TURBULENCE AND DROPLET CLUSTERING IN SHALLOW CUMULUS: THE EFFECTS OF AEROSOLS AND CLOUD HEIGHT

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

ATMOSPHERIC SCIENCES

MAY 2017

By

Dillon S. Dodson

Thesis Committee:

Jennifer D. Small Griswold, Chairperson
Christina Karamperidou
Pao-Shin Chu

ACKNOWLEDGMENTS

I would like to thank the crews of the CIRPAS Twin Otter aircraft whose hours of hard work and dedication to collecting high quality observations and data during the GoMACCS campaign made this research possible. I am deeply grateful for my adviser Jennifer Griswold who provided me the amazing opportunity to seek out my research interests under her guidance. Her willingness to help and insightful conversations constantly motivated and challenged me to think critically of my work. Furthermore, funding my travel to the 2017 AMS conference in Seattle was vital in allowing me to receive feedback from the scientific community, which significantly improved the work presented. Jennifer by no means needed to take on an extra student, but allowed me to work with her while being a TA for the Atmospheric Sciences Department here at the University of Hawaii, Manoa. And for that, I am deeply grateful. I would also like to thank my committee members, Christina Karamperidou and Pao–Shin Chu for their helpful comments and constructive criticism which challenged me to further improve my research. Both Pao–Shin and Christina accepted my constant interruptions to help answer my question, whether about statistics or basic formatting. I would also like to recognize Dr. Micheal Larsen, as he went more than out of his way to provide me with information on the pair-correlation function. Dr. Larsen also provided with me with code which was used throughout this thesis for the calculation of the pair-correlation function. He gave insightful questions that constantly challenged my understanding and opened my mind to consider all aspects of the problem at hand. Without Dr. Larsen, my thesis would not be what it is. I also need to thank my undergraduate professors at the University of Wyoming for challenging me to pursue my master's degree. Robert Kelly, Jefferson Snider, and Thomas Parish helped shape me into a better student, scientist, and person. Their dedication to helping me through the research process and applying to graduate schools brought me further than I would've ever gone on my own. I would also like to thank my family and friends whose constant belief in my abilities motivated me to keep succeeding. Time spent with them during breaks kept me sane and the happiness they brought me throughout this journey was invaluable. Finally, I would like to thank Hanna, Naufal, and Jon for their intellectual conversations and contributions of ideas and motivation. Without their friendship, these past two years would have been much longer than they were.

ABSTRACT

Aerosol–cloud interactions are complex including albedo and lifetime effects that cause modifications to cloud characteristics, as well as the resulting effects on climate. With most cloud–aerosol interactions focused on the previously stated phenomena, few studies have focused explicitly on how aerosols can effect turbulence within a cloud, especially cloud edge entrainment which impacts cloud lifetime and size. Along with turbulence, little research has been conducted outside the lab on droplet clustering within clouds and its relationships with turbulence. Cloud droplets are impacted by turbulence through droplets being preferentially concentrated into regions of lower vorticity, leaving regions of higher vorticity relatively free of droplets. This research aims to gain a better understanding on how turbulence can be used to map droplet clustering within a cumulus cloud and how aerosol number concentration affects droplet clustering. Not only is turbulence a tool for analyzing droplet clustering, but it can also have implications related to precipitation formation through collision–coalescence.

In this work, aerosol–cloud relationships are derived from warm continental cumuli subjected to various levels of anthropogenic influence sampled during the 2006 Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS) by the Center for interdisciplinary Remotely–Piloted Aircraft Studies (CIRPAS) Twin Otter aircraft. Drop size distributions, cloud liquid water content (LWC), and a time stamp (allowing for droplet calculations down to 10^{-4} m scale) for when each cloud droplet was encountered were measured using the Artium Flight Phase–Doppler interferometry (PDI). These data along with other meteorological observations are used to investigate turbulence and droplet spacing within the cumuli. The pair–correlation function (PCF) is used to identify the scale of preferential concentration, with more clustering signifying a more turbulent environment and vice versa.

Results using four complete days of data with 81 non–precipitating cloud penetrations (minimum 300 m in length) organized into two flights of low (L1, L2) and high (H1, H2) pollution data show a more turbulent environment near cloud edge as compared to the cloud center for all four cases, with low polluted clouds showing more droplet clustering for both cloud edge and center. Although low polluted clouds have more overall clustering, high polluted clouds have more statistically significant clustering than low polluted clouds. Further analysis shows that the higher clustering experienced in low polluted clouds is due to L2, which has a significantly higher amount of clustering than the other three flights. It is believed that most of the clouds within the L2 flight are decaying, making cloud lifetime more significant to clustering than the aerosol number concentration, although more research

must be done. Using a single cloud in which multiple flight passes were made at different altitudes (ranging from 806 to 3381 m) shows that clustering increases particularly at cloud top where cloud top entrainment is occurring. Negligible change in the PCF as a function of altitude is seen before the cloud top entrainment zone. This work helps lay the foundation for being able to model the affects that droplet clustering has on collision–coalescence and rain formation time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background

Clouds are an essential part of Earth's environmental system, playing a role in the hydrological cycle and other physical processes. They are responsible for transporting fresh water through precipitation, the amount and frequency of which controls the type of climate that is present. On larger scales, clouds are involved in the radiation budget, responsible for transporting energy poleward and equatorward due to latent heat release associated with different phase changes of water. Clouds also have a large albedo, reflecting between 60 and 90 percent of shortwave radiation, depending on cloud type and height (Ahrens 2008), while also absorbing longwave radiation that is emitted from Earth's surface. On average, about 67 percent of our planet is covered by clouds (Rossow and Schiffer 1999). The amount of cloud cover changes depending on latitudes and location (land vs. ocean). On average, about 72 percent of the ocean is covered by clouds while 59 percent of land surfaces are covered (Rossow and Schiffer 1999). These clouds form due to convection, large–scale vertical air motions associated with frontal systems, and forced convection through orographical lifting. Depending on updraft intensity, the abundance and intensity of precipitation can vary. It is clear that clouds play a major role in controlling climate and determining day–to–day weather.

Clouds experience different temporal and spatial scales. Satellite views of Earth reveal organized cloud patterns extending over hundreds to thousands of kilometers. They move with the wind field or stay in place, some existing for hours, while others exist for days at a time (Rogers and Yau 1989). A cloud field can extend over several kilometers or more, but the processes that lead to rain formation take place on scales of the cloud droplets themselves, the micrometer scale. Thus cloud processes cover spatial scales that span nine orders of magnitude. Within the clouds themselves the motions of individual cloud droplets can have an influence on the overall cloud size and lifetime. This makes the implementation of clouds in general circulation models (GCMs) a difficult task because of the various types of clouds that exist and the millions of cloud droplets that exist within a single cloud that interact with each other and the environment around them. In order to better understand clouds and their global impact, we need to understand cloud processes at all scales. All phenomena such

as cloud microphysics (smallest spatial scale) to entrainment mixing (larger spatial scale) occurring in clouds are equally important to cloud lifetime and precipitation development.

There are two types of precipitating clouds, defined by the presence or lack thereof of ice crystals. In the mid–latitudes, rain is often produced in the presence of ice via the Wegener–Bergeron–Findeisen process. Clouds often extend to altitudes with temperatures below 0°C where ice particles form through either freezing of droplets or sublimation of water vapor on ice nuclei (Vali 1996). The second type, mostly seen in the trade wind regime and the tropics are warm clouds. Warm clouds are an important part of the global water cycle and significantly modify the atmospheric radiation budget. Condensation and collision–coalescence are the main processes in which droplets grow in warm clouds. Supercooled clouds that contain only droplets and no ice particles can be found at temperatures below the freezing point. The same process as in warm clouds still applies as long as no ice crystals are present. The annual mean cloud cover of cumulus clouds is about 5 percent over land and 12 percent over the ocean (Warren and Hahn 2002), most of which are warm. The trade wind regions are particularly characterized by shallow maritime cumulus convection. Thirty-one percent of the total rain and 72 percent of the total rain area in the tropics are associated with warm clouds (Lau and Wu 2003). Although the processes that lead to rain formation, whether from a cold or warm cloud, are well known, one of the main challenges of cloud physics is the fast formation of rain in warm clouds. Laird et al. (2000) and Szumowski et al. (1997) made radar measurements in Florida and offshore the island of Hawaii, respectively, indicating that warm shallow cumulus clouds can produce rain in 15 to 20 minutes. It has been shown by Jonas (1996) that a supersaturation of 0.2 percent with cloud droplets of 10 μm radii will need about 20 minutes to grow to a size of 20 μm through condensation. To further grow to drizzle size ($\sim$100 μm) in a cloud with a Liquid Water Content (LWC) of 1 gm^{-3} requires another hour, assuming that the droplet collection efficiency is one. Clearly these calculations are in disagreement with the aforementioned observations.

In order to explain the fast formation of rain, it has become more and more accepted that the turbulent nature of the wind velocity field and the fields of water vapor, temperature and liquid water are an important aspect of cloud formation and evolution (Shaw 2003). Thinking about turbulence and mixing, it is a common observation, for instance when stirring cream in a cup of coffee, that turbulence produced by the stirring action would cause the initially inhomogeneous field of cream to become homogenously mixed. Such mixing ability is, in fact, a hallmark of turbulent flows. In this mixing process, the randomly rotating and stretching fluid motion in turbulence causes the initially simple geometry of the cream field to be

distorted and stretched, thus increasing the surface area bordering the creamed and clear fluid. When the bordering area grows to become large enough, the rate of molecular diffusion across it becomes dominant and the mixture becomes homogenous down to sub–micrometer scales (Villermaux et al. 2001). Similar mixing processes can be seen throughout nature, such as in the mixing of smokestack plumes, pollutants in turbulent rivers, and so on.

Clouds are well known as areas of enhanced turbulence, with the intensity of the turbulence varying widely depending on the type of cloud and its age (Pinsky and Khain 1996). Different turbulent dissipation energies have been observed for different types of clouds. Focusing on small cumuli, the turbulent energy dissipation rate is about 100 cm^2s^{-3} (Mazin et al. 1984). One might then expect that the same situation described in the previous paragraph occurs in turbulent atmospheric clouds. Mixing does occur in clouds, however, it does not follow the examples explained in the previous paragraph. Instead, the mixing in clouds has been shown to be non–homogeneous at small, less than a meter scales (Shaw 2003).

The key to understanding this clustering phenomenon is the inertia of the particles. Unlike the cream discussed earlier, which is made of particles with sizes comparable to molecular scales, water droplets have finite size and have a much higher mass–density than the surrounding fluid. Thus while the former is passively advected by the fluid, and its motion at any time mimics that of the local fluid motion, the latter may de–correlate from the advecting fluid motion due to its inertia alone (Villermaux et al. 2001). According to Shaw (2003), it is generally accepted that this inertial de–correlation leads to the clustering of inertial particles in turbulent flow at fine scales (comparable to the smallest structures of turbulence motion). This phenomenon, referred to as inertial clustering or preferential concentration (Shaw 2003) can lead to increased collision rates of water droplets in clouds, which may help explain the anomalous short time of rain formation discussed earlier (Pinsky and Khain 1996; Falkovich et al. 2002; Shaw 2003). For example, the stochastic collection equation, which describes the growth of droplets via collision–coalescence, assumes that the cloud droplets are spatially distributed in a perfectly random fashion, and not preferentially concentrated (Kostinski and Shaw 2001). It was also shown that, according to observations, the rate of spectrum broadening, or the broadening of the cloud droplet size range, is greatest in the areas of cloud with high intensity turbulence (Pinsky and Khain 1996; Politovich 1993).

Cloud responses to changes in aerosol remain a major uncertainty in the radiative forcing of climate. There are two main theories related to aerosol effects on clouds. The first being the albedo effect, where anthropogenic aerosol results in increased droplet concentrations that generate increases in cloud albedo. The second is the lifetime effect, whereby anthropogenic

aerosol suppresses precipitation and results in clouds with more liquid water, higher fractional cloudiness, and longer lifetimes (Small et al. 2009). These two constructs are also known as the indirect effects.

According to the Intergovernmental Panel on Climate Change (IPCC) (Ramaswamy et al. 2001), clouds remain one of the greatest sources of uncertainty in estimating climate sensitivity. It is clear therefore, that more research must be conducted to understand how aerosols can change the radiative forcings of clouds. This thesis focuses on the fundamental investigation of inertial clustering, specifically how it changes within clouds (cloud edge vs. cloud center) and different cloud environments, such as low pollution clouds and high pollution clouds. The latter of these foci is motivated by the effect that aerosols have on the formation and evolution of cumulus clouds. It is known that the effects of aerosols on clouds can lead to increased cloud lifetime due to the suppression of rain formation and can also lead to higher cloud albedo due to smaller cloud droplets, better known as the cloud lifetime and albedo effects, respectively, as discussed earlier (Small et al. 2009). This begs the question, can aerosol concentration effect droplet clustering, which could then have an influence on the lifetime and/or albedo effects just discussed? Specifically, clustering could cause larger droplets to form through enhanced collision–coalescence leading to a dampening influence of the cloud albedo effect. The former of these foci is motivated by the different turbulence values commonly seen at cloud edge vs. cloud center. The main sources of turbulent kinetic energy (see Section 1.2.2) in clouds are shear evaporative cooling due to entrainment of dry air (at cloud edge and cloud top) and heating due to condensational growth (mostly at cloud base) (Shaw 2003). With the main source of turbulence being at cloud edge, you would expect to see larger inertial clustering in entrainment zones. Stronger clustering at cloud edge could result in slower evaporation times for droplets, or even result in a larger rate of coalescence which acts to slow down the rate of evaporation of droplets at cloud edge (due to the formation of larger droplets) and therefore enhance the lifetime effect. It has been shown by Small et al. (2009) that when small, non–precipitating cumulus experience aerosol perturbations, they respond in a manner consistent with enhanced evaporation around the shell of the cumulus and a stronger horizontal buoyancy gradient is produced, known as the evaporation–entrainment feedback mechanism. This leads to stronger vorticity around the core of the cloud and stronger entrainment. Perhaps, the stronger vorticity produced in this example can lead to a decrease in the enhanced evaporation due to increased droplet clustering leading to heightened coalescence, acting as a negative feedback. See Section 2.2 for more information on scientific motivation.

Data was collected during the 2006 Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS) by the Center for interdisciplinary Remotely–Piloted Aircraft Studies (CIRPAS) Twin Otter aircraft. Drop size distributions, cloud LWC, and a time stamp for when each cloud droplet was encountered were measured using the Artium Flight phase-Doppler interferometer (PDI). These data along with other meteorological observations are used to investigate the inertial clustering of droplets within 81 non–precipitating shallow cumulus clouds that are a minimum of 300 meters in length, organized into two flights of low pollution (L1, L2) and two flights of high pollution (H1, H2) data. First, the basic theoretical concepts utilized within this work are introduced starting in Section 1.2, including turbulence basics, droplet clustering, and tools used to analyze droplet clustering. Chapter 2 will cover cloud microphysics basics (entrainment mechanisms and types), scientific questions regarding the benefits this research can have on cloud microphysics, pair–correlation function (PCF) characteristics, and the methods and instrumentation used. Chapter 3 will cover the results, including cloud properties from different flight days. Results will cover questions including: Does droplet clustering depend on the aerosol number concentration? Does clustering change as a function of height? Is there a difference in clustering between cloud center and cloud edge? Hypotheses to these questions can be found in Section 2.7. The results are summarized and an outlook is given in Chapter 4.

1.2 Turbulence Theory

Mathematical and statistical tools utilized in this work will be looked at. First, the fundamental concepts of classic turbulence theory which are critical to the understanding of this work will be analyzed. This section is not a guide through the complete turbulence theory, rather it is an extract of what is considered relevant to this work. For further details on turbulence theory, see (e.g., Frisch (1995); Tennekes and Lumley (1972)). Fluid turbulence, turbulent kinetic energy, the energy cascade, and the Kolmogorov microscale will give the reader a basic understanding of turbulence, which leads into inertial particle clustering in turbulent flow and the tools which can be used to measure said clustering.

1.2.1 Fluid Turbulence

This thesis concerns itself with the type of fluid motion that could accurately model many fluid dynamical phenomena in our atmosphere, oceans, and rivers. In practical terms, this is the motion of an incompressible Newtonian fluid and its equation of motion is the

Navier–Stokes equation coupled by the (incompressible) continuity equation. The continuity equation can be shown from Pruppacher and Klett (1980) by:

$$\frac{1}{\rho}\frac{d\rho}{dt} = -\nabla \cdot u \tag{1.1}$$

where ρ and u are the fluid density and velocity, respectively. The above equation represents the continuity equation in differential form (Granger 1995). It is well known that flow past an object can be considered incompressible ($\frac{d\rho}{dt} = 0$) whenever the following four conditions are met according to Pruppacher and Klett (1980):

- The characteristic flow speed U satisfies $U \ll c$, where c is the speed of sound in the fluid ($c \sim 340$ msec^{-1} in air at 15°C and 1 atm);
- The dominant flow oscillation frequency f satisfies $f \ll c/L$, where L is the order of the size of the falling particle;
- $L \ll \frac{g}{c^2}$, where g is the magnitude of the gravitational acceleration (being equivalent to the condition that the static pressure difference between two points separated in the vertical by length L must be very small compared to the absolute pressure (L is a small fraction of the atmospheric scale height, $H \sim 10$ km);
- The fractional temperature difference between obstacle and stream is small;

As these inequalities hold for all cloud particle motions, we shall henceforth assume the flows under consideration are incompressible, so that Equation 1.1 can be re–written, leading to the governing equations for a three–dimensional flow of an incompressible fluid, where the fluid can be either a gas or a liquid, given by the Naiver–Stokes equations. The Navier–Stokes equations for air are given by:

$$\begin{gathered} \frac{\partial u}{\partial t} + u \cdot \nabla u = -\nabla P + \frac{1}{Re}\nabla^2 u \\ \nabla \cdot u = 0 \end{gathered} \tag{1.2}$$

(Pruppacher and Klett 1980; Chorin 1968; Granger 1995), where the Navier–Stokes equation is shown in its non–dimensional form with $u(x,t)$ as the fluid velocity field and $P(t)$ as the fluid pressure field. Re is the dimensionless Reynolds number defined from the characteristic scales of the problem at hand. Meaning that the flow becomes turbulent when the Reynolds

number of the flow, given by McIlveen (1986) and Granger (1995),

$$Re = \frac{UL}{\nu} \tag{1.3}$$

is larger than a critical value that depends on the problem itself. Here U is the characteristic velocity scale, L is the characteristic length scale, and ν is the fluids kinematic viscosity, which can be defined as:

$$\nu = \frac{\mu}{\rho} \tag{1.4}$$

where ρ is the fluid density and μ is the dynamic viscosity. The kinematic viscosity for air is roughly $1.5 \cdot 10^5$ ms^{-2}. The characteristic length scale L is the largest possible scale for a turbulent eddy in the flow, and is also referred to as the outer or integral scale. The Reynolds number, which can be interpreted as the ratio of the flows inertial energy to the fluid viscous dissipation (Stull 1988) is an important quantity that characterizes the nature of the fluid motion. Typical Reynolds numbers in the atmosphere are on the order of 10^6 to 10^7 (Shaw 2003). Flows with the same Reynolds number behave in a statistically similar manner, whereas when the Reynolds number is very high, the fluid motion becomes turbulent and has the general characteristic of being chaotic and solenoidal (having many vortices). If we consider a problem like mixing, Re can also be interpreted as the ratio of the time scale for molecular diffusion to the time scale for turbulent mixing (Tennekes and Lumley 1972).

