaluation of Phase, Aerosol-Cloud Interactions and Microphysical Properties of Single- and Multi-Layer Clouds Over the Southern Ocean Using in Situ Observations From SOCRATES". In: *Journal of Geophysical Research: Atmospheres* 128.15. e2023JD038610 2023JD038610, e2023JD038610. DOI: https://doi.org/10.1029/2023JD038610. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2023JD038610. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023JD038610.

Doelling, D. R. et al. (2013). "Geostationary Enhanced Temporal Interpolation for CERES Flux Products". In: *Journal of Atmospheric and Oceanic Technology* 30.6, pp. 1072–1090. DOI: https://doi.org/10.1175/JTECH-D-12-00136.1. URL: https://journals.ametsoc.org/view/journals/atot/30/6/jtech-d-12-00136_1.xml.

Dusek, U. et al. (2006). "Size Matters More Than Chemistry for Cloud-Nucleating Ability of Aerosol Particles". In: *Science* 312.5778, pp. 1375–1378. DOI: `10.1126/science.1125261`. eprint: `https://www.science.org/doi/pdf/10.1126/science.1125261`. URL: `https://www.science.org/doi/abs/10.1126/science.1125261`.

Evans, S. M. et al. (2012). "Identification and analysis of atmospheric states and associated cloud properties for Darwin, Australia". In: *Journal of Geophysical Research Atmospheres* 117 (6), pp. 1–12. ISSN: 01480227. DOI: `10.1029/2011JD017010`.

Farmer, G. T. and J. Cook (2013). "Earth's Energy Budget". In: *Climate Change Science: A Modern Synthesis: Volume 1 - The Physical Climate*. Dordrecht: Springer Netherlands, pp. 81–95. ISBN: 978-94-007-5757-8. DOI: `10.1007/978-94-007-5757-8_4`. URL: `https://doi.org/10.1007/978-94-007-5757-8_4`.

Feingold, G. et al. (2016). "New approaches to quantifying aerosol influence on the cloud radiative effect". In: *Proceedings of the National Academy of Sciences of the United States of America* 113 (21), pp. 5812–5819. ISSN: 10916490. DOI: `10.1073/pnas.1514035112`.

Forster, P. et al. (2021). "The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity". In: ed. by V. Masson-Delmotte et al. Cambridge University Press. URL: `https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_07.pdf`.

Gelaro, R. et al. (2017). "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)". In: *Journal of Climate* 30.14, pp. 5419–5454. DOI: `https://doi.org/10.1175/JCLI-D-16-0758.1`. URL: `https://journals.ametsoc.org/view/journals/clim/30/14/jcli-d-16-0758.1.xml`.

Gerber, H. et al. (2013). "Entrainment rates and microphysics in POST stratocumulus". In: *Journal of Geophysical Research: Atmospheres* 118.21, pp. 12, 094–12, 109. DOI: `https://doi.org/10.1002/jgrd.50878`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50878`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50878`.

Ghan, S. et al. (2016). "Challenges in constraining anthropogenic aerosol effects on cloud radiative forcing using present-day spatiotemporal variability". In: *Proceedings of the National Academy of Sciences* 113.21, pp. 5804–5811. DOI: `10.1073/pnas.1514036113`. eprint: `https://www.pnas.org/doi/pdf/10.1073/pnas.1514036113`. URL: `https://www.pnas.org/doi/abs/10.1073/pnas.1514036113`.

Glassmeier, F. et al. (2021). "Aerosol-cloud-climate cooling overestimated by ship-track data". In: *Science* 371, p. 2023. URL: `https://www.science.org`.

Gorman, E. T. et al. (2019). "The NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission: an emerging era of global, hyperspectral Earth system remote sensing". In: *Sensors, Systems, and Next-Generation Satellites XXIII*. Ed. by S. P. Neeck, P. Martimort, and T. Kimura. Vol. 11151. International Society for Optics and Photonics. SPIE, 111510G. DOI: `10.1117/12.2537146`. URL: `https://doi.org/10.1117/12.2537146`.

Govender, P. and V. Sivakumar (2020). "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)". In: *Atmospheric Pollution Research* 11 (1), pp. 40–56. ISSN: 1309-1042. DOI: `https://doi.org/10.1016/j.apr.2019.09.009`. URL: `https://www.sciencedirect.com/science/article/pii/S1309104219304556`.

Grosvenor, D. P. and R. Wood (2014). "The effect of solar zenith angle on MODIS cloud optical and microphysical retrievals within marine liquid water clouds". In: *Atmospheric Chemistry and Physics* 14.14, pp. 7291–7321. DOI: `10.5194/acp-14-7291-2014`. URL: `https://acp.copernicus.org/articles/14/7291/2014/`.

Grosvenor, D. P. et al. (2018). "Remote sensing of droplet number concentration in warm clouds: A review of the current state of knowledge and perspectives". In: *Reviews of Geophysics* 56.2, pp. 409–453.

Gryspeerdt, E., J. Quaas, and N. Bellouin (2016). "Constraining the aerosol influence on cloud fraction". In: *Journal of Geophysical Research* 121 (7), pp. 3566–3583. ISSN: 21562202. DOI: `10.1002/2015JD023744`.

Gryspeerdt, E. et al. (July 2015). "Wet scavenging limits the detection of aerosol effects on precipitation". In: *Atmospheric Chemistry and Physics* 15 (13), pp. 7557–7570. ISSN: 16807324. DOI: `10.5194/acp-15-7557-2015`.