1.2.2 Turbulent Kinetic Energy

Turbulent kinetic energy, hereafter known as TKE, is one of the most important variables in turbulence/micrometeorology because it is a measurement of the intensity of turbulence. TKE is directly related to the momentum, heat, and moisture transport through the turbulent medium and is also used as a starting point for approximations of turbulent diffusion. As shown in Stull (1988), TKE can be written as:

$$\frac{\partial \bar{e}}{\partial t} + U_j \frac{\partial \bar{e}}{\partial x_j} = \delta_{i3} \frac{g}{\theta_v} (\overline{u_i' \theta_v'}) - \overline{u_i' u_j'} \frac{\partial \bar{U}_i}{\partial x_j} - \frac{\partial (\overline{u_j' e})}{\partial x_j} - \frac{1}{\rho} \frac{\partial (\overline{u_i' p'})}{\partial x_i} - \varepsilon \tag{1.5}$$

where $\overline{e} = \frac{1}{2}(\overline{u'^2} + \bar{v'^2} + \overline{w'^2})$, which is equal to (using summation notation) $\frac{1}{2}\overline{u_i'^2}$. Here the first term on the left represents local storage or the tendency of TKE, and the second term describes the advection of TKE by the mean wind. Looking at the right side of Equation 1.5, it is recognized that the rate of change of TKE is due to (looking at terms on the right

hand side of the equation, going from left to right):

- The buoyant production or consumption, depending on whether the heat flux is positive (during the day over land) or negative (at night over land), where θ is the potential temperature;
- Mechanical or shear production/loss. Also known as deformation;
- The turbulent transport of TKE, describing how TKE is moved around by the turbulent eddies;
- Pressure correlation term that describes how TKE is redistributed by pressure perturbations. It is often associated with oscillations in the air such as buoyancy or gravity waves;
- The viscous dissipation of TKE; i.e., the conversion of TKE into heat, also known as the viscous dissipation or energy dissipation rate (see Section 1.2.3 for more information).

The shear production/loss term ($\overline{u_i' u_j'} \frac{\partial \overline{U_i}}{\partial x_j}$) is usually negative. The momentum flux ($\overline{u_i' u_j'}$) is usually of opposite sign from the mean wind shear because the momentum of the wind is usually lost downward to the ground. Thus, the shear term results in a positive contribution to TKE when multiplied by a negative sign as it is in the equation. This term takes kinetic energy out of the mean flow and redistributes it to turbulence. This production term of energy increases with eddy size, and thus the energy taken from the mean flow transforms to turbulence mainly at large scales comparable to L. We will see in Section 1.3.4 that the shear production/loss term is important for the spatial distribution of cloud droplets.

1.2.3 Energy Cascade

Turbulence is dissipative. The viscous dissipation term (ε) can be written as:

$$\varepsilon = \frac{U^3}{L} \tag{1.6}$$

where ε is a loss term that always exists whenever TKE is nonzero. Physically, this means that turbulence will tend to decrease and disappear with time unless it can be generated locally or transported by some process (Stull 1988). The dissipation of energy is most effective at small scales, which together with the fact that energy is mainly provided at large scales, implies that energy is transferred from larger scales to smaller scales. The

governing process is the input of energy by large eddies. Following this picture, an energy cascade model for turbulence was proposed (see e.g. Tennekes and Lumley (1972)), where the largest scales of motion receive kinetic energy from the forcing mechanism and this energy is cascaded to a long series of smaller scales. The cascade mechanism is thought to be due to inviscid instability in which large vertical structures (such as eddies) become unstable thus breaking into smaller eddies. At the smallest scale of turbulent motion, viscosity becomes dominant and kinetic energy is dissipated into heat. Kolmogorov (1941) states that the smallest scales of turbulence are universal and do not depend on the specifics of the large scale forcing. The memory of the large scales is lost in the random nonlinear cascade. It is then deduced that the small scale motions are fully governed by the fluid viscosity and the kinetic energy dissipation rate per unit mass of the fluid. Following this energy cascade we can talk about the role of vorticity in the origin of turbulence, where the development of an unstable shear results in the growth of two–dimensional disturbances with periodic fluctuations of vorticity. The transformation of the disturbance to three–dimensionality leads to a power spectrum (a plot of the portion of a signals power (energy per unit time) falling within given frequency bins (Welch 1967)) broadening by vortex filament interactions. In the formation of vortex filaments, the vorticity is transferred across the spectrum to smaller and smaller scales (Granger 1995). See Section 1.3.3 for more information on vortex filaments, also known as vortex "tube".

1.2.4 Kolmogorov Microscale

At small scales the flow is governed by energy dissipation and viscosity. Using both parameters, Kolmogorov derived a measure of the smallest scales of motion, a length scale l_K and a time scale τ_K for the smallest turbulent eddies, defined as:

$$l_K = \left(\frac{\nu^3}{\varepsilon}\right)^{\frac{1}{4}} \quad \text{and} \quad \tau_K = \left(\frac{\nu}{\varepsilon}\right)^{\frac{1}{2}} \tag{1.7}$$

The length scale l_K is called the Kolmogorov microscale or "inner scale", and τ_K is the Kolmogorov time–scale (Shaw et al. 1998; Stull 1988; Marshak et al. 2004). To provide some sense for the numerical values of these scales, under typical conditions found in cumulus clouds the average energy dissipation rate varies over the range of 0.001 to 0.1 $\mathrm{m^2 s^{-3}}$. The corresponding range of Kolmogorov length and time scales are 0.4–1.25 mm and 0.01–0.1 seconds, respectively. TKE dissipation rates are on the order of 0.01 $\mathrm{m^2 s^{-3}}$ in the adiabatic cores of cumulus clouds (Shaw et al. 1998; Meischner et al. 2001). This scaling assumes that

the smallest eddies see only turbulent energy cascading down the spectrum at rate ε, and feel only the viscous damping of ν. When studying such small scales of turbulence, the Reynolds number based on the Taylor microscale can be used. The Taylor microscale is given by:

$$l_T = U\sqrt{\frac{15\nu}{\varepsilon}} \tag{1.8}$$

Using l_T and U, the Taylor Reynolds number is defined by:

$$Re_T = U^2\sqrt{\frac{15}{\varepsilon\nu}} \tag{1.9}$$

Re_T can be interpreted as the ratio of the large eddy time scale to the time scale of the strain rate fluctuations. When the Reynolds number is high, the Taylor Reynolds number is related to the Reynolds number by:

$$Re = 0.15{Re_T}^2 \tag{1.10}$$

as derived in Pope (2000).

The nature of the atmospheric turbulence spectrum is directly related to the fact that production and dissipation are not happening at the same scales. Production is feeding only the larger size eddies, but dissipation is acting only on the smaller sizes. Thus, the rate of transport across the middle part of the turbulence spectrum is equal to the rate of dissipation, ε, at the small–eddy end. Such transfer can be thought of as happening by inertially–larger eddies creating or bumping into smaller ones, and transferring some of their inertia in the process. This middle portion of the spectrum is called the inertial subrange.

1.3 Inertial Particle Clustering in Turbulent Flow

1.3.1 Poisson Process

Poisson's probability formula is useful when specific events occur infrequently (Langley 1971). Imagine a few blades of grass in a square mile of desert. The proportion of grass to bare ground is so small that it is virtually impossible to express it as a fraction; in such a case it is more convenient and accurate to work in terms of the incidence of the feature, so one can count the blades of grass per square mile. Other examples include the number of flaws in the insulation of a length of electric cable or the number of misprints per page of

a book. The situation can also be considered as a function of time, such as the arrival of people at a telephone booth, the frequency of hurricane formation, or the arrival of droplets through instrumentation. All such cases are described as being isolated occurrences in time or space. The number of occurrences can be stated, but not the number of non–occurrences. The Poisson distribution describes the numbers of discrete events occurring in a series, or a sequence, and so pertains to count data that can take on only non–negative integer values. Poisson events occur randomly, but at a constant average rate. That is, the average rate at which Poisson events are generated is stationary (Wilks 2011). The fact that the data must be stationary plays an important role in determining the significance of droplet clustering, which will be discussed in Section 1.4. The individual events must be independent, in the sense that their occurrences do not depend on how many other events may have occurred elsewhere in nonoverlapping portions of the sequence.

For the case of cloud droplets, the spatial distribution of droplets at scales as small as the droplet diameter d_p are of interest, because it is at these scales where droplet collisions occur, leading to larger scale cloud processes such as precipitation. If the influence of turbulent motion and large scale processes related to turbulence, such as entrainment are neglected, it is reasonable to assume that droplets are randomly distributed in space. The following conditions define a Poisson process, characterized by the Poisson distribution on any scale. In other words, the spatial distribution of droplets is called random if the following assumptions apply (Kostinski and Shaw 2001):

- The process is statistically homogeneous;
- The probability of detecting more than one particle in a given volume dV is vanishingly small for sufficiently small dV, in other words, in an infinitesimally small volume either zero or one droplet is found;
- Particle counts in non–overlapping volumes are statistically independent random variables at any length scale.

The last assumption is the true origin of the Poisson point process. Given these assumptions, the probability distribution of particle counts is given by:

$$p(N) = \frac{\bar{N}^N exp(-\bar{N})}{N!} \tag{1.11}$$

from (Kostinski and Shaw 2001; Shaw 2003; Shaw et al. 2002). For a Poisson process, *p(N)* is the probability of finding N particles in a test volume while the mean value ($\bar{N}$) is equal

to the variance of counts $\overline{(N-\bar{N})^2} = \overline{(\delta N)^2}$ (Kostinski and Jameson 2000). Note that $\overline{(\delta N)^2}$ can hold only for unitless, integer–valued random variables such as counts (Kostinski and Shaw 2001). It is important to note that the validity of the Poisson distribution on some spatial scale L does not imply the Poisson process. For example, positive and negative deviations from the Poisson process on smaller spatial scales can cancel each other and result in Poisson statistics on longer scales (as was shown in Kostinski and Shaw (2001)). The Poisson process, or perfect randomness, implies the absence of correlations on all scales.

1.3.2 Inertial Clustering Basics

Evidence of particle clustering is already abundant (see e.g. Maxey (1987); Eaton and Fessler (1994); Shaw et al. (1998, 2002); Shaw (2003)) and will be examined using simple figures and techniques. Figure 1.1 demonstrates, in the simplest form, that inertial clustering exists. Each circle represents a cloud droplet, and each line with the squares represents the distance traversed for each sampling volume, with each square corresponding to a count reaching 10 droplets. Each representation of real data from a randomly selected cloud and Poisson point data contains 280 drops over a 15 meter distance. From visually examining Figure 1.1 it can be seen that the real data is preferentially concentrated or patchy, whereas the Poissonian point data is perfectly homogeneous. The range between droplets for the real data is 0.001 to 0.29 m while the range between droplets for the Poisson process data is 0.055 to 0.055 m (each droplet is equally spaced). Again, looking at the range for the distance between every 10th droplets, the range for the real data is 0.25 to 0.76 m and for the Poisson process data it is 0.55 to 0.55 m. Visually and mathematically, there is clear clustering in the real data with the Poissonian point data being as expected, perfectly homogeneous. See Kostinski and Jameson (2000) Appendix B for an identical comparison. From Figure 1.1 the assumption can be made that droplets are non–homogeneously spaced on the order of less than a meter.

Another easy way to detect inertial clustering on small (less than 1 meter) temporal scales is to use one of the assumptions from Kostinski and Shaw (2001) that the probability of detecting more than one particle in a given volume dV is vanishingly small for sufficiently small dV. This statement is true for homogeneously distributed droplets. Figure 1.2 (a) shows a time series from a randomly selected cloud from the GoMACCS campaign that shows the number of drops per every half meter over a span of 1000 m. The green line represents a cubic smoothing spline of the supplied data. Data between the two red lines represents data that is reasonably stationary with a standard deviation of 0.97, as compared to the standard

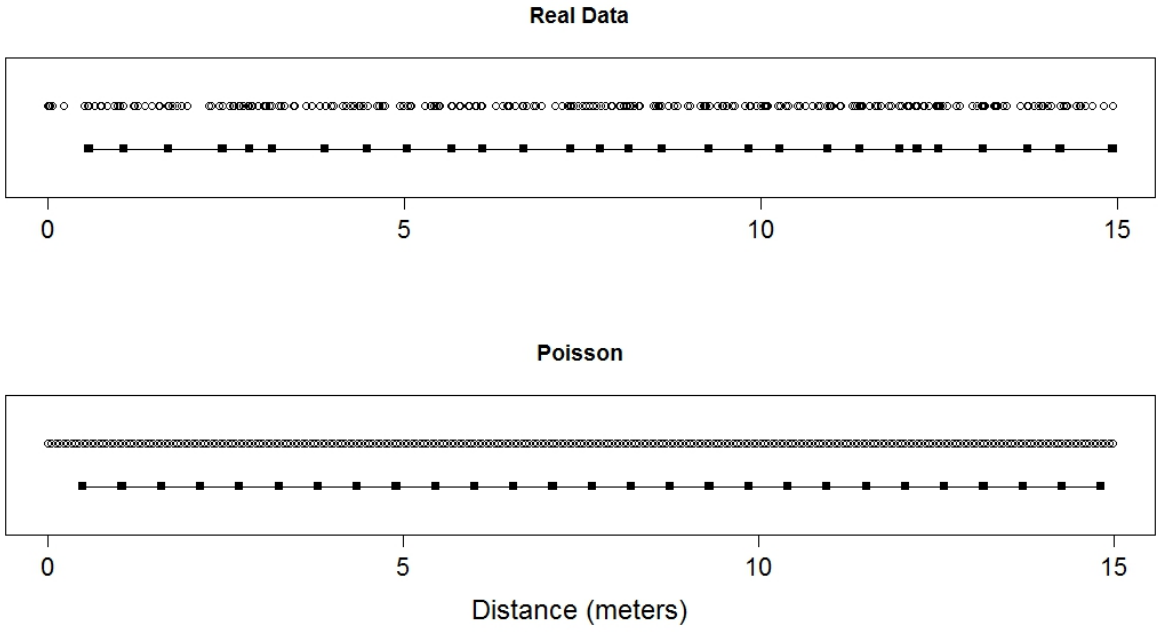

Figure 1.1: Illustration of two clouds: the top plot showing real data from a random cloud during the GoMACCS campaign, and the second plot showing a homogeneous Poissonian cloud. Each circle represents a drop while each line with a square represents the distance traversed for each sampling volume, with each square corresponding to a count of 10 droplets.

deviation of the entire time series which is 3.37, and is used in (b) of Figure 1.2. For (b), beginning with the first droplet in the selected region of the time series, the distances to all the subsequent 3172 droplets were measured. The total 175 m path is then divided into 5 cm increments and each droplet is assigned to one of these bins, which transforms the data into a time series of the number of droplets per 5 cm path increment (corresponding to an infinitesimally small length where either zero or one droplet should be found for a Poisson process; the 1–D equivalent of the 3–D sampling volume dV). This data is compared to the Poissonian simulated point data in (c) which has the same amount of droplets covering the same distance and is divided into the same sampling length as (b). As can be seen in (c), the Poisson point data when divided into a sampling length of 5 cm only has droplet counts of zero or one, which obeys the assumption stated previously. Looking at the real data in (b), when divided into sampling lengths of 5 cm the droplets range from zero to five for each length bin. For (b), 76.3 percent of the bins contain zero or one droplet, with the other 24.7

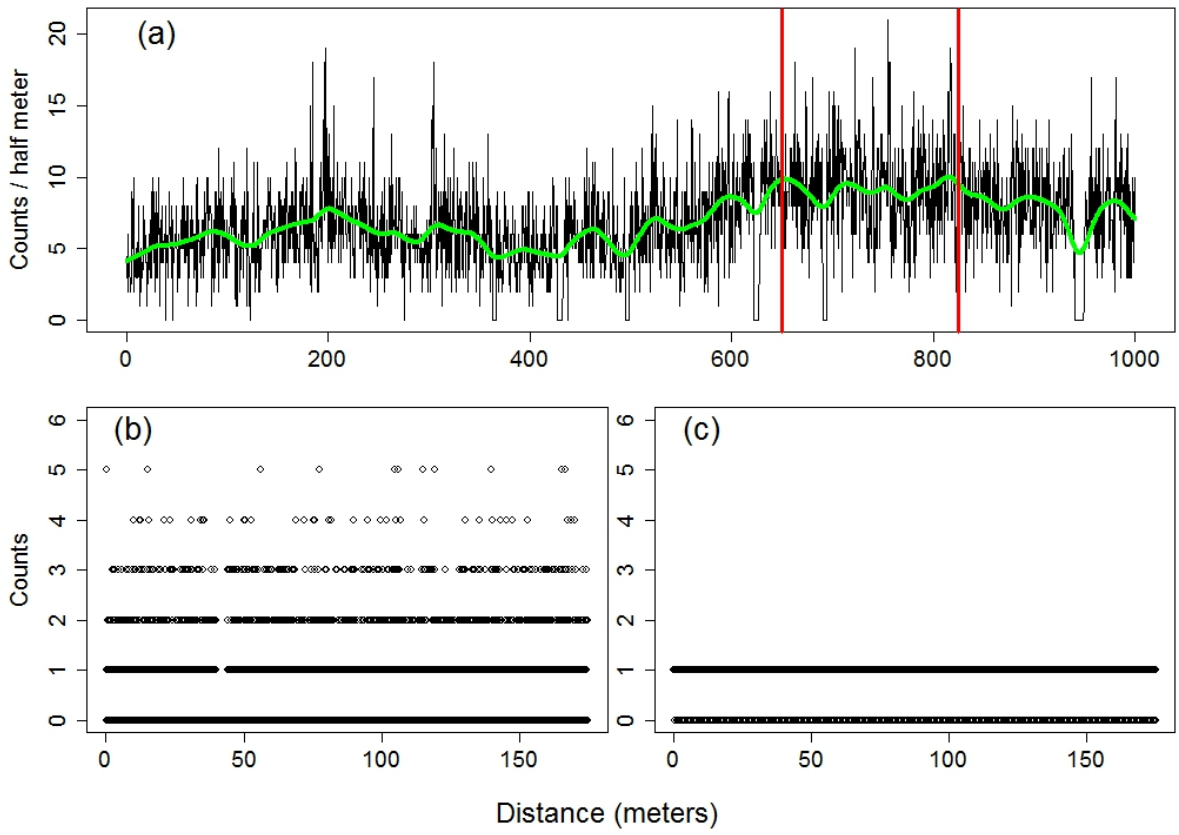

Figure 1.2: (a) A time series from a random cloud during the GoMACCS campaign, with a cubic smoothing spline in green and the data between the red lines indicating the data used in (b). (b) and (c) represent real data, and Poisson data, respectively. Beginning with the first droplet in each series, the distances to the subsequent droplets were measured. The total path was then divided into 5 cm increments and each droplet was assigned to one of these bins.

percent of the bins containing more than one droplet, which defies the assumption stated previously. This shows that real cloud data, again, is not perfectly homogeneous. Further descriptions of this process and similar graphing techniques can be found in Kostinski and Jameson (2000).

Both Figures 1.1 and 1.2 show inertial clustering, with Figure 1.1 showing clustering in 1–D using simple droplet counts. Figure 1.2 shows clustering by binning the count data into sampling lengths of 5 cm, a different technique, but again using data that is 1–D. All data from the GoMACCS campaign is represented in 1–D due to the measuring technique used (see Section 2.5 for instrumentation and measurement details). To get a slightly better

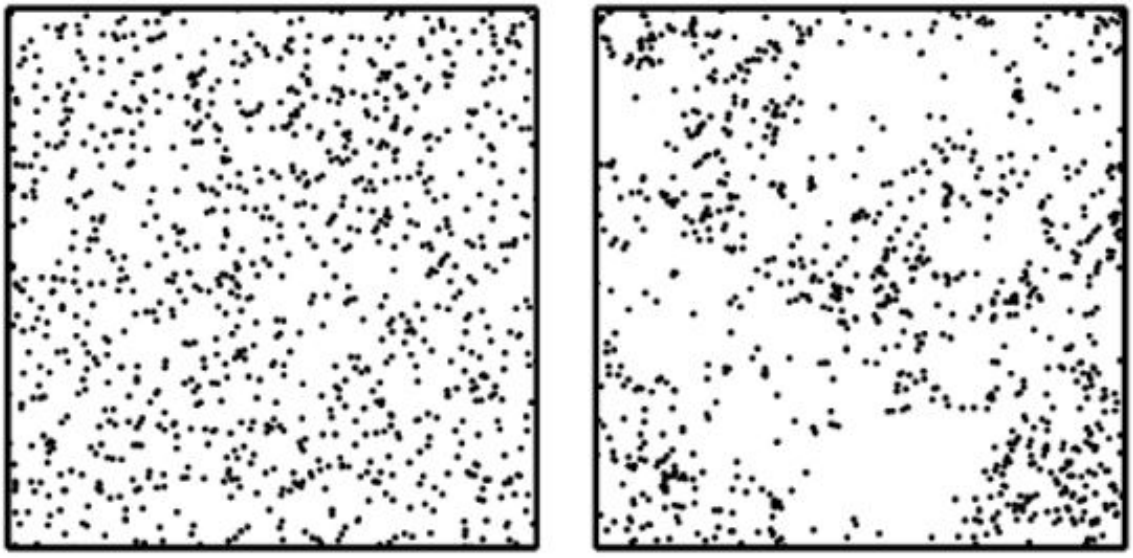

Figure 1.3: Two examples of 1000 particles with spatial positions in two dimensions chosen randomly. The left panel shows a statistically homogeneous Poisson process. The right panel shows a statistically homogeneous but spatially correlated (not Poisson) random process (Figure 5 in Shaw (2003)).

understanding of inertial clustering however, Figure 1.3 from Shaw (2003) shows inertial clustering in 2–D, with each panel having 1000 droplets. The left panel represents perfect randomness for a collection of droplets. It is referred to as perfect because particle positions are uniformly distributed and statistically independent, plus the process is statistically homogeneous (Shaw et al. 2002). A statistically homogeneous but spatially correlated random process is illustrated in the right panel of Figure 1.3. This is not a Poisson random process because particle positions are no longer statistically independent owing to the presence of clusters and voids.

1.3.3 Inertial Clustering Theories

It has been proposed that inertial clustering of particles in turbulence is the result of particles being centrifuged out of regions of high fluid vorticity (highly rotating) as a result of their inertia and thus preferentially concentrating into regions of low fluid vorticity or high strain (Shaw et al. 1998). Sundaram and Collins (1997) have consistently shown that the most responsible scale for preferential concentration is the Kolmogorov scale. This is partially supported by the fact that vorticity plays a key role in concentrating the particles,

and vorticity is predominantly concentrated in the smallest eddies (Tennekes and Lumley 1972). Dissipative structures that appear within turbulence can be referred to as vortex "tubes" that infuse isotropic turbulence, where turbulence is said to be isotropic if rotation and buoyancy are not important and can be neglected, and there is no mean flow (Granger 1995). Vortex tubes are relevant to the present discussion because they are thought to be responsible for ejecting particles into high strain regions of the flow (Shaw et al. 1998). It should be noted that, also according to Shaw et al. (1998) that there is still no comprehensive theory that fully describes the vortex tubes, but what is known is the vortex tubes are thin filaments with diameter/length ratios much smaller than unity. They are stabilized against viscous dissipation by large–scale straining motions that stretch the filaments along their axes, producing vorticity. Evidence of inertial particles preferentially concentrating in regions of low vorticity and high strain is abundant (see e.g. Maxey (1987); Eaton and Fessler (1994)), how this occurs and its relation to vortex tubes is still being researched however.

1.3.4 Governing Equations for Droplet Motion

The motion of a small droplet in a turbulent flow field is governed by an equation with several complex terms, however, under the conditions relevant for a cloud (i.e., $= \frac{\rho_d}{\rho} \gg 1$, where ρ_d and ρ are the density of the droplet and air, respectively) the equation can be simplified to:

$$\frac{dv_i}{dt} = \frac{(u_i - v_i)}{\tau_d} + g_i \tag{1.12}$$

from Siebert et al. (2010), where v_i is droplet velocity, u_i is the fluid velocity, g_i represents the gravitational acceleration vector, with i representing the i^{th} component of each term, τ_d is the droplet response time, and $\frac{dv_i}{dt}$ is the Lagrangian droplet acceleration. The droplet response time is the time a droplet needs to reach 63 percent of its terminal fall velocity in the absence of air motion, and is given by Siebert et al. (2010) as:

$$\tau_d = \frac{\beta {d_p}^2}{18\nu} \tag{1.13}$$

where d_p is the droplet diameter. Equation 1.12 is valid only for droplet Reynolds numbers (Equation 1.3) that are $\ll 1$ (Siebert et al. 2010). The droplet Reynolds number is calculated by replacing the characteristic length scale L in Equation 1.3 with the droplet diameter d_p. It is assumed, for Equation 1.12 that the ratio of Lagrangian and gravitational acceleration is small. Since these assumptions might not be valid in turbulent clouds, it is useful to study

the impact of droplet inertia on its velocity in general. According to Siebert et al. (2010), for small values of τ_d, a perturbation analysis of Equation 1.12 yields the following explicit expression for droplet velocity:

$$v_i = u_i + \tau_d g_i - \tau_d a_i \tag{1.14}$$

where $a_i = \frac{du_i}{dt}$ is the Lagrangian acceleration of the fluid at the particle location. According to Granger (1995), $a_i = \frac{\partial u_i}{\partial t} = \frac{\partial^2 u_i}{\partial t^2}$ (where acceleration results from the derivative of velocity). Qualitatively, Equation 1.14 infers cloud droplets tend to move with the surrounding air, but with a relative motion resulting from gravitational settling (second term) and from the inertial response of the droplet to fluid accelerations (third term). Equation 1.14 has important implications for the mechanism of relative droplet motions in a local region of the fluid. Relative motion of droplets will be driven by (1) spatial variations in the fluid velocity u_i (for small separations, this is controlled by the velocity gradient $\frac{\partial u_i}{\partial x_j}$); and (2) induced motion between droplets of different τ_d. In other words, droplets of different diameters respond to fluid accelerations differently (Siebert et al. 2010).

Before we continue with Equation 1.14, we must first return to the shear production/loss term (also known as the deformation term) presented in Equation 1.5. The term $\frac{\partial \bar{U}_i}{\partial x_j}$ is the strain rate tensor, defined by Mei and Vernescu (2010) as:

$$\frac{\partial U_i}{\partial x_j} = \frac{1}{2}\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}\right) \tag{1.15}$$

The strain rate tensor is a physical quantity that describes the rate of change of the deformation of a substance (air) at a specific interest point. According to Mei and Vernescu (2010), any velocity gradient tensor can be separated into a symmetric part, which is the strain rate tensor (s_{ij}) and an asymmetric part, in this case the rotation tensor or vorticity tensor (Ω_{ij}):

$$\frac{\partial U_i}{\partial x_j} = \frac{1}{2}\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}\right) + \frac{1}{2}\left(\frac{\partial U_i}{\partial x_j} - \frac{\partial U_j}{\partial x_i}\right) = s_{ij} + \Omega_{ij} \tag{1.16}$$

For further information on the derivation of the strain rate tensor and vorticity tensor see Granger (1995) pages 172–175. Returning to Equation 1.14, omitting the first and second term, the velocity gradient can be calculated by taking the divergence of Equation 1.14, where divergence is $\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}$ from Holton and Hakim (2013), resulting in:

$$\frac{\partial v_i}{\partial x_j} = \frac{\partial U_i}{\partial x_j} = -\tau_d\left(\frac{\partial U_i}{\partial x_j}\frac{\partial U_j}{\partial x_i}\right) \tag{1.17}$$

where this defines the divergence of a tensor (Flugge 1972). With the help of Equation 1.16, Equation 1.17 can be rewritten as:

$$\frac{\partial U_i}{\partial x_j} = -\tau_d({s_{ij}}^2 - {\Omega_{ij}}^2) \tag{1.18}$$

The conclusion can be made from Equation 1.18 that the droplet velocity field is divergent in regions of high vorticity and convergent in regions where strain dominates (Maxey 1987; Shaw 2003). To state another way, particles will tend to accumulate in regions of low vorticity and high strain. Keep in mind that divergence (convergence) results when the divergence equation is positive (negative)(Holton and Hakim 2013).

The particle response time τ_d is not the only value that determines whether or not particles will tend to cluster. The ratio of τ_d to the relevant timescale of fluid accelerations is also of importance. Because clustering is associated with vorticity, and the vorticity spectrum peaks at small scales (see Section 1.2.3), the Kolmogorov timescale τ_K (see Section 1.2.4) is the relevant fluid time scale, where

$$S_t = \frac{\tau_d}{\tau_\eta} = \frac{\rho_w d^2 \varepsilon^{\frac{1}{2}}}{18\rho_a \nu} \tag{1.19}$$

is the Stokes number (Vaillancourt et al. 2002). It is expected that maximum preferential concentration results when $S_t \approx 1$. Particles with $S_t \gg 1$ react very slowly to the changes in the flow due to large particle response times, while particles with $S_t \ll 1$ behave like Lagrangian floats, following the flow exactly due to the very small particle response times (Vaillancourt et al. 2002).

1.4 Clustering Measurement Methods

Although the understanding of the basic mechanisms causing clustering are constantly improving due to direct numerical simulations and laboratory investigations, the characterization of deviations from perfect randomness for real clouds in the atmosphere remains primarily statistical in nature (Larsen 2012). There are multiple tools that can be used to measure droplet clustering using a time series of droplet detection times including but not limited to: the Fishing statistic, the clustering index, the volume–averaged pair–correlation function, and the pair–correlation function (PCF). Many studies (e.g., Shaw et al. (2002); Shaw (2003); Larsen (2006)) have attempted to examine multiple statistical techniques in

the same paper so as to show connections between the methods or to argue for the primacy of one tool over others for a particular set of data. It is important to understand the basic statistical tools listed above that are used to characterize deviations from perfect randomness in clouds. It will become apparent that a careful understanding of statistical stationarity/homogeneity is critical when examining cloud droplet structures. This is due to the fact that the particular counting statistic used may deviate from zero (where zero represents a perfectly random dataset (Poisson)) not from a statistical correlation or clustering between subvolumes of scale t (where t refers to the temporal lag), but rather due to concentration fluctuations that may occur at scales equal to, larger than, or less than t (Larsen (2012) Appendix B). Unfortunately, there are no good techniques in proving if a system of data is statistically stationary or not. However, information on some possible techniques to determine if a time series may or may not be stationary is discussed in Section 2.4.

As was just stated, to use the statistical tools listed above, the counting process has to be statistically homogeneous (stationary). In other words, the mean and variance of the number of counts are assumed to be constant over the analyzed time interval. This requirement can be met when sampling horizontally homogeneous clouds, such as stratus. For more turbulent cumulus clouds, where stronger entrainment and mixing processes occur, this condition is usually only fulfilled for subsections of the cloud. The longer the analyzed cloud section, the more statistically significant the calculation will become due to a larger droplet population (see Section 2.3.2 for a better understanding), but the stationarity of the data will be compromised for longer cloud sections. Thus, for calculations with the tools listed above, a compromise between statistical homogeneity and counting statistics has to be found. For better comparisons, all PCF measurements will be normalized such that they decay to zero (see Section 1.4.3) to represent stationary data.

1.4.1 The Clustering Index and Fishing Statistic

The Clustering Index (CI) is the first of several statistical tools that rely on binning the data at variable scales. The total measurement time T is divided into m equal intervals of duration $\frac{T}{m} = t$, where t may be varied (temporal lag) and is the independent variable for the hypothesis testing statistics (Baker and Lawson 2010). After binning the data into these intervals, quantities of V_t and M_t are defined as the variance ($\overline{(\delta N)^2} \equiv \overline{(N - \bar{N})^2}$) and mean number of droplets ($\bar{N}$) associated with each interval. The CI can then be shown as the

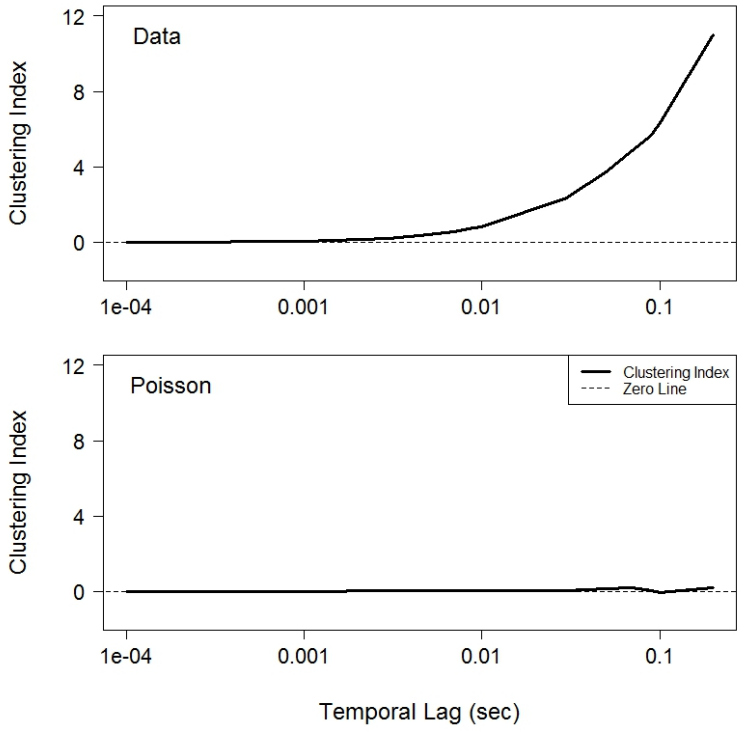

Figure 1.4: Illustration of the clustering signature for the clustering index. The top panel represents real data, with a temporal scale of 16 seconds and 13943 droplets. The bottom panel represents Poisson simulated data with the same temporal scale and droplet count.

relationship between the sample variance and sample mean:

$$CI(t) = \left(\frac{V_t}{M_t} - 1 \right) \tag{1.20}$$

For a perfectly random, statistically stationary dataset (Poisson), CI will equal zero. The CI becomes positive when positive spatial correlations are dominant (Larsen 2012; Shaw et al. 2002). An important point to be made is that one must interpret the CI with caution. Due to its inherent volume dependence it is not uniquely related to a single spatial scale, but instead represents contributions from a range of scales. The clustering index is said to contain memory of all scales within the volume (bins) (Kostinski and Shaw 2001). The top portion of Figure 1.4 shows the clustering index for data from a randomly selected cloud from the GoMACCS campaign, covering 16.8 seconds ($\sim$ 1000 m) with a population of 13943 drops (this is the same cloud that was used to create Figures 1.1–1.2). The bottom portion of Figure 1.4 shows simulated Poisson point data with the same temporal scale and

droplet population. In Figure 1.4 the cumulative nature of CI is clearly seen. For example, it increases monotonically as the temporal scale gets larger. The top portion of Figure 1.4 confirms that the relatively small magnitude of CI at small scales does not necessarily imply a lack of clustering at those scales. Similarly, the relatively large magnitude of CI at larger scales is a manifestation of clustering at smaller scales adding up. The Poisson point data shown in the bottom panel is consistently zero with slight deviations at larger temporal scales. The zero values confirm the data follows the Poisson process, as expected.

The Fishing statistic is very similar to the CI, and therefore will only briefly be mentioned. The Fishing statistic was introduced to the cloud droplet detection community by Baker (1992). It is structured similarly to the CI but scaled differently. It is designed to be a hypothesis testing statistic to determine whether a set of data points is Poisson distributed (Baker and Lawson 2010). Using the same notation as developed above, we can write the Fishing statistic at scale t, shown as:

$$F(t) = \left(\frac{V_t}{M_t} - 1 \right) \sqrt{\frac{T-t}{2t}} \tag{1.21}$$

The Fishing statistic also gives a value of zero at all values of t for a stationary Poisson process, but has a t–dependent weighting factor to help $F(t)$ from increasing with t for non–Poisson systems (as is seen in the top portion of Figure 1.4 (Larsen 2012)).