Gryspeerdt, E. and P. Stier (2012). "Regime-based analysis of aerosol-cloud interactions". In: *Geophysical Research Letters* 39 (21), pp. 1–5. ISSN: 00948276. DOI: `10.1029/2012GL053221`.

Gryspeerdt, E. et al. (2017). "Constraining the instantan eous aerosol influence on cloud albedo". In: *Proceedings of the National Academy of Sciences of the United States of America* 114 (19), pp. 4899–4904. ISSN: 10916490. DOI: `10.1073/pnas.1617765114`.

Gryspeerdt, E. et al. (Apr. 2019). "Constraining the aerosol influence on cloud liquid water path". In: *Atmospheric Chemistry and Physics* 19 (8), pp. 5331–5347. ISSN: 16807324. DOI: `10.5194/acp-19-5331-2019`.

Gryspeerdt, E. et al. (Jan. 2020). "Surprising similarities in model and observational aerosol radiative forcing estimates". In: *Atmospheric Chemistry and Physics* 20 (1), pp. 613–623. ISSN: 16807324. DOI: `10.5194/acp-20-613-2020`.

Hamilton, D. S. et al. (2014). "Occurrence of pristine aerosol environments on a polluted planet". In: *Proceedings of the National Academy of Sciences* 111.52, pp. 18466–18471. DOI: `10.1073/pnas.1415440111`. eprint: `https://www.pnas.org/doi/pdf/10.1073/pnas.1415440111`. URL: `https://www.pnas.org/doi/abs/10.1073/pnas.1415440111`.

Hasekamp, O. P., E. Gryspeerdt, and J. Quaas (2019). "Analysis of polarimetric satellite measurements suggests stronger cooling due to aerosol-cloud interactions". In: *Nature Communications* 10 (1), pp. 1–7. DOI: `10.1038/s41467-019-13372-2`. URL: `http://dx.doi.org/10.1038/s41467-019-13372-2`.

Hasekamp, O. P., P. Litvinov, and A. Butz (2011). "Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements". In: *Journal of Geophysical Research Atmospheres* 116 (14), pp. 1–13. ISSN: 01480227. DOI: `10.1029/2010JD015469`.

Jia, H., X. Ma, and Y. Liu (2019). "Exploring aerosol–cloud interaction using VOCALS-REx aircraft measurements". In: *Atmospheric Chemistry and Physics* 19.12, pp. 7955–7971. DOI: `10.5194/acp-19-7955-2019`. URL: `https://acp.copernicus.org/articles/19/7955/2019/`.

Klein, S. A. and D. L. Hartmann (1993). "The Seasonal Cycle of Low Stratiform Clouds". In: *Journal of Climate* 6.8, pp. 1587–1606. DOI: `https://doi.org/10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/clim/6/8/1520-0442_1993_006_1587_tscols_2_0_co_2.xml`.

Klein, S. A. et al. (2018). "Low-Cloud Feedbacks from Cloud-Controlling Factors: A Review". In: *Shallow Clouds, Water Vapor, Circulation, and Climate Sensitivity*. Ed. by R. Pincus et al. Cham: Springer International Publishing, pp. 135–157. DOI: `10.1007/978-3-319-77273-8_7`. URL: `https://doi.org/10.1007/978-3-319-77273-8_7`.

Koren, I., G. Feingold, and L. A. Remer (2010). "The invigoration of deep convective clouds over the Atlantic: aerosol effect, meteorology or retrieval artifact?" In: *Atmospheric Chemistry and Physics* 10.18. DOI: `10.5194/acp-10-8855-2010`. URL: `https://acp.copernicus.org/articles/10/8855/2010/`.

Kuma, P. et al. (2023). "Machine learning of cloud types in satellite observations and climate models". In: *Atmospheric Chemistry and Physics* 23.1, pp. 523–549. DOI: `10.5194/acp-23-523-2023`. URL: `https://acp.copernicus.org/articles/23/523/2023/`.

L'Ecuyer, T. S. et al. (2019). "Reassessing the Effect of Cloud Type on Earth's Energy Balance in the Age of Active Spaceborne Observations. Part I: Top of Atmosphere and Surface". In: *Journal of Climate* 32.19, pp. 6197–6217. DOI: `https://doi.org/10.1175/JCLI-D-18-0753.1`. URL: `https://journals.ametsoc.org/view/journals/clim/32/19/jcli-d-18-0753.1.xml`.

Lebsock, M. D., G. L. Stephens, and C. Kummerow (2008). "Multisensor satellite observations of aerosol effects on warm clouds". In: *Journal of Geophysical Research: Atmospheres* 113.D15. DOI: `https://doi.org/10.1029/2008JD009876`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD009876`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD009876`.

Ma, P. L. et al. (2018). "Observational constraint on cloud susceptibility weakened by aerosol retrieval limitations". In: *Nature Communications* 9.1, p. 2640. DOI: `10.1038/s41467-018-05028-4`. URL: `https://doi.org/10.1038/s41467-018-05028-4`.

Ma, X., F. Yu, and J. Quaas (Sept. 2014). "Reassessment of satellite-based estimate of aerosol climate forcing". In: *Journal of Geophysical Research* 119 (17), pp. 10, 394–10, 409. ISSN: 21562202. DOI: `10.1002/2014JD021670`.

Marchant, B. et al. (2016). "MODIS Collection 6 shortwave-derived cloud phase classification algorithm and comparisons with CALIOP". In: *Atmospheric Measurement Techniques* 9.4, pp. 1587–1599. DOI: `10.5194/amt-9-1587-2016`. URL: `https://amt.copernicus.org/articles/9/1587/2016/`.

McComiskey, A. et al. (2009). "An assessment of aerosol-cloud interactions in marine stratus clouds based on surface remote sensing". In: *Journal of Geophysical Research: Atmospheres* 114.D9. DOI: `https://doi.org/10.1029/2008JD011006`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD011006`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD011006`.

McCoy, D. T. et al. (2015). “Natural aerosols explain seasonal and spatial patterns of Southern Ocean cloud albedo”. In: *Science Advances* 1.6, e1500157. doi: 10.1126/sciadv.1500157. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.1500157. url: https://www.science.org/doi/abs/10.1126/sciadv.1500157.

McCoy, I. L. et al. (2020). “The hemispheric contrast in cloud microphysical properties constrains aerosol forcing”. In: *Proceedings of the National Academy of Sciences* 117.32, pp. 18998–19006. doi: 10.1073/pnas.1922502117. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1922502117. url: https://www.pnas.org/doi/abs/10.1073/pnas.1922502117.

Mishchenko, M. I. et al. (2007). “Accurate Monitoring of Terrestrial Aerosols and Total Solar Irradiance: Introducing the Glory Mission”. In: *Bulletin of the American Meteorological Society* 88.5, pp. 677–692. doi: https://doi.org/10.1175/BAMS-88-5-677. url: https://journals.ametsoc.org/view/journals/bams/88/5/bams-88-5-677.xml.

Mülmenstädt, J. and G. Feingold (Mar. 2018). “The Radiative Forcing of Aerosol–Cloud Interactions in Liquid Clouds: Wrestling and Embracing Uncertainty”. In: *Current Climate Change Reports* 4 (1), pp. 23–40. issn: 21986061. doi: 10.1007/s40641-018-0089-y.

Nakajima, T. et al. (2001). “A possible correlation between satellite-derived cloud and aerosol microphysical parameters”. In: *Geophysical Research Letters* 28.7, pp. 1171–1174. doi: https://doi.org/10.1029/2000GL012186. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2000GL012186. url: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GL012186.

Nishant, N. and S. C. Sherwood (2017). “A cloud-resolving model study of aerosol-cloud correlation in a pristine maritime environment”. In: *Geophysical Research Letters* 44.11, pp. 5774–5781. doi: https://doi.org/10.1002/2017GL073267. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL073267. url: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL073267.

Platnick, S., M. King, and P. Hubanks (2017a). *MODIS Atmosphere L3 Daily Product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center*. doi: 10.5067/MODIS/MOD08_D3.061.

Platnick, S. et al. (2017b). “The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua”. In: *IEEE Transactions on Geoscience and Remote Sensing* 55.1, pp. 502–525. doi: 10.1109/TGRS.2016.2610522.

Prospero, J. M. et al. (1983). “The atmospheric aerosol system: An overview”. In: *Reviews of Geophysics* 21.7, pp. 1607–1629. doi: https://doi.org/10.1029/RG021i007p01607. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/RG021i007p01607. url: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/RG021i007p01607.

Quaas, J., O. Boucher, and U. Lohmann (2006). “Constraining the total aerosol indirect effect in the LMDZ and ECHAM4 GCMs using MODIS satellite data”. In: *Atmospheric Chemistry and Physics* 6.4, pp. 947–955. doi: 10.5194/acp-6-947-2006. url: https://acp.copernicus.org/articles/6/947/2006/.

Quaas, J. et al. (2009). “Aerosol indirect effects - general circulation model intercomparison and evaluation with satellite data”. In: *Atmospheric Chemistry and Physics* 9 (22), pp. 8697–8717. issn: 16807324. doi: 10.5194/acp-9-8697-2009.

Quaas, J. et al. (2008). “Satellite-based estimate of the direct and indirect aerosol climate forcing”. In: *Journal of Geophysical Research: Atmospheres* 113.D5.

Raes, F. et al. (2000). “Formation and cycling of aerosols in the global troposphere”. In: *Atmospheric Environment* 34.25, pp. 4215–4240. issn: 1352-2310. doi: https://doi.org/10.1016/S1352-2310(00)00239-9. url: https://www.sciencedirect.com/science/article/pii/S1352231000002399.

Ramanathan, V. and G. Carmichael (2008). “Global and regional climate changes due to black carbon”. In: *Nature Geoscience* 1.4, pp. 221–227. issn: 1752-0908. doi: 10.1038/ngeo156. url: https://doi.org/10.1038/ngeo156.

Romps, D. M. (2017). “Exact expression for the lifting condensation level”. In: *Journal of the Atmospheric Sciences* 74 (12), pp. 3891–3900. issn: 15200469. doi: 10.1175/JAS-D-17-0102.1.

Rossow, W. B. and R. A. Schiffer (1991). “ISCCP Cloud Data Products”. In: *Bulletin of the American Meteorological Society* 72.1, pp. 2–20. doi: https://doi.org/10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2. url: https://journals.ametsoc.org/view/journals/bams/72/1/1520-0477_1991_072_0002_icdp_2_0_co_2.xml.

Saxena, A. et al. (2017). “A review of clustering techniques and developments”. In: *Neurocomputing* 267, pp. 664–681. ISSN: 0925-2312. DOI: `https://doi.org/10.1016/j.neucom.2017.06.053`. URL: `https://www.sciencedirect.com/science/article/pii/S0925231217311815`.