1.4.2 Volume-Averaged Pair–Correlation Function

The volume-averaged pair–correlation function (VAPC, $< \eta >$ or $\bar{\eta}$ depending on the source) was introduced in Shaw et al. (2002) and used again in several other studies including Shaw (2003) and Baker and Lawson (2010). The VAPC as defined by Baker and Lawson (2010) is shown to be:

$$VAPC(t) = \frac{1}{M_t} \left(\frac{V_t}{M_t} - 1 \right) = \frac{1}{M_t} [CI(t)] \tag{1.22}$$

This measure has previously been used because it is computationally easier to measure than the PCF (see Section 1.4.3) at the expense of not being a localized measure of deviations from pure randomness (Larsen 2012). The advantage of the VAPC over CI is the fact that it possesses greater reduced scale memory (or greater localization) relative to CI due the weighting of CI by $\bar{N}$. Figure 1.5 uses the same data as Figure 1.4, except it shows the clustering signature for the VAPC. As can be seen in Figure 1.5, the clustering can no longer

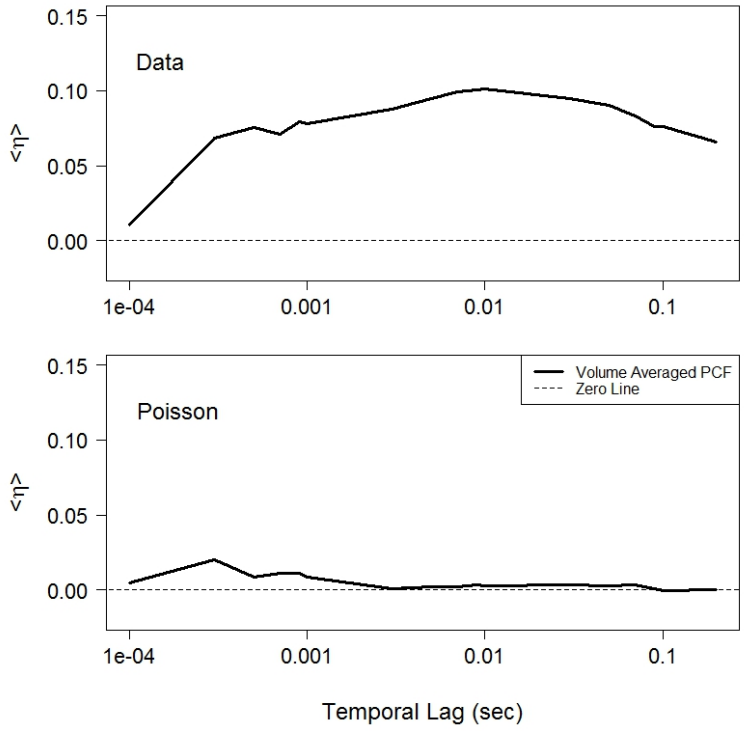

Figure 1.5: As in Figure 1.4, but for the VAPC.

be completely localized or attributed to a given scale. Because of the volume averaging, it has memory of sub–scales. This is reflected in the peak delay and slower descent (as compared to the PCF in Section 1.4.3). Again, the Poisson data is around zero with slight deviations. Although VAPC has a greatly reduced scale memory compared to the CI, this does not lead to a perfect lack of scale memory, only reduced scale memory.

1.4.3 Pair–Correlation Function (PCF)

The PCF physically can be introduced as a scale–localized deviation from a stationary Poisson distribution. The PCF has been discussed in great length in multiple publications including but not limited to (Shaw et al. (2002); Larsen (2006, 2012)). Due to the finite size of droplets it is clear that the inter–droplet separation distance can never be less than one droplet diameter, implying that projected separation distances cannot be exponentially distributed (Kostinski and Shaw 2001). In other words we know, because of finite size effects, that clouds are not perfectly random at spatial–scales less than one droplet diameter; but we would still like to know whether they are perfectly random at mm or cm scales (Shaw

et al. 2002).

In defining the PCF, we consider a volume element dV sufficiently small so that it can contain only one droplet, and so that the probability of finding a droplet in that volume is $\bar{N}dV$. If droplet positions are uncorrelated then it is expected that the probability of finding droplets in each of two volume elements dV_1 and dV_2 separated by distance r is, from Shaw et al. (2002)

$$P_r(1,2) = (\bar{N}dV_1)(\bar{N}dV_2) \tag{1.23}$$

If, however, the probabilities are not independent we define the PCF ($\eta(r)$) as:

$$P_r(1,2) = (\bar{N}dV_1)(\bar{N}dV_2)\{1+\eta(r)\} \tag{1.24}$$

The PCF is zero for perfect randomness and has a lower limit of –1, e.g. for scales less than the diameter of particles. If the PCF is greater than zero, it implies that if a droplet is encountered at a given position in a cloud, there is an enhanced probability of finding another droplet distance r away (Shaw et al. 2002).

Note that the current definition of the PCF is based on distance, and also note that timescales have been used for the definitions of both the VAPC and CI. Just as the VAPC and CI can be defined in terms of distance, so to can the PCF be defined in terms of time, which is used in this thesis. Following the same basis for Equations 1.23 and 1.24, given a particle detected at some time t_o, the probability of a particle detection in some later time window $(t_o + t, t_o + t + dt)$ would, for a stationary Poisson process, be given by:

$$[p(t_o + t|t_o)dt]_{poisson} = \lambda dt \tag{1.25}$$

from Larsen (2012). Where Equation 1.25 is the time equivalent to the length defined Equation 1.23. Note that λ is the mean number of particles per bin. Equation 1.25 results from the fact that the presence or absence of a particle at t_o has no bearing on the presence of a particle in (t_o+t, t_o+t+dt) for a stationary Poisson process. For a statistically stationary (but not Poisson) process, we may have an enhanced or diminished probability of finding a second particle in the specified window. Therefor Equation 1.25 can be modified to:

$$p(t_o + t|t_o)dt = \lambda dt[1+\eta(t)] \tag{1.26}$$

where Equation 1.26 is the time equivalent to the length defined Equation 1.24. Here $\eta(t)$ is the PCF evaluated at t. Equation 1.26 can be rearranged to produce an expression for $\eta(t)$

from Larsen (2012):

$$\eta(t) = \frac{p(t_o + t|t_o)}{\lambda} - 1 \tag{1.27}$$

Figure 1.6 shows the PCF for real data (top panel) and Poisson simulated data (bottom panel). The data used has the same characteristics discussed for Figures 1.4 and 1.5. The red line represents a running mean of five for the PCF values (shown in black). The real data shows a peak at smaller temporal scales and a steady decrease to a PCF value of 0.04. This begs the question, why don't we see a decrease to zero (Poisson process) at larger temporal scales? The answer is based on the data set. The data is not stationary, resulting in the data, even at larger temporal scales, being non–Poissonian. As was discussed in Section 1.4, a deviation is shown from a perfectly random dataset not from a statistical correlation or clustering at scale t, but rather due to concentration fluctuations at different scales of t. This can be accounted for through normalizing the PCF curve, taking the average value at which the PCF decays to and normalizing it to zero. The Poisson simulated point data shows PCF values varying around the zero line, as is expected. The main takeaway from Figure 1.6 is the fact that the largest droplet clustering can be seen at smaller temporal scales below 0.05 seconds (this corresponds to the strongest clustering being below 30 cm), whereas the data starts to decrease towards a Poisson distribution at larger temporal scales at values of 0.05 seconds.

The advantage of the PCF is the fact that it is scale localized: $\eta(t)$ depends only on the presence or absence of particles separated by $t \pm dt$ in time. Though both the CI and Fishing statistic give values of zero for all scales from a stationary Poisson dataset, they can both have inflated values at scale t due to clustering on scales smaller than t (Larsen 2012). It is shown in both Shaw et al. (2002) and Larsen (2012) that the CI, VAPC, and PCF can all be related to each other mathematically. Although these relationships will not be derived here, it is important that each of the clustering statistical tools discussed above have mathematical relationships. Shaw et al. (2002) gives a nice summary of why the PCF is ideal for quantifying droplet clustering. These can be summarized as: (1) the PCF can be used to derive other commonly used measures of droplet clustering (not shown or discussed here); (2) the PCF is scale localized; and (3) the PCF can be extended to the theory of radiative transfer, collision–coalescence, and rain characterization. Due to the advantages of using the PCF over the other statistical tools, the PCF will be used throughout this thesis to measure droplet clustering.

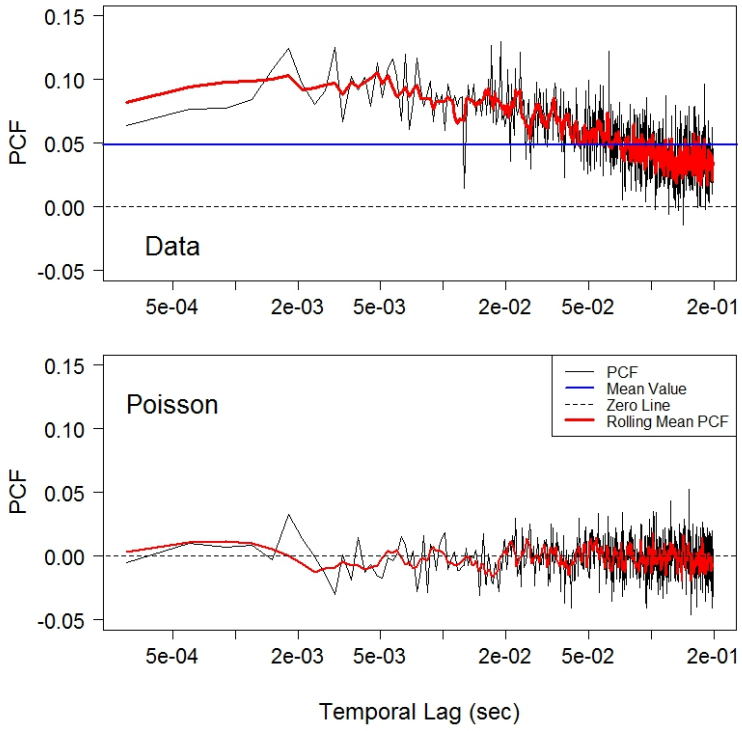

Figure 1.6: As in Figure 1.4, but for the PCF.

1.4.4 Poisson vs. Real Data

In order to get a better understanding of Poisson point data versus real data collected from a cloud, the data that was used to create Figures 1.4–1.6 will be analyzed. All four statistical tools discussed depend on binning the data at a variable time scale (the x–axis). A temporal lag of 0.001 results in binning the 16 second time series used above into bins of equal time (0.001 second bins) whereas larger temporal lags result in binning the data into larger time bins such as 0.02 second bins. The number of droplet counts that fall into each bin can then be recorded and used to create a histogram which shows the range of droplet counts in each bin (see Figure 1.7).

Figure 1.7 was created using a time lag of 0.02 seconds. Instead of the frequency of occurrences for the number of drops in the bins, the density is shown. A density curve is a graph of a continuous probability distribution and gives a total area under a curve equaling one. Every point on the curve must also have a frequency of occurrences that is equal to or greater than zero. The blue curve represents a Poisson distribution, where the lambda used to calculate it is equal to the mean of the number of drops recorded in each bin. It can

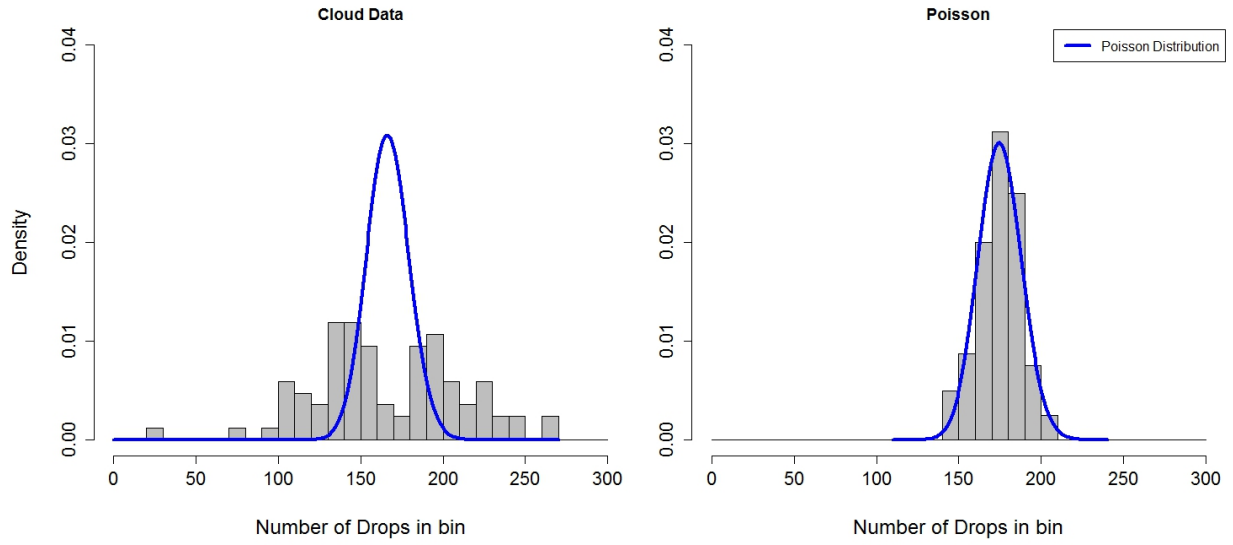

Figure 1.7: Shows the frequency of the number of drops recorded for each bin (with a temporal lag of 0.02 seconds for a 16 second time interval), with a Poisson distribution overlaid in blue.

clearly be seen that the Poisson simulated point data follows the Poisson distribution very closely, with droplet counts per bin ranging from 140 to 200. However, the cloud data has a much more variable range (29 to 269) of droplet counts per bin and does not follow the Poisson distribution. It is because of this departure from a perfect Poissonian distribution that the "data" panels in Figures 1.4–1.6 deviate from zero. Thinking about the PCF and Figure 1.6, at smaller temporal lags it is expected that the data will have a much larger range of droplet counts per bin (greater clustering). For larger temporal scales (on the order of 0.1 seconds) the data will have a much smaller range of droplet counts per bin, representing that of a Poisson distribution seen in the right panel of Figure 1.7.

Going back and thinking about Figure 1.1, an examination of why clustering is more significant at smaller length scales can be made. Assuming the data is stationary, at larger temporal scales an equal amount of droplets per bin should be seen. Table 1.1 shows three different bin sizes, with the corresponding mean number of drops per bin, the range of the drops, and the difference in the range. It is shown that smaller bin sizes result in a larger difference in the range of droplets per bin. Although smaller bins result in a lower number of drops per bin, you still get a superior range. It is this greater variability at smaller scales that is responsible for the greater PCF values at these scales and defines the basic inner–workings of the PCF. If the data is non–stationary, you would then see a larger range of droplets per bin at greater temporal/length scales not due to clustering, but due to concentration

fluctuations at a different scale than that being examined.

Table 1.1: Different size bins are recorded with the mean number of drops per size bin, the range of the drops in each size bin, and the difference in the range. The data used is from the top panel of Figure 1.1.

Bin Size (m)	Mean Number of Drops	Range of Drops	Difference in Range
5	94	89–98	9
1	19	13–25	12
0.3	12	4–19	15

CHAPTER 2
MOTIVATION AND METHODS

2.1 Framework for Chapter 2

There are several hypotheses to explain the observed time of rain formation in shallow warm cumulus clouds (which is in disagreement with theory) as introduced in Chapter 1. These hypotheses will be reviewed and a look at how inertial clustering can affect collision–coalescence will be examined. Additional information on the PCF along with properties that are typically observed when using the PCF will be looked at. Added information on data stationarity is discussed along with the instrumentation used to collect data. Thereafter, the objective and hypotheses of this work is explained.

2.2 Motivation

2.2.1 Entrainment and Mixing

The entrainment process describes the mixing of subsaturated, ambient air with saturated cloud air. The fraction and the thermodynamic properties of the entrained air influence the development of the cloud and may determine whether a cloud persists or decays. The characteristics of the mixing process are crucial for the evolution of the cloud droplet size distribution and in the formation of precipitation. For a turbulent mixing process involving a phase change, the ratio of the mixing time scale τ_{mix} to the time for complete phase change τ_{phase} is of importance (Dimotakis 2004):

$$Da = \frac{\tau_{mix}}{\tau_{phase}} \tag{2.1}$$

This ratio is called the Damköhler number (Da). In the case of clouds, dry subsaturated air is entrained in discrete patches, which then break down to smaller and smaller scales, while the boundaries between saturated cloud air and subsaturated air remain intact (Blyth 1993) until the Kolmogorov microscale l_K (see Section 1.2.4) is reached. The time scale for

an eddy of size l to break down to l_K is given by Baker et al. (1984) as:

$$\tau_{mix} = \left(\frac{l^2}{\varepsilon}\right)^{\frac{1}{3}} \tag{2.2}$$

where l is typically taken to be the Taylor microscale l_T ($\sim 2 \cdot 10^4$ in deep convective clouds and $\sim 5 \cdot 10^3$ in stratiform clouds (Pinsky et al. 2006)) (see Section 1.2.4 for information on the Taylor microscale). When the mixing has proceeded and scales close to l_K are reached, the last step of homogenization is done by molecular diffusion, with the time–scale

$$\tau_{diff} = \frac{l_2^K}{C_{diff}} \tag{2.3}$$

where C_{diff} is the diffusion coefficient ($\sim 2.2 \cdot 10^{-5}$ m^{2}s^{-1} for water vapor in air). Usually, $\tau_{mix} \gg \tau_{diff}$, and τ_{mix} is thus the appropriate timescale that determines if droplets can evaporate before mixing is complete. In Equation 2.1, τ_{phase} is given by the evaporation time for droplets, which for a droplet of diameter d_p is given by Pruppacher and Klett (1997) as:

$$\tau_{phase} = \frac{{d_p}^2 \rho_w R^* T}{8 M_w C_{diff}(e_s(T) - e)} \tag{2.4}$$

where $R^* = 8.314$ J mol^{-1}K^{-1} is the universal gas constant, T is the temperature, $M_w = 18$ g mol^{-1} is the molecular weight of water, e is the water vapor pressure of the air surrounding the droplet in Pascals, and e_s is the saturation water vapor pressure at T.

If the Damköhler number is $\ll 1$, the mixing is fast compared to the evaporation and all droplets experience similar conditions with respect to saturation. Thus, the cloud drop number concentration (CDNC) is diminished by dilution and the droplet size distribution is shifted towards smaller sizes. This is referred to as homogeneous mixing. If mixing is slow compared to evaporation, $Da \gg 1$, the boundaries of the entrained air exist long enough that droplets in the vicinity of these boundaries may completely evaporate before the volumes are completely mixed, while other droplets remain unaffected by the entrained air. During this process, known as inhomogeneous mixing, the CDNC is diminished by dilution and additionally by the evaporation of some droplets, while the droplet size distribution retains its mode diameter.

2.2.2 Adiabatic Droplet Growth

A simplified equation for the condensational growth of an activated cloud droplet is given by

$$r\frac{dr}{dt} = \frac{(S-1)}{F_K + F_D} \tag{2.5}$$

from Rogers and Yau (1989) and Pruppacher and Klett (1997), where r is the droplet radius, $S = \frac{e}{e_s}$ is the ambient water vapor saturation ratio with the water vapor pressure e and the saturation water vapor pressure e_s, and F_K and F_d are two terms associated with heat conduction and vapor diffusion, respectively. In Equation 2.5, the curvature and the solution term of the Köhler curve are neglected (these are reasonable assumptions after the droplets are activated and have grown to a few micrometers). From looking at Equation 2.5, it can be seen that a supersaturation greater than zero will lead to droplet growth. It can also be seen that the droplet growth rate is inversely proportional to the droplet size. This leads to a narrowing of the drop number size distribution (DNSD) due to droplets on the small end of the number size distribution growing faster than the droplets on the large end.