Schmale, J. et al. (2019). “Overview of the Antarctic Circumnavigation Expedition: Study of Preindustrial-like Aerosols and Their Climate Effects (ACE-SPACE)”. In: *Bulletin of the American Meteorological Society* 100.11, pp. 2260–2283. DOI: `https://doi.org/10.1175/BAMS-D-18-0187.1`. URL: `https://journals.ametsoc.org/view/journals/bams/100/11/bams-d-18-0187.1.xml`.

Schutgens, N. A. J., D. G. Partridge, and P. Stier (2016). “The importance of temporal collocation for the evaluation of aerosol models with observations”. In: *Atmospheric Chemistry and Physics* 16.2, pp. 1065–1079. DOI: `10.5194/acp-16-1065-2016`. URL: `https://acp.copernicus.org/articles/16/1065/2016/`.

Scott, R. C. et al. (2020). “Observed sensitivity of low-cloud radiative effects to meteorological perturbations over the global oceans”. In: *Journal of Climate* 33 (18), pp. 7717–7734. ISSN: 08948755. DOI: `10.1175/JCLI-D-19-1028.1`.

Seinfeld, J. H. et al. (2016). “Improving our fundamental understanding of the role of aerosol- cloud interactions in the climate system”. In: *Proceedings of the National Academy of Sciences* 113.21, pp. 5781–5790.

Shupe, M. D. et al. (2016). “Cloud Property Retrievals in the ARM Program”. In: *Meteorological Monographs* 57, pp. 19.1–19.20. DOI: `https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0030.1`. URL: `https://journals.ametsoc.org/view/journals/amsm/57/1/amsmonographs-d-15-0030.1.xml`.

Snedecor, G. W. and W. G. Cochran (1989). “Statistical Methods, eight edition”. In: *Iowa state University press, Ames, Iowa* 1191.2.

Stap, F. A., O. P. Hasekamp, and T. Röckmann (2015). “Sensitivity of PARASOL multi-angle photopolarimetric aerosol retrievals to cloud contamination”. In: *Atmospheric Measurement Techniques* 8.3, pp. 1287–1301. DOI: `10.5194/amt-8-1287-2015`. URL: `https://amt.copernicus.org/articles/8/1287/2015/`.

Stephens, G. L. et al. (2012). “An update on Earth’s energy balance in light of the latest global observations”. In: *Nature Geoscience* 5 (10), pp. 691–696. ISSN: 17520894. DOI: `10.1038/ngeo1580`.

Stevens, B. and G. Feingold (2009). “Untangling aerosol effects on clouds and precipitation in a buffered system”. In: *Nature* 461 (7264), pp. 607–613. ISSN: 00280836. DOI: `10.1038/nature08281`.

Stier, P. (2016). “Limitations of passive remote sensing to constrain global cloud condensation nuclei”. In: *Atmospheric Chemistry and Physics* 16.10, pp. 6595–6607. DOI: `10.5194/acp-16-6595-2016`. URL: `https://acp.copernicus.org/articles/16/6595/2016/`.

Stratmann, F. et al. (2004). “Laboratory Studies and Numerical Simulations of Cloud Droplet Formation under Realistic Supersaturation Conditions”. In: *Journal of Atmospheric and Oceanic Technology* 21.6. DOI: `https://doi.org/10.1175/1520-0426(2004)021%3C0876:LSANSO%3E2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/atot/21/6/1520-0426_2004_021_0876_lsanso_2_0_co_2.xml`.

Stubenrauch, C. J. et al. (July 2013). “Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel”. In: *Bulletin of the American Meteorological Society* 94 (7), pp. 1031–1049. ISSN: 00030007. DOI: `10.1175/BAMS-D-12-00117.1`.

Tselioudis, G. et al. (2013). “Global weather states and their properties from passive and active satellite cloud retrievals”. In: *Journal of Climate* 26 (19), pp. 7734–7746. ISSN: 08948755. DOI: `10.1175/JCLI-D-13-00024.1`.

Twomey, S. (1977). “The Influence of Pollution on the Shortwave Albedo of Clouds”. In: *Journal of Atmospheric Sciences* 34 (7), pp. 1149–1152. DOI: `10.1175/1520-0469(1977)034<1149:TIOPOT>2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/atsc/34/7/1520-0469_1977_034_1149_tiopot_2_0_co_2.xml`.

Wagner, R. et al. (2009). “A review of optical measurements at the aerosol and cloud chamber AIDA”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 110.11. Light Scattering: Mie and More Commemorating 100 years of Mie’s 1908 publication, pp. 930–949. ISSN: 0022-4073. DOI: `https://doi.org/10.1016/j.jqsrt.2009.01.026`. URL: `https://www.sciencedirect.com/science/article/pii/S0022407309000326`.

Wall, C. J., T. Storelvmo, and A. Possner (2023). “Global observations of aerosol indirect effects from marine liquid clouds”. In: *Atmospheric Chemistry and Physics* 23.20, pp. 13125–13141. DOI: `10.5194/acp-23-13125-2023`. URL: `https://acp.copernicus.org/articles/23/13125/2023/`.

Wall, C. J. et al. (2022). “Assessing effective radiative forcing from aerosol-cloud interactions over the global ocean”. In: *Proceedings of the National Academy of Sciences of the United States of America* 119 (46). issn: 10916490. doi: `10.1073/pnas.2210481119`.

Williams, J. et al. (2002). “Application of the variability-size relationship to atmospheric aerosol studies: estimating aerosol lifetimes and ages”. In: *Atmospheric Chemistry and Physics* 2.2, pp. 133–145. doi: `10.5194/acp-2-133-2002`. url: `https://acp.copernicus.org/articles/2/133/2002/`.