When droplets move relative to each other in a cloud, they may collide and coalesce. The collision kernel (κ) in still air has the form

$$\kappa = \pi\left(\frac{d_p}{2} + \frac{d_p{}'}{2}\right)^2 E|v_{rel}| \tag{2.6}$$

from Sundaram and Collins (1997), with d_p and $d_p{}'$ the diameters of the two colliding droplets, the collision efficiency E defined as the ratio of the actual to the geometric collision cross section ($E = \frac{{y_c}^2}{(r_1+r_2)^2}$, where y_c is the difference between the distance between the center of r_1 and r_2), and the relative velocity of the two droplets $|v_{rel}|$. In the absence of turbulence, $|v_{rel}|$ is given by the difference in the terminal fall velocities of the two droplets. Equation 2.6 describes the volume of smaller droplets of size $d_p{}'$ that a droplet of diameter d_p sweeps out and collects during a given time unit. The collision kernel varies approximately as ${d_p}^6$ for $20 \leq d_p \leq 100$ μm and as ${d_p}^3$ for $d_p > 100\mu$m (Pruppacher and Klett 1997). Thus growing by collision–coalescence is a fast process. Once started, the collision process increases rapidly with droplet size. From Equation 2.6 it can be seen that a difference in fall velocity (in the absence of turbulent air motion) is needed to onset the collision–coalescence process. Thus the DNSD must be broad enough for some droplets to fall faster than others. It is this broadness that is not predicted by growth through vapor diffusion alone and raises the questions in warm rain formation. The crossover in dominance from condensational to

collection growth occurs for $20 \leq d_p \leq 40\mu$m (Pruppacher and Klett 1997).

2.2.3 Alternative Droplet Growth Mechanisms

Several hypotheses have been postulated to explain the fast onset of warm rain. Some of them will only be briefly mentioned, whereas the main possibility (enhanced collision–coalescence) analyzed in the course of this work will be discussed in more detail in Section 2.2.4. Typical droplet concentrations vary between a couple tens cm^{-3} in fog (Wendisch et al. 1998) to a few hundred cm^{-3} in maritime clouds and values up to 2300 cm^{-3} in continental cumulus clouds (Conant et al. 2004). The concentration of raindrops is typically 10^{-3} cm^{-3}. Thus only about one of 10^6 to 10^7 droplets must grow to larger than average size and onset the coalescence process in order to initiate precipitation (these droplets are sometimes referred to as the "lucky" ones (Kostinski and Shaw 2005)).

Many scientists have suggested giant (GN) and ultragiant (UGN) nuclei (with diameters greater than 2 μm and 10 μm, respectively) to be responsible for the fast onset of precipitation in warm clouds. If these nuclei are hygroscopic and are exposed to relative humidities above their deliquescence relative humidity, they grow to sizes that are large enough to initiate coalescence with smaller droplets in the cloud without the limitations imposed by the condensation process. There is no consensus about the overall importance of GN and UGN however. Laird et al. (2000) and Knight et al. (2002) found GN and UGN composed of sea–salt aerosol at moderate wind–speeds in sufficient concentration to explain the drizzle and rain drop concentration in maritime cumulus clouds. On the contrary, several model studies showed that GN and UGN are negligible (Rosenfeld et al. 2002; Feingold et al. 1999).

It was suggested that turbulent fluctuations of the supersaturation ($S - 1$) field might contribute to the broadening of the DNSD (e.g., Srivastava (1989)) due to the enhanced growth of some droplets. These fluctuations can either be the result of fluctuations of the vertical wind velocity or of the temperature and water vapor field. Srivastava (1989) found that fluctuations of the supersaturation arising from fluctuations of vertical velocity cannot significantly contribute to the broadening of the DNSD, because high values of $S - 1$ are usually correlated with high values of vertical velocity, thus reducing the time a drop can grow before it reaches a certain cloud level. Also, Vaillancourt et al. (2002) concluded that fluctuations of the supersaturation field alone cannot explain the observed broadness of the DNSD.

There are several mechanisms by which turbulence can affect collision rates. Based on the work of Sundaram and Collins (1997), Wang et al. (1998) suggested that the collision

kernel in a turbulent flow field should be approximated as:

$$\tau_{turb} = 2\pi \left(\frac{d_p}{2} + \frac{d_p{}'}{2} \right)^2 E \overline{v_{rel} \frac{L}{|L|}} [\eta \left(\frac{d_p}{2} + \frac{d_p{}'}{2} \right) + 1] \tag{2.7}$$

where L is the separation vector between the droplets positions, and $\eta \left(\frac{d_p}{2} + \frac{d_p{}'}{2} \right)$ is the PCF as defined in Section 1.4.3 at contact. Here, $\overline{v_{rel} \frac{L}{|L|}}$ denotes the average over all orientations of L and the spatial average. Comparison with Equation 2.6 reveals that turbulence can enhance the collision rates by several mechanisms:

- The combined effect of droplet inertia, turbulent fluid accelerations, and shear leads to an increased relative radial droplet velocity (Sundaram and Collins 1997; Pinsky and Khain 1996; Zhou et al. 2001)
- The deviation of the droplet spatial distribution from randomness caused by the accumulation of inertial droplets in regions of high strain and low vorticity as quantified by the PCF (Reade and Collins 2000).

Additionally, the collision efficiency E itself can be enhanced by turbulence (Pinsky et al. 1999; Wang et al. 2005). Also, Ghosh et al. (2005) showed that interactions of droplets with the smallest eddies of length scale comparable to l_K yield an enhanced settling rate of droplets smaller than a critical size, leading to higher collision rates than in still air. All of these processes depend on the turbulence intensity of the flow, the droplet response time and the Reynolds number. It is this enhanced collision kernel that could possibly explain the difference between the rain formation times observed in shallow cumulus clouds by Laird et al. (2000) and Szumowski et al. (1997) of 15 to 20 minutes and the calculations made by Jonas (1996) who found that rain formation ($\sim 100\mu$m) should take upwards of 85 minutes.

Entrainment and subsequent mixing of subsaturated cloud–free air at the lateral cloud edges or at cloud top is thought to have a significant influence on the cloud microphysics and may alter the shape of the DNSD by broadening it towards smaller and larger sizes. Although entrainment of dry air lowers the cloud water content (and thus has a negative impact on rain development), it can result in a dramatic impact on the cloud droplet spectra. In particular, broad spectra are typically produced as a result of entrainment and mixing through the evaporation of drops (leading to smaller sizes) and the enhancement of some drops (some drops experience no evaporation, and are therefore larger than droplets that do

experience evaporation) in inhomogeneous mixing (Khain et al. 2000). The resulting effects on large droplet formation and coalescence initiation remain unknown however.

2.2.4 Turbulence and Collison–Coalescence

Observations of radar reflectivity in tropical regions suggest that rain could form in cumulus clouds by the warm rain process in approximately 15 to 20 minutes (Szumowski et al. 1997; Laird et al. 2000). While there are still uncertainties in linking radar reflectivity and the rain initiation time as the radar reflectivity depends on both the average droplet size and the LWC (Knight et al. 2002), the above rain initiation time is usually quoted as the average time for the initiation of warm rainfall in unpolluted, maritime clouds (i.e., from the initial development of a cumulus cloud to the first appearance of rain). Theoretical predictions require a time interval on the order of 85 minutes for droplets to grow from 20 to 100 μm in radius (the actual time depends on the cloud water content, initial droplet size spectrum, updraft speed, etc.; see Jonas (1996)). Therefore, there appears to be a factor of 2 or more difference between the predicted growth time and the observed growth time. Cloud droplets of radii less than 10 to 15 μm grow efficiently through diffusion of water vapor (see Section 2.2.2) while droplets larger than 30 to 50 μm in radii grow efficiently through gravitational collisions (Langmuir 1948; Kogan 1993; Pruppacher and Klett 1997).

In general, it is difficult to explain the rapid growth of cloud droplets in the size range from 10 to 50 μm in radius (the so–called size gap) for which neither the condensation nor the gravitational collision–coalescence mechanism is effective. The inclusion of turbulence allows droplets to collide and coalesce even when the droplets are of the same size. Equation 2.6 shows the collision kernel under still conditions. In other words, Equation 2.6 gives the collision kernel under gravitational collision–coalescence. Gravitational collision–coalescence forms the basis of many weather and climate models, and assumes that two independent droplets (D1 and D2) can only collide when the diameter of D1 $\neq$ D2. Although it has been proven that turbulence has a significant effect on the motion and collisions of drops, a quantitative evaluation of the effect presents a complicated problem. The main difficulties as explained from Pinsky et al. (2006) are related to:

- No measurements of the fine turbulent structure in clouds have been carried out so far. Most laboratory experiments are conducted for Reynolds number values that are much smaller than those typically found for atmospheric turbulence (especially that of clouds, since clouds are known as areas of enhanced turbulence as compared with cloud–free

atmosphere). Since the structure of a turbulent flow depends on the Reynolds number, it is not clear how the results obtained in laboratory experiments can be extended to atmospheric conditions.

- There is no known appropriate numerical model of a turbulent flow that is designed specifically for the purpose of simulating droplet collisions. Turbulent diffusion is often simulated using the Langevin stochastic differential equations (Pope 2000). In these simulations, the parcel velocity generated by a finite–difference analog of the Langevin equation is not a differentiable function, which does not allow realistic reproduction of particle accelerations and collisions.

The work conducted here therefore becomes important due to (1) A dataset is provided that gives *in situ* cloud droplet spatial data that allows for an analysis of spatial variability and (2) droplet spatial variations can be compared to laboratory measurements to see if laboratory experiments can indeed be extended to atmospheric conditions. A measurement of the PCF (see Chapter 3 for results) also allows for the evaluation of the enhanced collision kernel from Equation 2.7. Although preliminary, these results can generate a base for being able to determine how turbulence enhances collision–coalescence and eventually explain the time difference between rain formation in observation (20 minutes) and theory (85 minutes).

2.2.5 Indirect Effects

Most of what is discussed here is dependent on the assumptions that droplet clustering enhances collision-coalescence (Section 2.2.4). Also, most scientific work has focused on the physics of inertial clustering and how droplet clustering in the lab can be related to atmospheric conditions. Consequently, very little work has looked at how droplet clustering can effect cloud aerosol responses. Therefore, the effects that aerosols have on droplet clustering will only briefly be mentioned and are based on assumptions and theory. It was discussed in Section 1.1 that aerosol–cloud interaction include cloud albedo and cloud lifetime effects, along with an evaporation–entrainment feedback mechanism

- **Cloud Albedo Effect**

  The cloud albedo effect is defined by an increased cloud albedo due to the increased droplet concentration that results with greater anthropogenic aerosol forcing. Can droplet clustering enhance this effect? According to Matsuda et al. (2012), turbulent clustering causes the formation of void regions of droplets and hence increases the

direct transmittance of radiation. This effect decreases as the turbulent Reynolds number increases and is estimated to be negligibly small under the conditions in real clouds however. Other research (Borovoi 2002; Kostinski 2002) has concluded that the effect of inhomogeneity due to turbulence on radiative transfer is unclear. If there is enhanced collision–coalescence however, then one can assume that the droplet sizes will tend to be larger and therefore the albedo of the cloud will decrease. This is due to the fact that clouds that have a small number of larger droplets tend to be less translucent. One can then theorize, that if clustering does change as a function of aerosol number concentration, that the clustering will either enhance or diminish the albedo effect.

- **Cloud Lifetime Effect**

  The lifetime effect is defined by anthropogenic aerosol suppression of precipitation which results in clouds with more liquid water, higher factional cloudiness, and longer lifetimes (Small et al. 2009). Again, if the assumption that collision–coalescence is enhanced by droplet clustering is made, then the lifetime effect could be reduced through enhanced precipitation formation, depending on if/how clustering changes as a function of aerosol number concentration.

- **Evaporation–Entrainment Feedback**

  When small, non–precipitating cumulus experience aerosol perturbations, they respond in a manner consistent with enhanced evaporation around the shell of the cumulus and a stronger horizontal buoyancy gradient is produced. This leads to stronger vorticity around the core of the cloud and stronger entrainment (Small et al. 2009). Enhanced evaporation occurs due to the presence of smaller, more numerous droplets that are easier to evaporate than larger droplets. However, if enhanced collision–coalescence is present due to droplet clustering, (especially with evaporation leading to stronger vorticity and enhanced entrainment), larger droplets will be produced which will lead to less evaporation and a decrease in the entrainment. A negative feedback loop is created, which will result in a cloud lifetime that is longer than it would otherwise be with increased entrainment and evaporation. Again, keep in mind that this depends on the assumption that collision-coalescence is enhanced, and also depends on how droplet clustering responds as a function of aerosol number concentration.

Overall, these three effects taken into account at the same time, for a cloud that is forced with anthropogenic aerosols, and if collision–coalescence is enhanced which leads to

larger droplets results in: (1) The cloud albedo will decrease (2) the cloud lifetime will decrease (larger drops could lead to rain formation) (3) the cloud lifetime will increase (less evaporation will lead to less entrainment and decrease the rate of cloud deterioration). The issue becomes how the lifetimes of the cloud from (2) and (3) interact with each other and which will be dominant. These three effects and the overall impact on the radiation budget will be looked at again at the end of Chapter 3, after the results have been analyzed and it is known how the clustering changes as a function of aerosol number concentration (and whether clustering changes from cloud center to cloud edge). An outlook will be given to determine work that can be done to explore these ideas in future research.

2.3 PCF Characteristics

2.3.1 PCF Derivation

There are many forms of the PCF that can be used for different occurrences. 3–D, 2–D, and 1–D versions of the PCF exist, reliant on the data that is being used. The 1–D PCF that is used in this thesis is derived by considering a time point process which is a random distribution of time instants t_i, and $N(t,\tau)$ (where t is equivalent to t_o and τ is equivalent to t (the time change) from Section 1.4.3) is the random variable equal to the number of points in the time interval $[t, t+\tau]$. It is assumed that this process is perfectly random and a small interval cannot contain more than one point of the process or, more explicitly, that

$$P[N(t,\Delta t)=1] = \lambda(t)\Delta t + o(\Delta t) \tag{2.8}$$

from Picinbono and Bendjaballah (2005), where P is the probability and $\lambda(t)$ is the density of the point process. If $\lambda(t)$ is constant, the process is stationary. The assumption that the data is stationary will be assumed in all that follows. In coincidence measurements one considers two or more small time intervals at different time instants. A coincidence event is characterized by the fact that there is one point of the point process in each of these intervals. Bicoincidence is described by a coincidence function $\gamma(t,t^{'}), t \neq t^{'}$, defined by (Picinbono and Bendjaballah 2005) as:

$$P\{[N(t,\Delta t)=1][N(t^{'},\Delta t^{'})=1]\} = \gamma(t,t^{'})\Delta t\Delta t^{'} + o(\Delta t,\Delta t^{'}) \tag{2.9}$$

The left hand side of Equation 2.9 is indicating what the probability of the random variable in both time intervals being equal to one is. Since the point process is stationary, this coincidence function only depends on $\tau = t - t^{'}$ and is equivalent to a function defined for $\tau \neq 0$ by

$$c(\tau) = c(-\tau) = \gamma(t, t - \tau), \tau \neq 0 \tag{2.10}$$

Picinbono and Bendjaballah (2005) state that the random variables $N(t, \Delta t)$ and $N(t^{'}, \Delta t^{'})$ become uncorrelated for large values of τ, which is expressed as

$$\lim_{\tau \to \infty} c(\tau) = \lambda^2 \tag{2.11}$$

Note that the basic property of Poisson processes is that the variables $N(t, \Delta t)$ and $N(t^{'}, \Delta t^{'})$ are independent. This yields $c(t) = \lambda^2$ for any τ and is the simplest example of coincidence function.

The coincidence function has a direct application in the description of the bunching or clustering effect in a point process (Picinbono and Bendjaballah 2005). This effect is related to the fact that the presence of a point of the process at $t = 0$ can modify the probability $P[N(t, \Delta t) = 1]$. This doesn't occur in a Poisson point process due to the fundamental property that Poisson processes are to be memoryless. Nonetheless, since the fact that there is a point at $t = 0$ is not a statistical event, a small interval $\Delta t^{'}$ at $t = 0$ is introduced and the conditional probability (Equation 2.12) is used.

$$\hat{F}(t, \Delta t, \Delta t^{'}) = P\{[N(t, \Delta t) = 1][N(0, \Delta t^{'}) = 1]\} \tag{2.12}$$

The clustering effect is described by the limit of Equation 2.12 when $\Delta t^{'} \to 0$ or

$$F(t, \Delta t) = \lim_{\Delta t^{'} \to 0} \hat{F}(t, \Delta t, \Delta t^{'}) \tag{2.13}$$

It results from Equations 2.8 and 2.9 that $F(t, \Delta t) = b(t)\Delta t$ with

$$b(t) = \frac{c(t)}{\lambda} \tag{2.14}$$

This clustering function $b(t)$ has the same properties of the coincidence function $c(t)$ and yields

$$\lim_{t \to \infty} b(t) = \lambda \tag{2.15}$$

For a Poisson process, $b(t) = \lambda$, therefor it can be stated that there is a clustering effect at t if $b(t) > \lambda$ and an anticlustering effect in the opposite case.

This leads to the understanding of the data itself. The lifetime L_k of order k is the random variable equal to the distance between a point t_i of the point process and the k^{th} point of this process posterior to t_i. Due to the assumption of stationarity, the probability distribution of L_K does not depend on t_i and therefor it can be assumed that $t_i = 0$. For all processes considered here the random variables L_K are continuous and characterized by their probability distribution function (PDF) $f_k(t)$. The quantity $f_k(t)\Delta t$ is by definition the probability to have one point of the point process in $[t, t + \Delta t]$ and $k - 1$ points in $[0, t]$, conditional to one point at $t = 0$. But it results from Equation 2.14 that $b(t)\Delta t$ is the probability to have one point in $[t, t + \Delta t]$ conditional to one point at 0. As a result we have:

$$b(t) = \frac{c(t)}{\lambda} = \sum_{k=1}^{\infty} f_k(t) \tag{2.16}$$

This equation yields the relationship between the coincidence function $b(t)$ and the set of PDFs $f_k(t)$ of the lifetimes of all orders. In order to understand this more thoroughly, we can compare Equation 2.16 to the work done by Larsen (2007). Rearranging Equation 1.27 from Section 1.4.3, we obtain:

$$p_t(t_o + t|t_o) = \lambda[1 + \eta(t)] \tag{2.17}$$

Equation 2.17 shows a PDF for the probability of observing a particle at time $t_o + t$ given an observation at t_o. Note that Equation 2.10 can be rewrote to match the syntax used in Section 1.4.3, where

$$\gamma(t, t - \tau) = p_t(t_o + t|t_o) \tag{2.18}$$

Resulting in

$$b(t) = \frac{p_t(t_o + t|t_o)}{\lambda} \tag{2.19}$$

Equation 2.16 can now be rewrote to show

$$p_t(t_o + t|t_o) = \sum_{k=1}^{\infty} f_k(t) \tag{2.20}$$

Note that Equations 2.18 through 2.20 show the same thing as shown in previous equations, now written in a different syntax. If a particle is observed at $t_o + t$ it must be the k^{th} particle

posterior to the particle at t_o where k is some positive integer. $f_k(t)$ again, is the probability density that the k^{th} particle posterior to a particle at t_o (the k^{th} nearest neighbor) is located at $t_o + t$. It is assumed that co–located particles are impossible. The PCF is obtained by combining Equations 2.17 and 2.20, resulting in:

$$\eta(t) = -1 + \frac{1}{\lambda} \sum_{k=1}^{\infty} f_k(t) \tag{2.21}$$

(Larsen 2007). Each of the $f_k(t)$ can be estimated from the observed inter–arrival distributions (time between droplet arrival), thus allowing a computationally simple way to compute the PCF from particle arrival time data.

2.3.2 PCF Variation

In order to calculate the k^{th} nearest neighbor (as mentioned above), a maximum time interval (t.max) and a time bin (dt) must be selected. Careful consideration must be given however. If dt is too small, the PCF will be too noisy. Set dt too large, you end up doing unnecessary scale averaging which results in a poor estimate of the PCF. Typically, t.max is an order of magnitude or so above the mean inter–arrival time (MIT) (mean time between each droplet within the data) of the particles and sets the maximum temporal lag (x–axis in Figure 1.6 (Section 1.4.3)). A dt of 0.0003 seconds and a t.max of 0.2 seconds was selected for all PCF calculations throughout this thesis. This results in a vector ranging from 0.000 to 0.2, over a sequence of 0.0003 (i.e. (0.0000, 0.0003, 0.0006, 0.0009,..., 0.1997, 0.2)). This vector gives the x–axis location for each PCF measurement. The PCF is calculated by binning the inter–arrival times of the droplets into the vector sequence seen above. An inter–arrival time is first determined between every subsequent droplet, then the inter–arrival time is binned and summed (the sum for each inter–arrival time per bin). An inter–arrival time is then determined for every other droplet, between every third droplet, every fourth droplet, and so on. The inter–arrival time is binned and added to the previously summed binned inter–arrival times. This is done up until the minimum inter–arrival time in the data is no longer less than t.max (0.2 seconds). The total summed binned data is then used to calculate the PCF from Equation 1.27. The process, on average, results in using between the 50^{th} and 700^{th} nearest neighbor (the k^{th} value), depending on the spatial variation (inter–arrival times) of the data (droplets).

Although briefly mentioned in Chapter 1, one of the main goals of this thesis is to compare

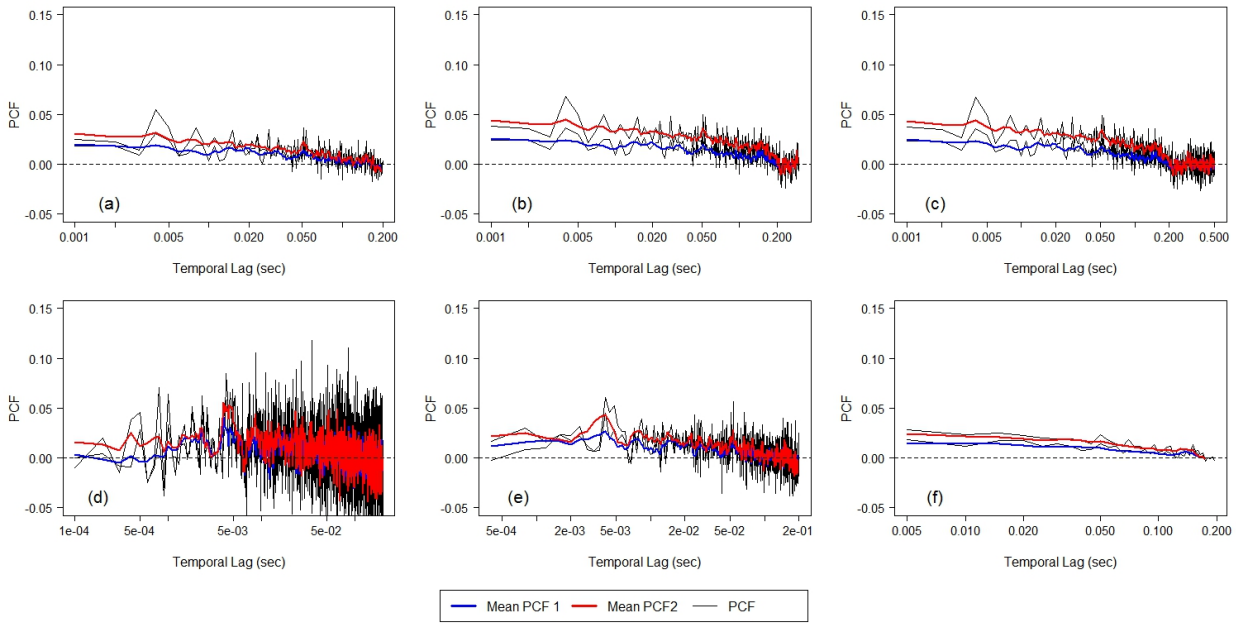

Figure 2.1: Each panel shows two sets of PCF functions for different data sets (subsets of data used in Figures 1.4–1.6). The mean PCF 1 is in blue and the mean PCF 2 is in red, representing a rolling mean of 5. Panels (a), (b), and (c) have a varying t.max of 0.2, 0.3, and 0.5 seconds respectively. Panels (d), (e), and (f) have a varying dt of 0.0001, 0.0004, and 0.005 seconds, respectively.

clustering, and therefore the PCF value, of the cloud entrainment region vs. the cloud center region. It is important to understand therefore, if, or how, the PCF changes as you change the values of dt and t.max and if changing these values affects being able to compare two different PCF values. Figure 2.1 shows how the PCF changes as the values of dt and t.max change. Note that the mean PCF 1 is blue and the mean PCF 2 is red, with a rolling mean of five (where the graphs were created using subsets of the same data as was used for Figures 1.4–1.6). The top row of Figure 2.1 (panels (a), (b), and (c)) shows PCF 1 and PCF 2 with a constant dt of 0.001 seconds and a varying t.max of 0.2, 0.3, and 0.5 seconds for panels (a), (b), and (c), respectively. The main takeaway from the top row is the fact that PCF 2 is always larger than PCF 1. The average difference between the endpoints of PCF 1 and PCF 2 is 0.015 with a standard deviation of 0.0043. The maximum PCF value for both PCF 1 and PCF 2 tends to increase as t.max increases, due to the fact that as you extend the time length of the x–axis, the PCF decreases more, and in normalizing the curves this results in a higher average PCF value. The second row of Figure 2.1 (panels (d), (e), and (f)) shows the two PCF curves with a constant t.max of 0.02 seconds and a varying dt of

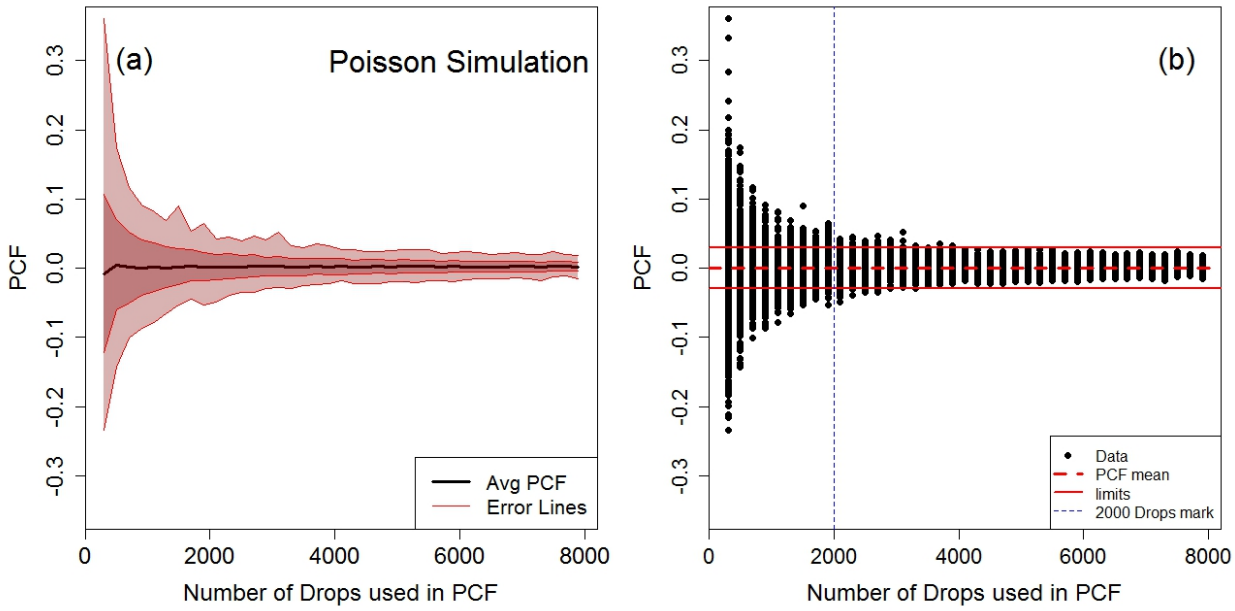

Figure 2.2: Shows 500 Poisson simulations for each variation of the number of droplets used. Panel (a) shows the number of Drops used in the PCF vs. the PCF value. The dark red envelope represents the 90 percent confidence interval and the light red envelope represents the maximum and minimum PCF values. Panel (b): as in panel (a), except each PCF value has now been transformed into a point. The solid red lines represent limits used to make statistical calculations (see text).

0.0001, 0.0004, and 0.005 seconds for panels (d), (e), and (f), respectively. Again, the main takeaway is that PCF 2 is always larger than PCF 1. It can now be seen that having an extremely small dt (panel (d)) results in a noisy PCF, while having too large of a dt (panel (f)) results in unnecessary scale averaging which results in excessively smoothing the PCF curves. Choosing the correct dt is critical, as choosing a dt that is too small can result in noise that may lead to uncertain results, as seen in panel (d). The average difference between the two PCF curves is 0.0098 with a standard deviation of 0.0049.

It is also important to understand how the PCF can change with the amount of data that is used. The longer the data series the more statistically significant the calculation of $\eta(t)$ will become (see Section 1.4). A data–set of only a few hundred droplets can become an issue for a second order counting statistical tool such as the PCF. In order to see how the average maximum PCF value (the average PCF of the first eight PCF values, corresponding to the first eight x–axis values) deviates as the number of droplets used in the calculation changes, 500 Poisson simulations were ran for each variation in the number of drops (see

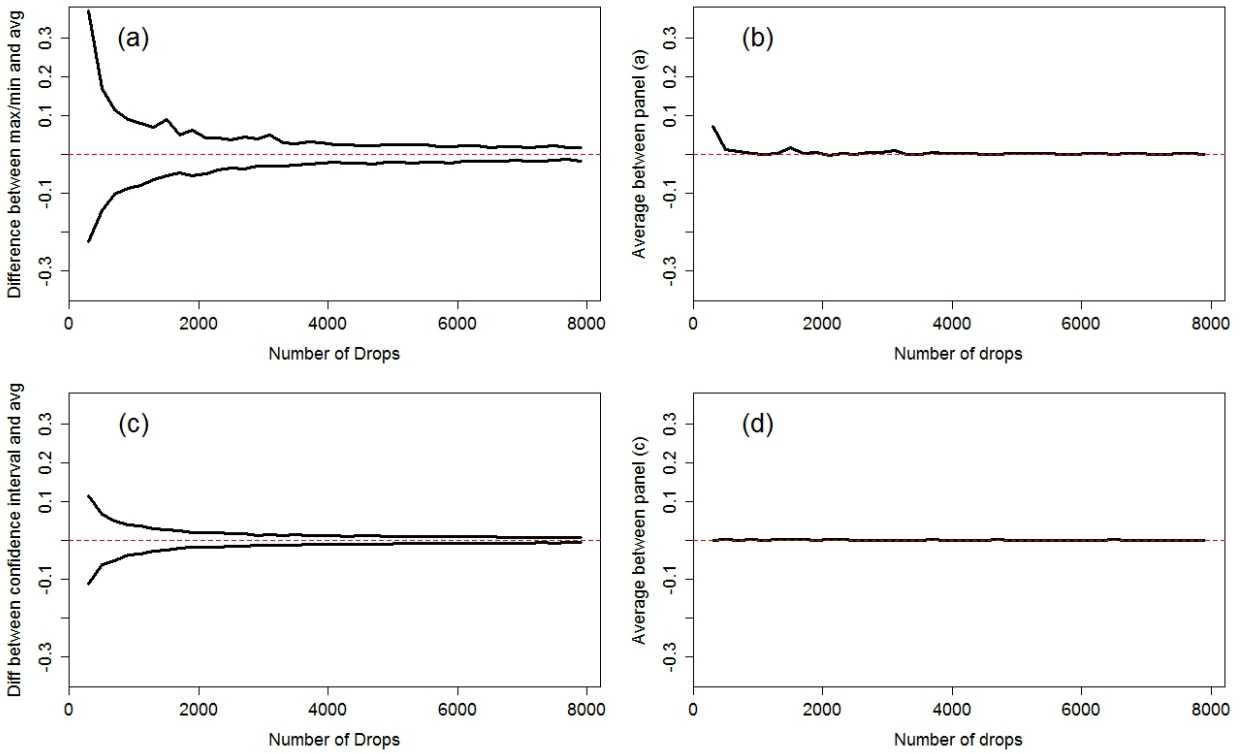

Figure 2.3: Panel (a) shows the difference between the maximum PCF and the average (top line) and the minimum PCF and the average (bottom line). Panel (b) shows the average between the two lines in panel (a). Panels (c) and (d): as in panels (a) and (b), except for the 90 percent confidence interval instead of the maximum and minimum.

Figure 2.2). The number of drops used ranges from 300 to 8000 moving at a sequence of 200 (i.e. (300, 500, 700,..., 7700, 8000)). Panel (a) in Figure 2.2 shows the average PCF value in black (zero except at the lower limit of the x–axis). The dark red envelope represents the 90 percent confidence interval while the light red envelope represents the maximum and minimum PCF values. Panel (b) gives a similar representation except each PCF value is graphed as a black dot. It can be seen that almost all PCF values lie between the two red lines located at y = 0.03 and -0.03 above 2000 drops used in the PCF. More variation is clear below 2000 drops used in the PCF. The percent of PCF values that lie outside the red lines for the whole plot is 7.86 percent, whereas that percent increases to 32.4 percent when looking at the data below 2000 drops used. After a visual examination of Figure 2.2 one would think that the variation present is symmetric. The percent of data that is below zero for drops used less than 2000 is 48.8 percent and 51.2 percent for data above zero.

Figure 2.3 proves that the variation is indeed symmetric. Panel (a) shows the difference

between the maximum PCF and the average (top line) and the minimum PCF and the average (bottom line). Panel (b) shows the average between the two lines in panel (a). Panels (c) and (d) represent the same, except for the 90 percent confidence interval instead of the maximum and minimum. From looking at panel (b), one can see that the average between the two lines in panel (a) is zero, again except for the lower limits of the x–axis where the smallest number of drops used are located. Looking at the average of the confidence intervals in panel (d), it is clear that the result is zero for the entire x–axis. This shows that, for at least a 90 percent confidence, that the variation seen in the PCF at a lower number of drops used is symmetric. One should still be cautious when interpreting PCF values for real data when the number of drops used is small (< 1000 drops), especially if one is only doing a single calculation. However, if multiple calculation are used, one can assume that the variation for smaller data series will be symmetric in terms of variation and result in PCF values that can be trusted when averaging over all of the calculation made. Examples of this will be given in the results section (see Section 3.8).

2.4 Data Stationarity

Data stationarity has been mentioned throughout Chapters 1 and 2, and a number of studies have carefully noted that the PCF can only be interpreted as a scale–localized deviation from perfect randomness for a statistically stationary/homogeneous system (see, e.g., Shaw et al. (2002); Larsen and Kostinski (2005); Kostinski et al. (2006); Larsen (2007, 2006)). A difficulty arises, however, when one needs to choose which statistical tools to use/apply to data collected in the field. If the PCF and related statistical tools can only be meaningfully used on data that is statistically stationary, then the first step would be to determine whether the data gathered is statistically homogeneous. Unfortunately, no unambiguous tool for determining statistical homogeneity for real data exists (Larsen 2012). There have been some efforts to determine practical conditions for statistical stationarity (applied to the statistical tools themselves), such as in Larsen and Kostinski (2005) and Anderson and Kostinski (2010), which will be discussed in the following paragraph.

Following appendix B of Larsen and Kostinski (2005), the following two quantities may help identify if the PCF was calculated in a situation where it shouldn't have been. These are not sufficient test for stationarity, but are necessary conditions which must be met. Analysis based on the PCF requires that the expected number of drops detected in the time interval dt be small enough so the probability of finding two or more drops in dt is negligible

(See Section 1.3.1 for assumptions of spatial distribution randomness). To appreciate the strictness of this condition, the Poisson process can be examined. The mean count, $\mu = \bar{N}dt$ should be much less than unity (Larsen and Kostinski 2005). Noting that the mean count is $\frac{Ndt}{T}$, where the number of drops is N and T is the time length of the time series, it can be stated that a requirement for computation of a meaningful PCF should be

$$\frac{N^2 dt}{T} = N\mu \gg 1 \tag{2.22}$$

and

$$\left(\frac{Ndt}{T}\right)^2 = \mu^2 \ll 1 \tag{2.23}$$

(Larsen and Kostinski 2005). The results of the PCF should be viewed as suspect if μ^2 is not substantially less than one and/or if $N\mu$ is not much larger than 1. These two equations signify that the PCF is most useful in the limit as $\mu^2 \to 0$ and $N\mu \to \infty$ and help to specify bounds for what values of dt ought to be chosen for the data at hand (Larsen and Kostinski 2005). Although every statistically homogeneous/stationary system will pass these tests, some statistically inhomogeneous/nonstationary systems will pass these tests as well. These tests form necessary, but not sufficient criteria for statistical stationarity (Larsen 2012). The difficulty in finding truly necessary and sufficient tests of stationarity for real data has proven tricky, at least in part because real data is finite and, therefore, formally statistically inhomogeneous/nonstationary even if the underlying system is inherently homogeneous/stationary. This serves as a reminder that no tool can unambiguously show when a measured dataset comes from a statistically stationary system.

To get a better understanding of μ^2 and $N\mu$ we can examine the values obtained for PCF 1 in Figure 2.1. Panels (a), (b), and (c) all have the same values of μ^2=0.76 and $N\mu$=19785. These values make sense since panels (a), (b), and (c) have a constant dt value and μ^2 and $N\mu$ only depend on dt. Although $N\mu \gg 1$, μ^2 is not $\ll$1. This signifies that the dt of 0.001 seconds that was used was not a reasonable choice and a smaller dt should have been chosen. Panels (d), (e), and (f) have $N\mu$ values of 1978, 7914, and 98926, respectively. The panels (same order) have μ^2 values of .0076, 0.121, and 19.02. Examining these results, one can tell that you get more sound values of μ^2 and $N\mu$ for smaller values of dt. However, having a value of dt that is too small results in a noisy PCF function (and smaller values of $N\mu$). A compromise must be made between a noisy PCF and more reasonable values of μ^2 ($<$0.1) and $N\mu$ ($>$100).