Williams, K. D. and M. J. Webb (2009). “A quantitative performance assessment of cloud regimes in climate models”. In: *Climate Dynamics* 33.1, pp. 141–157. issn: 1432-0894. doi: `10.1007/s00382-008-0443-1`. url: `https://doi.org/10.1007/s00382-008-0443-1`.

Wood, R. (2012). “Stratocumulus Clouds”. In: *Monthly Weather Review* 140.8, pp. 2373–2423. doi: `https://doi.org/10.1175/MWR-D-11-00121.1`. url: `https://journals.ametsoc.org/view/journals/mwre/140/8/mwr-d-11-00121.1.xml`.

Wood, R. and C. S. Bretherton (2006). “On the relationship between stratiform low cloud cover and lower-tropospheric stability”. In: *Journal of Climate* 19 (24), pp. 6425–6432. issn: 08948755. doi: `10.1175/JCLI3988.1`.

Wood, R. et al. (2012). “Precipitation driving of droplet concentration variability in marine low clouds”. In: *Journal of Geophysical Research: Atmospheres* 117. doi: `https://doi.org/10.1029/2012JD018305`. url: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012JD018305`.

Yamaguchi, T., G. Feingold, and J. Kazil (2019). “Aerosol-Cloud Interactions in Trade Wind Cumulus Clouds and the Role of Vertical Wind Shear”. In: *Journal of Geophysical Research: Atmospheres* 124.22, pp. 12244–12261. doi: `https://doi.org/10.1029/2019JD031073`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019JD031073`. url: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD031073`.

Zelinka, M. D. et al. (2014). “Quantifying components of aerosol-cloud-radiation interactions in climate models”. In: *Journal of Geophysical Research: Atmospheres* 119.12, pp. 7599–7615. doi: `https://doi.org/10.1002/2014JD021710`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014JD021710`. url: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD021710`.

A

Formulas and constants

A.1. Constants

Name	
Gravitational acceleration	$g = 9.81$ m/s^2
Earth's radius	$R_E = 6.378 \cdot 10^6$ m
Gas constant of air	$R_a = 287.04$ J/kg/K
Gas constant for water vapour	$R_v = 461.5$ J/kg/K
Specific heat capacity of air	$c_p = 1003.6 J/kg/K$
Specific heat capacity of dry air	$c_{va} = 719$ J/kg/K
Specific heat capacity of water vapour	$c_{vv} = 1418$ J/kg/K
Specific heat capacity of liquid water	$c_{vl} = 4119$ J/kg/K
Specific heat capacity of solid water	$c_{vs} = 1861$ J/kg/K
Temperature at triple point of water	$T_{trip} = 273.16$ K
Pressure at triple point of water	$p_{trip} = 611.65$ Pa
Internal energy diff. between liquid and solid at the triple point	$E_{0v} = 2.3740 \cdot 10^6$ J/kg
Internal energy diff. between vapour and liquid at the triple point	$E_{0s} = 0.3337 \cdot 10^6$ J/kg
Latent heat of vaporisation	$L_v = 2.50 \cdot 10^6$ J/kg/K

A.2. Formulas

Here formulas used in this thesis that are note defined in the text are defined. Mostly these formulas concern the computation of the Estimated Inversion Strength (EIS; Wood et al. 2006). **Potential temperature**

$$\theta = T \left(\frac{p_0}{p} \right)^{R_a/c_p}, \tag{A.1}$$

Height at 700 hPa, assuming an exponential decrease in pressure with height with a single scale height

$$z_{700} = \frac{R_a T}{g} \log \left(\frac{p_0}{700\ \text{hPa}} \right) \tag{A.2}$$

Moist-adiabatic potential temperature gradient

$$\Gamma_m(T, p) = \frac{g}{c_p} \left[1 - \frac{1 + L_v q_s(T, p)/R_a T}{1 + L_v^2 q_s(T, p)/c_p R_a T^2} \right], \tag{A.3}$$

Saturated vapour pressure

$$e_s(T) = 6.11 \cdot 10^{7.5 \cdot T/(237.7+T)}, \tag{A.4}$$

Saturation mixing ratio

$$q_s(T,p) = 621.97 \cdot \frac{e_s(T)}{p + e_s(T)}, \tag{A.5}$$

Lifted condensation level. Romps (2017) derived an exact expression of the lifted condensation level assuming constant heat capacities.

$$T_{LCL} = c\left[W_{-1}(\mathrm{RH}_l^{1/a} c e^c)\right]^{-1} T, \tag{A.6a}$$

$$p_{LCL} = p\left(\frac{T_{LCL}}{T}\right)^{c_{pm}/R_m}, \tag{A.6b}$$

$$z_{LCL} = z + \frac{c_{pm}}{g}(T - T_{LCL}), \tag{A.6c}$$

$$a = \frac{c_{pm}}{R_m} + \frac{c_{vl} - c_{pv}}{R_v}, \tag{A.6d}$$

$$b = -\frac{E_{0v} - (c_{vv} - c_{vl})T_{trip}}{R_v T}, \tag{A.6e}$$

$$c = \frac{b}{a}, \tag{A.6f}$$

Two-sample t-test for equal means assuming unequal variances. $t_{1-a/2,v}$ is the critical value of the t distribution with v degrees of freedom at significance level a.

$$H_0 : \mu_1 = \mu_2, \tag{A.7a}$$

$$H_a : \mu_1 \neq \mu_2, \tag{A.7b}$$

$$t = \frac{\mu_1 - \mu_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}, \tag{A.7c}$$

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2(n_1 - 1) + (s_2^2/n_2)^2(n_2 - 1)}, \tag{A.7d}$$

where the null hypothesis is rejected if :

$$|T| > t_{1-a/2,v}. \tag{A.7e}$$

If equal variances are assumed, the formulas of t and v change to:

$$t = \frac{\mu_1 - \mu_2}{s_p\sqrt{1/n_1 + 1/n_2}}, \tag{A.8a}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \tag{A.8b}$$

$$v = n_1 + n_2 - 2, \tag{A.8c}$$

Liquid water path. Formal definition:

$$\mathcal{L} = \int_{z=0}^{h} q_L dz = \frac{1}{2} f_{ad}\Gamma_{ad}h^2 \tag{A.9}$$

Liquid water path estimated using satellite retrievals

$$\mathcal{L} = \frac{5}{9} f_{ad} r_e \tau_c \tag{A.10}$$

B

Different clustering

Here we show figures for different number of clusters. Specifically for $k = 3$ and $k = 12$. We also show the susceptibility plot for $k = 8$.

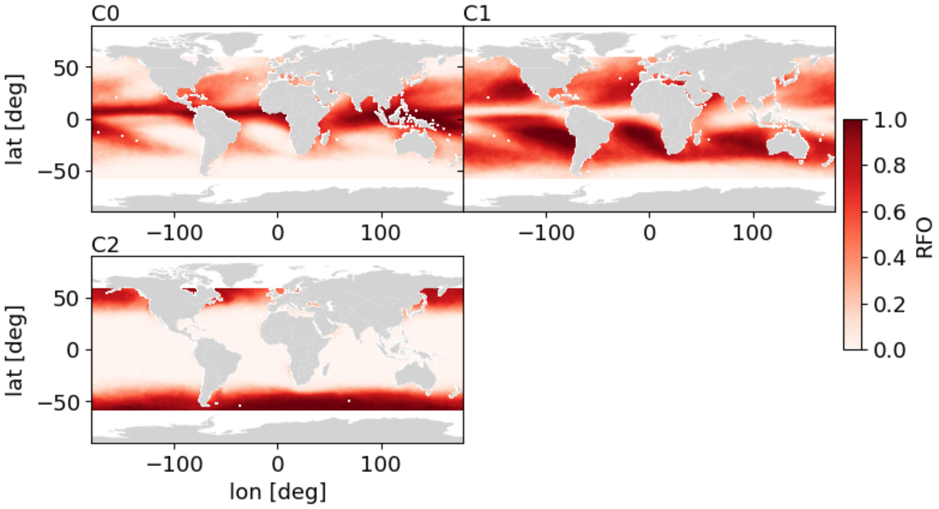


Figure B.1: Relative frequency of occurrence for k=3.

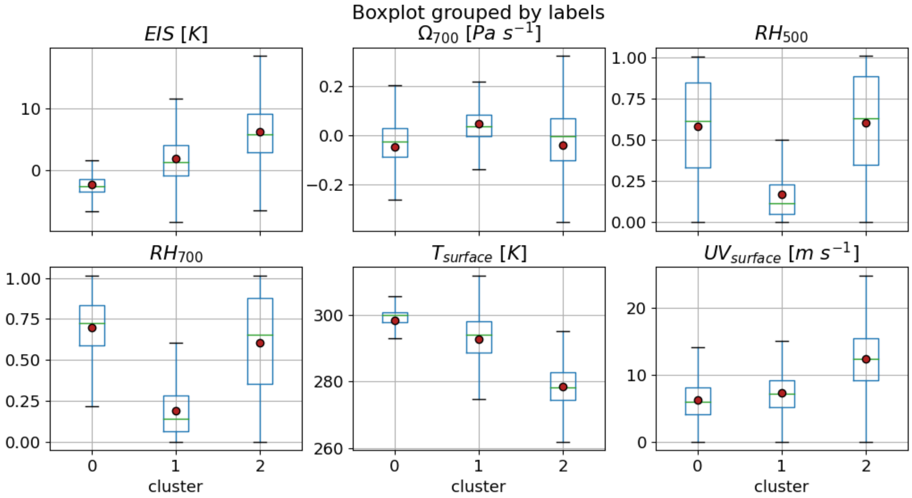


Figure B.2: Boxplot of the cloud-controlling factors for k=3.

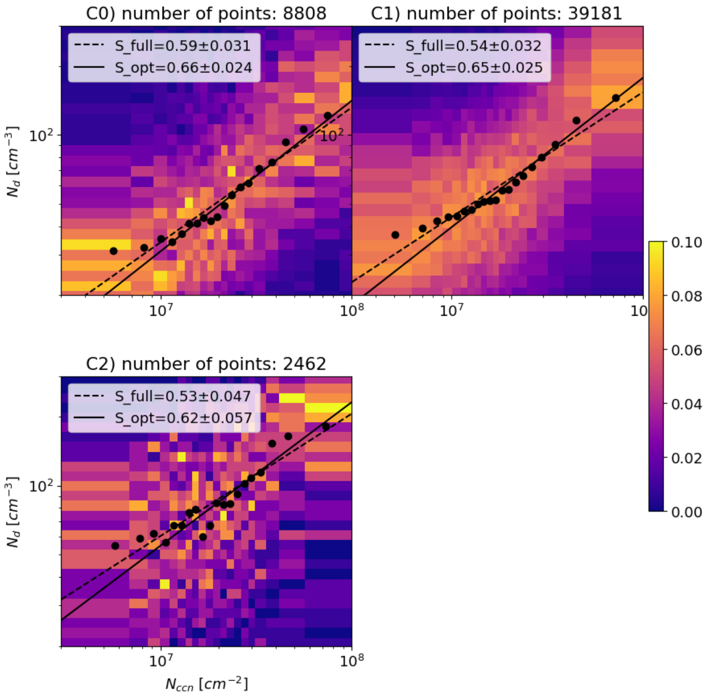


Figure B.3: Susceptibility of N_d to N_{ccn} for k=3.