2.5 Phase Doppler Interferometry

The Phase Doppler Interferometer (PDI) used in this research is the Artium Flight–PDI (F/PDI), designed and constructed by Artium Technologies, Inc. of Sunnyvale, CA. PDI is a laser–optical method capable of measuring velocity and diameter of individual, spherical particles as they traverse the instrument's measurement volume. A description of the PDI's general working principle will be described, followed by PDI measurement accuracy and an explanation of how this technique can be used for studying particle clustering in turbulence.

2.5.1 Basic PDI Working Principles

The measurement principle of the PDI is based on light scattering interferometry, which utilizes the wavelength of light as the measurement scale. Two important properties that particles must satisfy in order for the PDI to be a successful technique include: the droplets must be optically homogeneous at a length scale small relative to the size of the drop with a known refractive index, and they must be close to spherical (Chuang et al. 2008). These conditions are satisfied for liquid drops under atmospheric conditions for a diameter range of 2 μm to 1 mm. Once droplets grow larger than 1 mm, the drop shape becomes more "bun" shaped than spherical due to atmospheric aerodynamic drag forces, limiting the PDI range. According to Pruppacher and Klett (1997), droplets smaller than 300 μm are nearly perfect spheres, and a maximum uncertainty of 2 percent and 10 percent due to asphericity may be expected for drops of 1 mm and 2 mm, respectively. The limit at smaller droplet sizes is due to sizing ambiguities as well as signal to noise ratio (SNR)(Chuang et al. 2008).

Figure 2.4 from Chuang et al. (2008) shows the typical setup of a PDI system. Two laser beams which have the same frequency (derived from a single polarized laser which travels through a beam splitter) crossed at a known angle (γ) establish the measurement volume which takes on the approximate shape of a cylinder at a small γ, with dimensions D_{beam} and $L_{aperture}$. The scattered light is collected, spatially filtered using an aperture, and then imaged onto a set of three detectors labeled A, B, and C. The detector signals are then processed to produce individual drop size and velocity. The receiver is located at an angle θ from the transmitter optics. According to Chuang et al. (2008), in the intersection volume, the cross section of intensity has two components: (1) a low–frequency Gaussian profile that results from the Gaussian profile of each of the individual identical beams; and (2) a high frequency sinusoidal pattern that results from constructive and destructive interference of these two beams (Figure 2.5 from Chuang et al. (2008)). It is these two components that

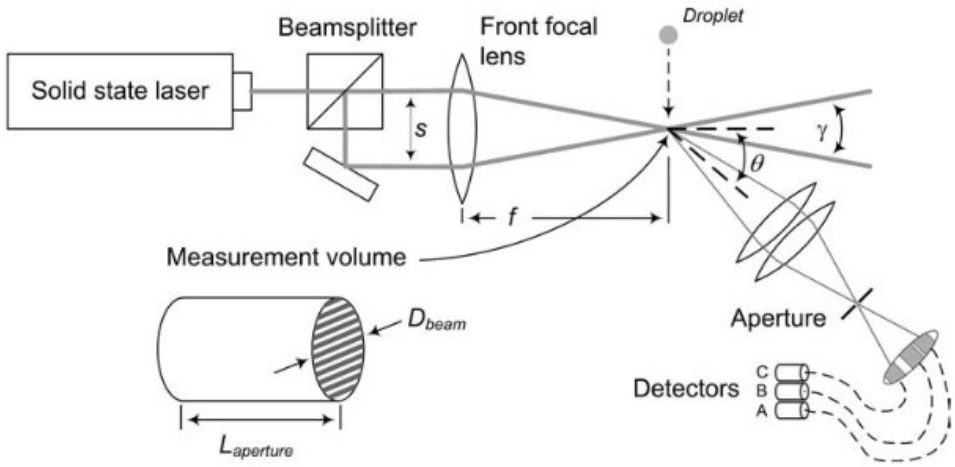

Figure 2.4: Schematic showing the basic operating principle of the PDI. A single laser is split by the beam splitter into two equal–intensity beams. The two beams are brought together at angle γ. The measurement volume is the intersection of the two beams. The scattered light is then imaged onto the three detectors A, B, and C (Figure 2 in Chuang et al. (2008)).

lead to the measurements of drop velocity and size.

2.5.2 Velocity Measurement

When a particle traverses the measurement volume in the forward direction, it scatters light projecting the intensity profile of the measurement volume. One or more light detectors collect the scattered light and turn it into an electrical signal. A typical temporal evolution of the observed signal is depicted in Figure 2.5 and is commonly termed a "Doppler burst", which represents the raw data collected. Figure 2.5 shows the Doppler burst (Doppler frequency f_d) received by two detectors. Frequency of either of the signals yields drop velocity. The frequency of each Doppler burst is directly proportional to the magnitude of the particle's velocity:

$$u = \delta f_d \tag{2.24}$$

where u is the component of the droplet velocity in the x–direction (Chuang et al. 2008). δ is the fringe spacing of the intensity profile in the measurement volume satisfying the relation:

$$\delta = \frac{\lambda}{2sin(\frac{\gamma}{2})} = \frac{\lambda F}{s} \tag{2.25}$$

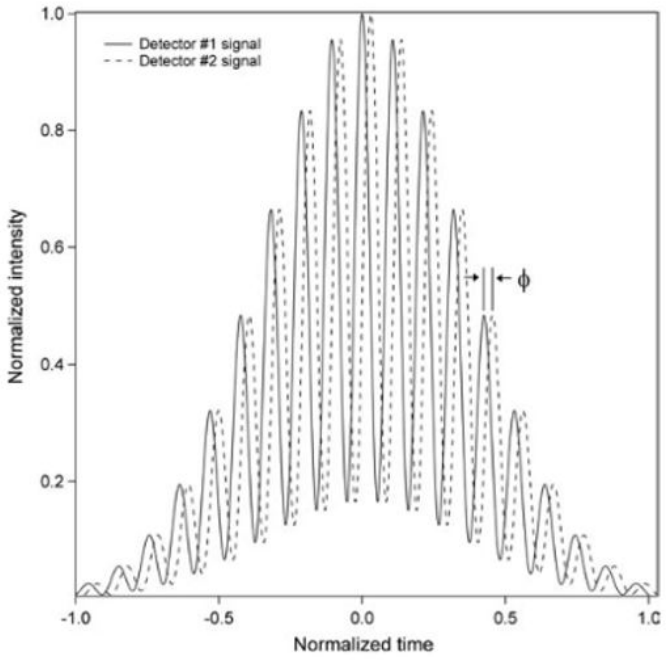

Figure 2.5: Schematic of the raw data collected through the PDI. Represents an idealized photodetector signal received by two detectors. Primary characteristics are Gaussian in intensity with a superimposed sinusoid frequency. θ is the phase shift between the two signals (Figure 3 in Chuang et al. (2008).

where λ is the wavelength of the laser, F is the focal length of the front focusing lens, and s is the beam separation before reaching the focusing lens.

2.5.3 Drop Size Measurement

The second property of the two signals is the phase shift ϕ between them (Figure 2.5), which has been shown to have a nearly monotonic, linear relationship with droplet diameter (Bachalo and Houser 1984; Albrecht et al. 2002). To understand this linear relationship, one can interpret the function of a detected particle as an optical lens (or mirror) that images the intensity fringes in the measurement volume to the surrounding space. The image of the fringes is then picked up by the detectors as depicted in Figure 2.6 from Chuang et al. (2008), which shows an instantaneous snapshot of how the fringes are projected in space by the particle. The magnification of the fringes scales inversely as the particle size, since particle size determines the curvature of the lens (the smaller the droplet, the more expanded the projected fringes will be because the radius of curvature of the lens is smaller).

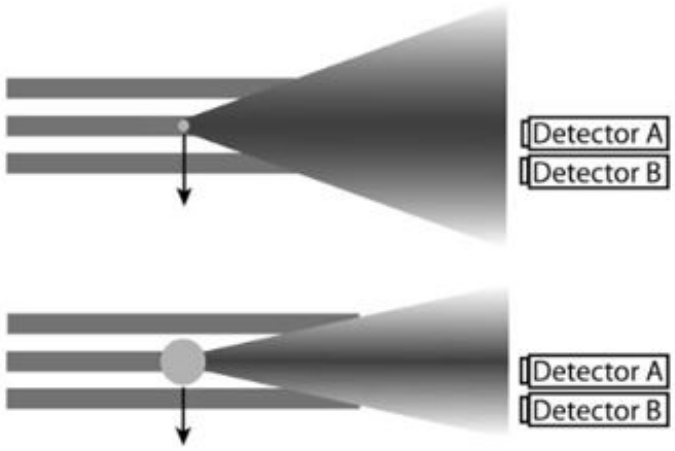

Figure 2.6: Shows a schematic illustrating the origin of the phase shift versus diameter relationship. A small drop (top) yields a small phase shift, whereas the larger drop (bottom) yields a larger phase shift (Figure 4 in Chuang et al. (2008).

Multiple detectors placed at different locations effectively measures the spatial frequency of the magnified fringes by registering different phases of the fringes. The phase shift between the Doppler burst measured by any detector pair is then monotonically related to the size of the particle. The exact relation between particle diameter and phase difference depends on the scattering mode considered (reflection, 1^{st} order refraction, etc.) and the physical setup of the instrument (Albrecht et al. 2002).

In practice, the detectors are placed at scattering angles that favor only one scattering mode. When this is done, it can be shown that the diameter–phase relation for any detector pair is linear by:

$$d_p = F_\phi(\lambda, \gamma, \phi, \theta_1, \theta_2)\Delta\phi \tag{2.26}$$

from Albrecht et al. (2002). $\Delta\phi$ is the phase difference, F_ϕ is the phase conversion factor, which is a function of laser wavelength (λ) and optical geometry of the system which includes beam–cross angle (γ) and the elevation angles of the detectors (θ_i), measured relative to the major plane.

2.5.4 PDI Accuracy

- **Sizing Accuracy**

The drop size measurement range is 0.5 to 2000 μm (obtained from the brochure published by the manufacturer, Artium) with an estimated accuracy of $\pm$ 0.5 μm. The F/PDI is also excellent at resolving particle coincidence. If two droplets of very different size are coincident, the most likely scenario is one where the signal from the large drop overwhelms that from the small drop, and the large drop will be detected while the small drop is omitted. This can result in under–counting of the most common drops in a cloud, but the larger, rarer drops will be counted accurately (Chuang et al. 2008). Other instrumentation can mistake two small coincident drops for a single larger drop, which cannot happen with the F/PDI. Note that when a drop whose size is comparable to or larger than the beam diameter passes through the view volume, significant errors in sizing can arise.

- **Velocity Accuracy**

  The velocity measurement range is from -100 to 200 m/s with an estimated accuracy of $\pm$ 0.2 percent. The accuracy depends on accurately deriving Equation 2.25. From Equation 2.25, λ and F are quantities that are known to a very high accuracy and are also essentially constant under normal operation (Chuang et al. 2008). Therefore, the uncertainty in δ depends on the ability to properly measure s. Normally, s is measured by directing the transmitter beams onto a distant wall. By measuring the distance to the wall, and the beam separation at this distance, s can be calculated to within less than 1 percent.

- **Arrival Time Accuracy**

  Especially important for this research is the accuracy of droplet arrival time or detection time which is typically not reported by manufacturers. It was found that the droplet arrival time can accurately be measured to < 3.5 μs from Saw (2008). Assuming that an aircraft is moving at roughly 60 m/s, this allows one to accurately map droplets down to $2.1 \cdot 10^{-4}$ m. Note that due to modern advances in electronics, the F/PDI has no dead time (time after each event during which the system is not able to record another event).

2.5.5 Observing Particle Clustering with PDI

The PDI method is suitable for studies of particle clustering when the probe is oriented such that there is a strong mean flow in the x–direction. Examples of this suitable situation

include when the probe is placed inside a wind tunnel, turbulent or otherwise, and when the probe is flown through atmospheric clouds (Albrecht et al. 2002). To be more precise, situations where the mean flow velocity is much stronger than the average deviation of particle velocities from the mean flow is of interest. Such deviations may be caused by the presence of turbulence or other forces like gravity. Under this condition, one can calculate inter–particle distances in the x–axis from the arrival times of each particle. This essentially contains information on the one dimensional distribution of the particles in space, from which further analysis can be made.

2.6 Flight Information

The GoMACCS campaign was conducted jointly with the 2006 Texas Air Quality Study (TexAQS) during August and September of 2006 as a combined climate change and air quality intensive field campaign. During the GoMACCS campaign, the CIRPAS Twin Otter (flight speed of about 60 m/s) performed 22 research flights to explore aerosol–cloud relationships over the Houston region and the northwestern Gulf of Mexico (Lu et al. 2008). Among the 22 research flight conducted, 14 intensive cloud measurements were carried out (where the clouds were all continental warm cumulus clouds subject to various levels of anthropogenic influence), including one flight in which an isolated cumulus cloud of sufficient size and lifetime existed to allow detailed sampling at different altitudes. The other 13 cases involved scattered cumuli that were sampled in such a manner as to provide statistical properties over the cloud field (Lu et al. 2008). Table 2.1 from Lu et al. (2008) shows the instrument payload that was on board the CIRAS Twin Otter. Note that only the PDI instrument was/will be talked about in this thesis. For further information on other instrumentation, see the references in the table (i.e., Chuang et al. (2008); Sorooshian et al. (2006); Pilewskie et al. (2003); Gerber et al. (1994); Roberts and Nenes (2005)).

Table 2.2 shows each flight conducted, with the corresponding Research Flight (RF) number, date, number of clouds in the flight (after filtering, including clouds that are only $>$ 300 m in length and non–precipitating), the total aerosol number concentration (N_a), and the aerosol number concentration (N_{acc}). N_{acc} includes aerosols that are only in the accumulation mode (0.1 μm $<$ particle size $<$ 2.5 μm). N_a was measured by the condensation particle counter (CPC) and N_{acc} was measured by the passive cavity aerosol spectrometer probe (PCASP). Note that values in parenthesis from Table 2.2 represent standard deviation.

Table 2.1: Shows the instrument payload on board the CIRPAS Twin Otter during the GoMACCS campaign. For more information on the instrumentation, see the imbedded references (Table 1 in Lu et al. (2008)).

Instrument	Data
Condensation particle counter (CPC)	aerosol number concentration (CPC 3025, >3nm; CPC 3010, >10 nm, dry)
Counterflow virtual impactor (CVI)	virtual impactor for isolating cloud droplets (cut size of 10 μm $\pm$ 20 percent
Dual automated classified aerosol detector (DACAD)	submicrometer aerosol size distribution (10.3–828 nm, 85 bins) at two RHs (dry and wet)
Passive cavity aerosol spectrometer probe (PCASP), forward scattering spectrometer probe (FSSP), cloud/aerosol/precipitation spectrometer (CASP consists of cloud and aerosol spectrometer (CAS) and cloud imaging probe (CIP)), phase Doppler interferometer (PDI [Chuang et al., 2008])	aerosol/cloud/precipitation droplet size distribution: PCASP (0.1–2.6 μm, dry), FSSP (2.55–42.7 μm, 19 bings), CAS (0.54–62 μm, 20 bins), CIP (40.45 μm–1.56 mm, 61 bins), PDI (2.07–200 μm, 127 bins)
Time–of–flight Aerodyne aerosol mass spectrometer (TOF-AMS)	nonrefractory aerosol chemistry
Particle–into–liquid sampler/ion chromatography (PILS–IC [Sorooshian et al., 2006])	submicrometer water–soluble aerosol chemistry (inorganic and organic acid ions)
Solar spectral flux radiometer (SSFR [Pilewskie et al., 2003])	upwelling and downwelling radiative fluxes above and below cloud soot absorption (multiwavelength/incandescence)
Particle soot absorption photometer (PSAP), photoacoustic spectrometer, CO_2	soot absorption (multiwavelength/incandescence)
Navigational/meteorology probes and Gerber liquid water content probe (PVM–100A [Gerber et al., 1994]), hotwire LWC probe	navigational data, temperature, dew point, RH, pressure, liquid water content, wind direction/speed, updraft velocity, etc.
Cloud condensation nuclei counter (CNN [Roberts and Nenes, 2005])	CCN at different supersaturations
Multiangle light scattering spectrometer (MLS)	shape and angular dependence of light scattering of individual particles

Table 2.2: Shows the flight information for 20 of the 22 flights that occurred during the GoMACCS campaign. Each flight corresponds to a RF number, date, the number of clouds (after filtering, see text), the total aerosol number concentration (N_a), and the aerosol number concentration (N_{acc}). Values in parentheses represent the standard deviation. Blue rows indicate low pollution flights, orange rows indicate high pollution flights, and the red row represents the Case flight.

Flight	RF Number	Date	Clouds	N_a cm^{-3}	N_{acc} cm^{-3}
Flight 1	1	8/21/06	1	NA	NA
Flight 2	2	8/22/06	11	NA	NA
Flight 3	3	8/23/06	9	2984 (588)	413 (66)
Flight 4	4	8/25/06	17	22667 (10672)	1797 (4184)
Flight 5	5	8/26/06	18	1409 (762)	310 (216)
Flight 6	6	8/27/06	0	NA	NA
Flight 7	7	8/28/06	0	NA	NA
Flight 8	9	8/29/06	28	3157 (1980)	360 (169)
Flight 9	11	8/31/06	37	3205 (650)	972 (214)
Flight 10	12	9/2/06	29	4768 (2826)	710 (169)
Flight 11	13	9/3/06	0	NA	NA
Flight 12	14	9/4/06	14	2770 (758)	697 (126)
Flight 13	15	9/6/06	10	1427 (398)	465 (147)
Flight 14	16	9/7/06	32	3547 (966)	949 (306)
Flight 15	17	9/8/06	21	4824 (1806)	821 (154)
Flight 16	18	9/10/06	69	1940 (2125)	1280 (4172)
Flight 17	19	9/11/06	27	6561 (5419)	1280 (4172)
Flight 18	20	9/13/06	0	NA	NA
Flight 19	21	9/14/06	25	2230 (1157)	653 (276)
Flight 20	22	9/15/06	13	2361 (613)	442 (127)

Following the methods in Small et al. (2013), two low (blue rows in Table 2.2) and two high (orange rows in Table 2.2) pollution flights were selected to compare the impacts that aerosols may have on droplet clustering and turbulence. A Case flight was also selected (red row in Table 2.2) to analyze how droplet clustering behaves as a function of altitude. Flights 5 and 20 were selected for the low pollution flights (L1 and L2), while flights 10 and 15 were selected for the high pollution flights (H1 and H2). Note that parts of flights 2 and 13 occurred over the Gulf of Mexico, resulting in marine cumulus clouds being measured. To keep all cumulus clouds continental, flights 2 and 13 were omitted from being a possible selection for low or high pollution flights.