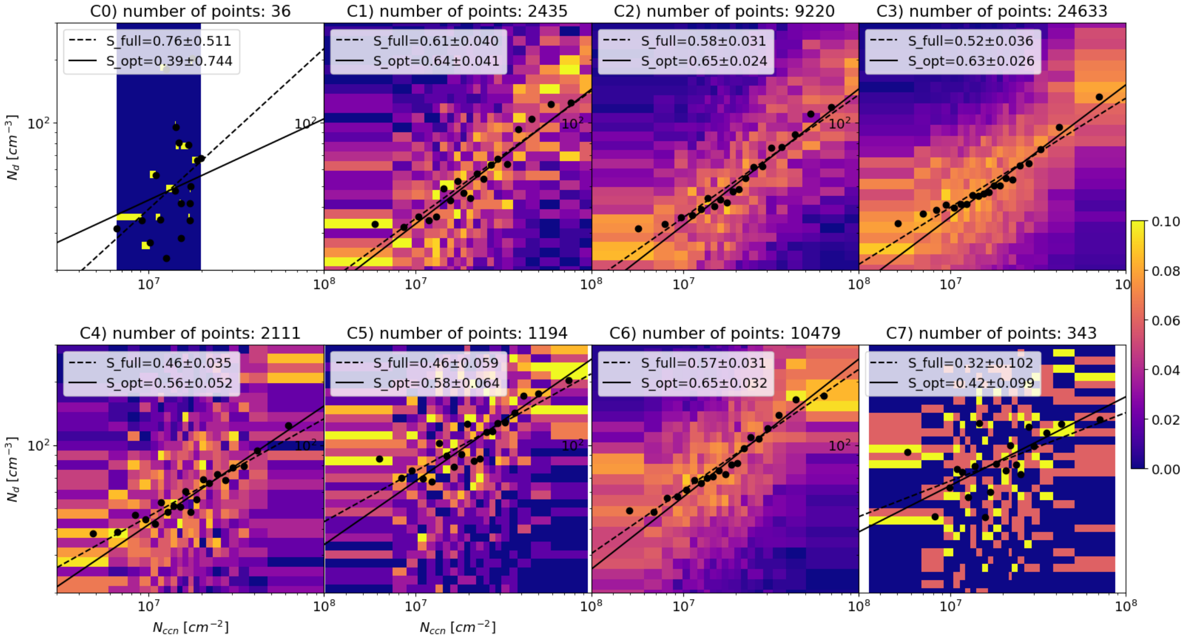


Figure B.4: Susceptibility of N_d to N_{ccn} for k=8.

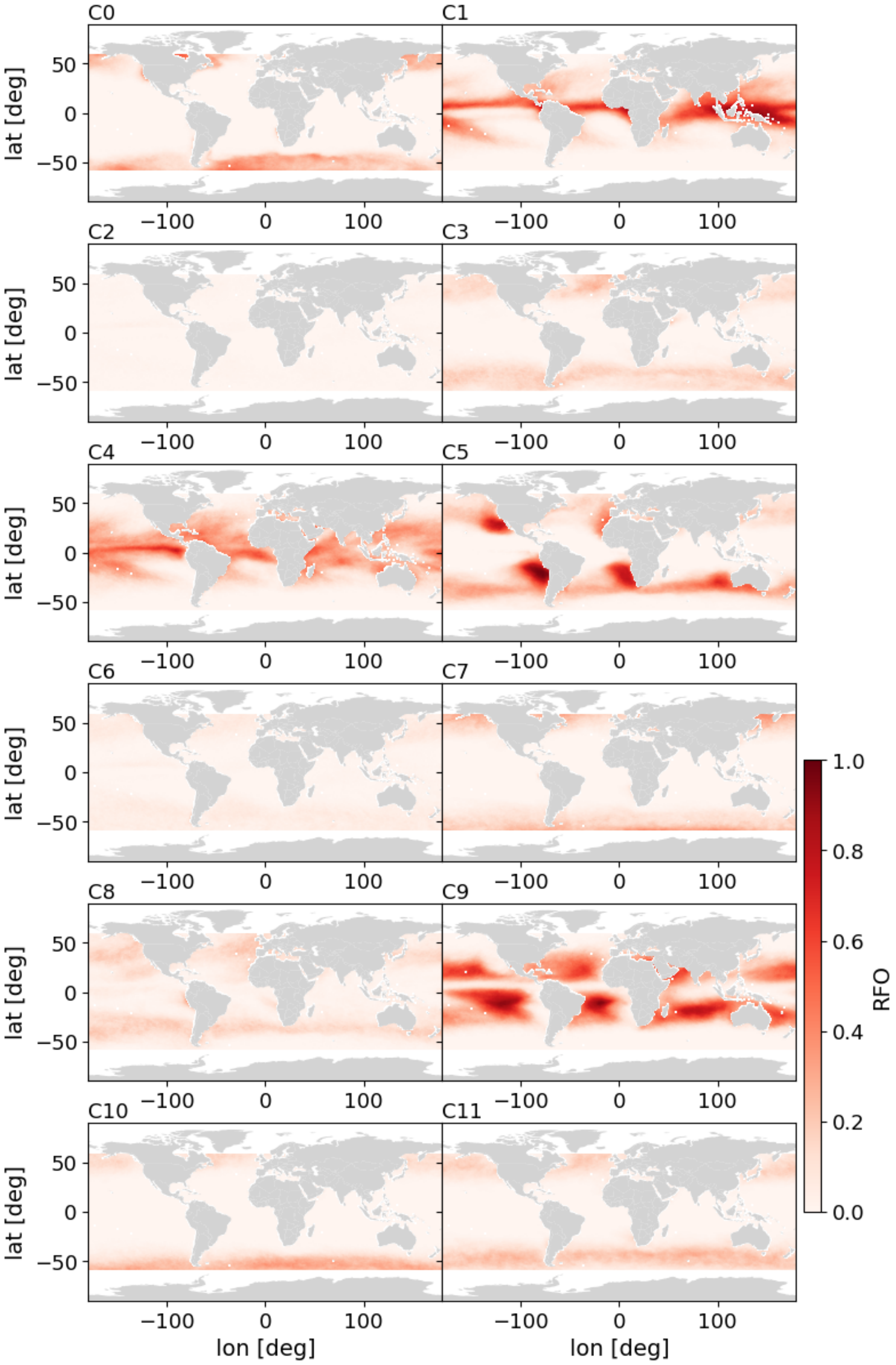


Figure B.5: Relative frequency of occurrence for k=12.

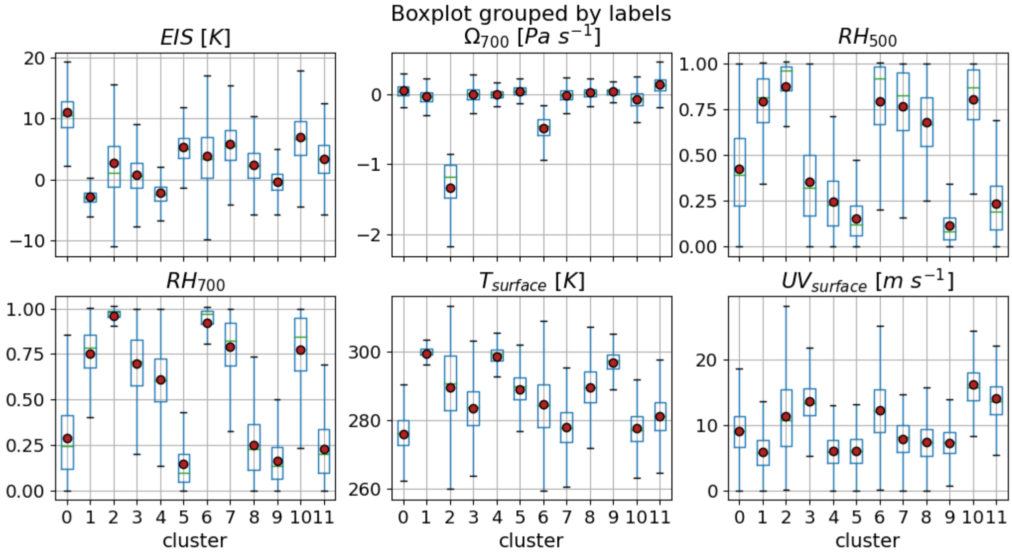


Figure B.6: Boxplot of the cloud-controlling factors for k=12.

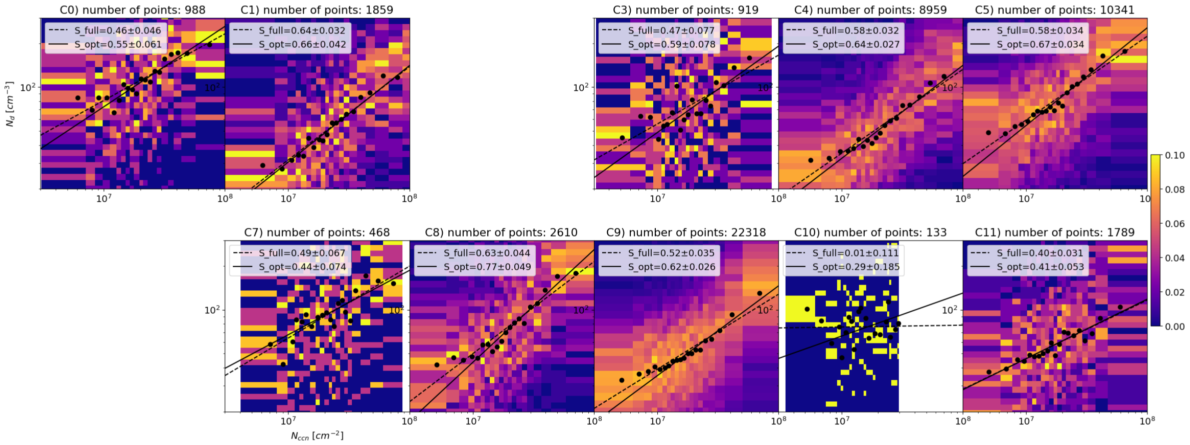


Figure B.7: Susceptibility of N_d to N_{ccn} for k=12.