For each cloud encountered, the first two seconds and last two seconds of each cloud were

used to collect droplet data for the entrainment zone, while the middle two seconds of each cloud were used for cloud center data. This results in twice as much data for entrainment zones as compared to center zones. Two, two second sections in the middle of the cloud would have been used but some clouds did not contain eight complete seconds of data. Two seconds were selected due to the turbulent nature of cumulus clouds. When stronger entrainment and mixing processes occur, stationary data is usually only fulfilled for subsections of the cloud. However, the longer the analyzed cloud section, the more statistically significant the calculation of the PCF will become due to larger droplet populations (as was discussed in Section 1.4 and can be seen in Section 2.3.2). Two seconds was found to be a good compromise between droplet populations and data stationarity.

2.7 Objectives and Hypotheses

The PDI and other instruments briefly mentioned in Section 2.5 offer the unique opportunity to gain insight into droplet clustering and turbulence described in Chapter 1 and to answer some of the open questions regarding the processes associated with warm clouds. The spatial distribution of cloud droplets is measured with the F/PDI, and deviations of the spatial distribution of droplets from pure randomness are quantified with the PCF. The data are then analyzed to infer possible dependencies of the droplet clustering quantified with the PCF and the turbulence of the air–flow. The focus will be on the following questions regarding the response that aerosols have on droplet clustering in shallow, warm continental cumulus clouds: (1) Does droplet clustering depend on aerosol number concentration? (2) Does droplet clustering change as a function of cloud center vs. cloud entrainment zone? (3) Does droplet clustering change as a function of altitude (normalized cloud height)? Based on the statement from Shaw (2003) that one of the main sources of TKE in clouds is shear evaporative cooling due to entrainment of dry air at cloud edge and cloud top; it is hypothesized that droplet clustering will be enhanced at cloud edge as compared to the middle of the cloud. Droplet clustering is therefore also expected to be enhanced at cloud top vs. the rest of the vertical profile of the cloud. There has been no research (to the best of the author's knowledge) on how droplet clustering will be affected by aerosol number concentration. One would think however that if given a higher number of drops (larger amount of aerosols), that the clustering would be greater due to the simple fact that there are more droplets available for clustering. These questions and hypotheses will be analyzed in Chapter 3.

CHAPTER 3
RESULTS AND DISCUSSION

3.1 Framework for Chapter 3

There is no shortage of research that has been conducted in the lab focusing on droplet clustering and the role that turbulence plays. Advances have been made in the statistical tools needed to measure said clustering, but the actual measurement approaches remain confined to cloud chambers and wind tunnels which use Reynolds numbers that are smaller than average Reynolds numbers experienced in the atmosphere (see Section 2.2.4). This chapter will look at data collected during the GoMACCS campaign using the PDI which was onboard the CIRPAS Twin Otter aircraft to analyze the amount of clustering that was occurring in shallow cumulus clouds that were exposed to different levels of anthropogenic pollution. Using the mathematical tools and concepts discussed throughout the first two chapters, answers to the following questions will be given:

- Does droplet clustering depend on the aerosol number concentration?
- Does clustering change as a function of height?
- Is there a difference in clustering between cloud center and cloud edge?

The environmental and flight characteristics for the selected days (L1, L2, H1, and H2) will be examined, followed by analysis conducted using the PCF to answer the aforementioned questions. Thereafter, possible errors in the data analysis and methods will be discussed.

3.2 Flight and Environmental Characteristics

3.2.1 Flight Path

During the GoMACCS campaign, 22 flights were conducted around the Houston, TX area. Figure 3.1 shows the location of the L1 flight on the left and the L2 flight on the right. Analyzing the top panels first, the "n=" represents the number of clouds that were sampled (after filtering), the blue line represents the flight altitude while the black lines represent the

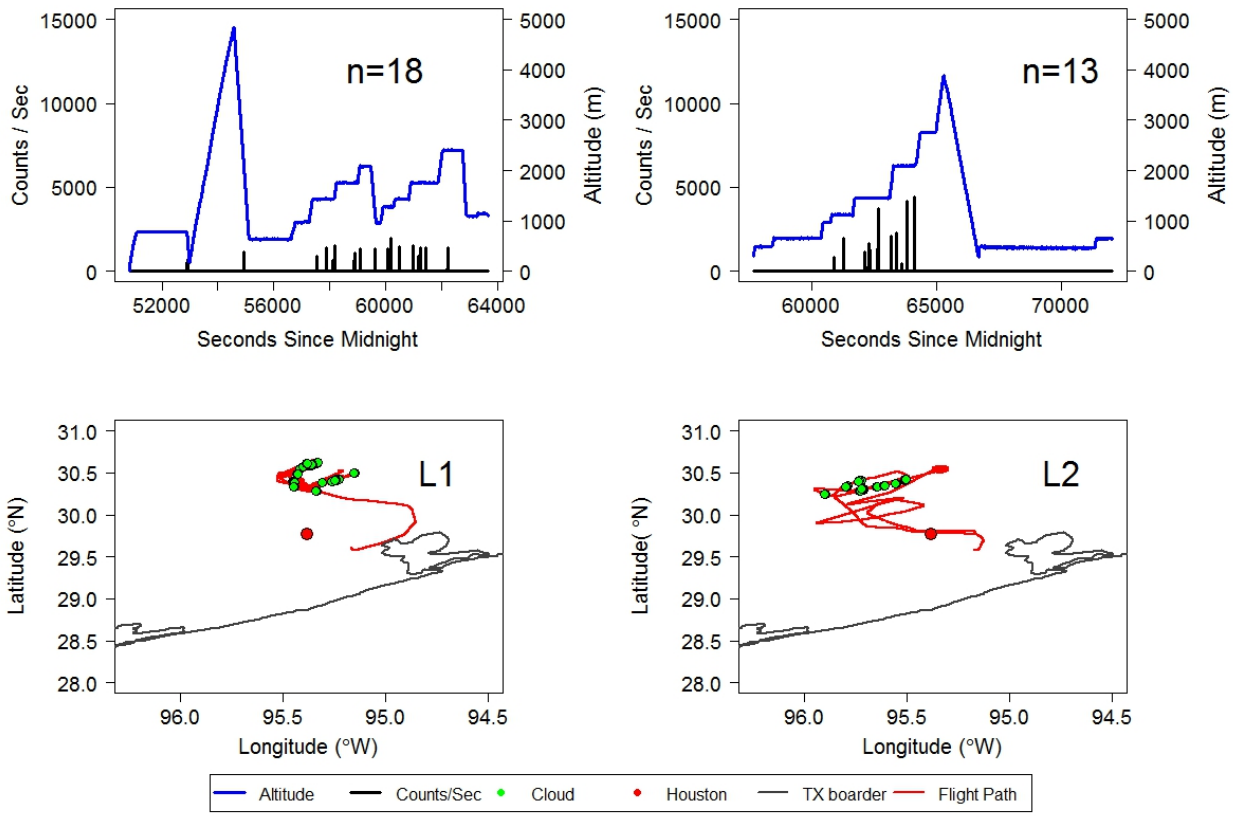

Figure 3.1: Shows L1 and L2 on the left and right, respectively. Flight altitude (blue) as a function of time is displayed in the top panels, with counts per second in black. The "n=" represents the number of clouds sampled for each flight after filtering. The bottom panels show the Texas coast in grey with the location of Houston represented by the red dot. The flight path is outlined in red with the location of clouds displayed by green dots.

cloud droplets per second. The number of clouds sampled for L1 and L2 were 18 and 13 clouds, respectively, with an average counts encountered per second of 660 and 1016 droplets (ignoring out of cloud sampling). Analyzing the bottom panel, the grey line represents the coast of Texas, the red dot represents the location of Houston (the main pollution source), the red line represents the flight path, and the green dots represent the location of every cloud encountered (again, after filtering). Figure 3.2 gives the same information as in Figure 3.1, except for flights H1 and H2, shown on the left and right, respectively. The average counts (clouds) encountered per second for H1 and H2 were 958 (29) and 3300 (21), respectively. Since H1 and H2 represent high pollution clouds, the higher amount of droplet encounters is expected. More aerosols (which often act as Cloud Condensation Nuclei (CNN)), for a given LWC, implies more, smaller droplets (Koren et al. 2015). The key difference between

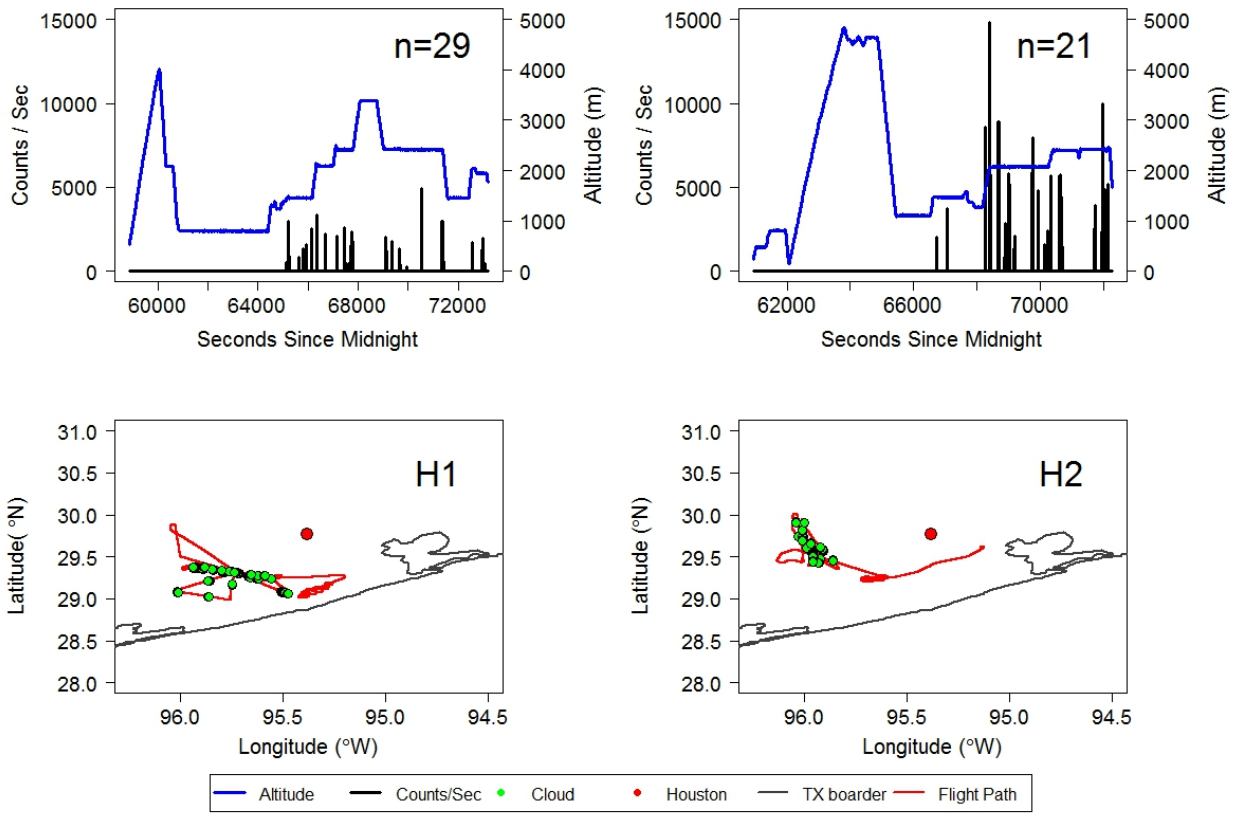

Figure 3.2: As in Figure 3.1, but for H1 and H2.

Figure 3.1 and 3.2 is the location at which the low and high pollution clouds were sampled. Low pollution clouds were sampled to the North of Houston, whereas the high pollution clouds were sampled to the Southwest (H1) and West (H2) of Houston. These locations agree with the dominant wind direction at the time of the flights (not shown here). Using archived wind data from the NOAA National Center for Environmental Information, the average surface wind speed on the days the flights took place was 4.4 mph moving from the East/Northeast to the West/Southwest. This resulted in the pollution being carried downwind to the West/Southwest of Houston, where the high pollution clouds were sampled. The low pollution clouds remained upwind from Houston. These results are also confirmed using HYSPLIT Trajectories from the Air Resources Laboratory (ARL) (Stein et al. 2015), which are not shown here. The entirety of each of the flight paths took place over land, excluding any possibility that maritime cumulus clouds could have been included in the analysis.

3.2.2 Cloud Characteristics

Table 3.1 shows variables highlighting the different characteristics of the clouds and environmental conditions within each flight, with the standard deviation of select variables represented in parentheses. Table 3.2 gives a summary of average values for low (L1 and L2) and high (H1 and H2) pollution cases for select properties from Table 3.1. Flights occurred between 1 pm and 3 pm Central Daylight Time (CDT) for H1 and H2, while the flights occurred slightly earlier at 9 am CDT for L1 and at 12 pm CDT for L2. The time that the flights took place has an important impact on atmospheric stability. Maximum daily temperatures are normally reached in the late afternoon (Ahrens, 2008), which can make the atmosphere more unstable than when surface temperatures are cooler during the morning hours. An unstable atmosphere can lead to more vigorous updrafts and stronger, more vertically developed clouds. This can make comparing clouds difficult (clouds that are similar in characteristics are ideal for comparing). Fortunately, as will be shown in Section 3.2.3, the environmental lapse rate on each of the flight days were similar to one another. Important characteristics to point out from Tables 3.1 and 3.2 include:

- Cloud width was lowest for clouds encountered during L2 at 520 m. L1, H1, and H2 had cloud widths of 700, 850, and 861 m, respectively. Possible reasons for the smaller horizontal extend of clouds in L2 will be given in Section 3.6.1.

- L2 had the lowest average in-cloud vertical velocity at 1.16 m/s. L1, H1, and H2 had in–cloud vertical velocities of 1.81, 2.34, and 1.61 m/s, respectively. It will be shown in Section 3.6.1 that these values are statistically different from one another.

- L1 and L2 had an average CDNC of 264 cm^{-3} while H1 and H2 had an average of 476 cm^{-3}. The ratio of CDNC for high vs. low is 1.80, which again, is expected since the clouds are sorted by their aerosol number concentrations.

- The mean aerosol number concentration for the low pollution clouds was 1814 cm^{-3} while the aerosol number concentration for high pollution clouds was 4491 cm^{-3}. High pollution clouds had roughly 2.5 times more aerosols per cubic centimeter than the low pollution clouds.

Table 3.1: A summary of cloud and flight properties from the L1, L2, H1, and H2 flights, moving from left to right, respectively. Standard deviation values are represented in parentheses.

Variable	**L1**	**L2**	**H1**	**H2**
Data	2006–Aug–26	2006–Sept–15	2006–Sept–02	2006–Sept–06
Flight Number	RF 5–2	RF 22	RF 12	RF 17
UTC for Cloud Sampling	1447–1717	1654–1748	1806–2018	1832–2002
Clouds > 300 m in width	18	13	29	21
Min cloud base height (m)	672	1120	1457	1476
Max clout top height (m)	2412	2101	2463	2451
Cloud thickness (m)	1740	981	1007	976
Cloud width (m)	700 (235)	520 (110)	850 (404)	861 (451)
Mean true air speed (m/s)	61.2 (1.5)	59.9 (1.3)	62.7 (2.3)	63.0(2.3)
Mean CDNC (cm^{-3})	318 (163)	210 (141)	421 (255)	531(363)
Max CDNC (cm^{-3})	819	526	1059	1630
Mean drops (s^{-1})	661 (448)	1016 (1037)	958 (895)	3300 (2704)
Cloud top LWC (gm^{-3})	0.97 (0.66)	0.55 (0.54)	0.47 (0.48)	0.60 (0.48)
Mean vertical velocity (ms^{-1})	181 (1.67)	1.16 (1.28)	2.34 (2.21)	1.61 (1.62)
$N_{aerosol}$ (cm^{-3})	1304 (699)	2323 (688)	4331 (2789)	4650(2057)
N_{acc} (cm^{-3})	290 (223)	436 (159)	671 (255)	771 (283)

Table 3.2: Average values for low (L1, L2) and high (H1, H2) pollution clouds for select variables from Table 3.1

Variable	**Low**	**High**
Mean CDNC (cm^{-3})	264	476
Mean Drops (s^{-1})	839	2129
$N_{aerosol}$ (cm^{-3})	1814	4491
N_{acc} (cm^{-3})	363	721
Cloud thickness (m)	1361	992
Cloud width (m)	610	856
Clouds > 300 m in width	31 (total)	50 (total)

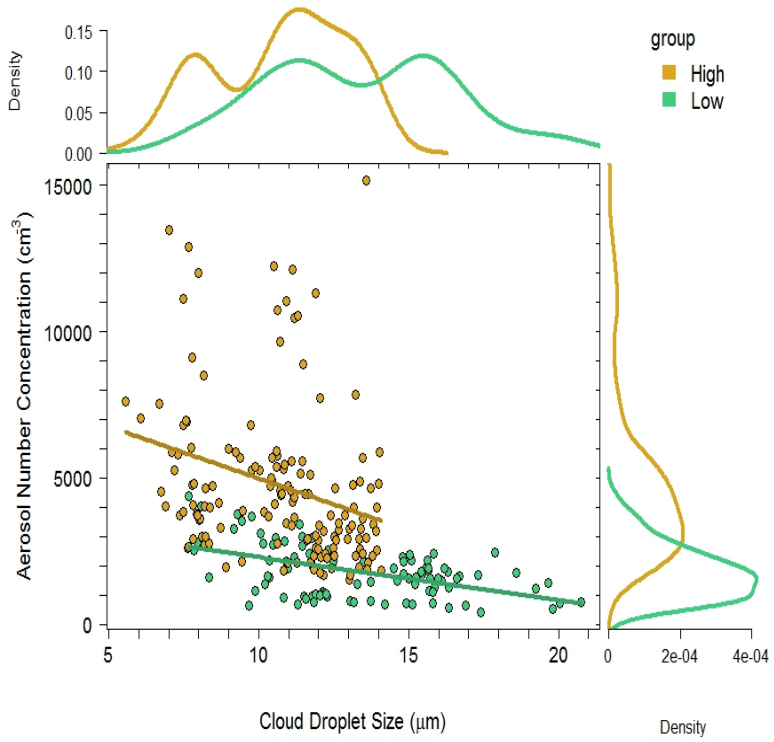

Figure 3.3: Shows cloud droplet size (μm) on the x–axis and aerosol number concentration (cm^{-3}) on the y–axis, with low pollution data in green and high pollution data in gold. The corresponding density curves of the high and low pollution data are given on the outer margins of the plot.

The difference in aerosol number concentration between the low and high pollution clouds produces clouds that are statistically different from one another using the Wilcoxon–rank–sum test (see Figure 3.3). The Wilcoxon–rank–sum test is ideal because it allows one to compare two unmatched random samples of measurements, such as two samples taken from different sources (Wilks 2011), in this case, two samples collected from two different clouds containing a different amount of cloud droplets. The null hypothesis of the Wilcoxon–rank–sum test is that two data samples have been drawn from the same distribution. If the two data samples have different distributions (the center of one sample is larger or smaller than another), than the null hypothesis is not true. Note that the Wilcoxon–rank–sum test tends to give slightly higher probability levels than more complicated alternatives (Mann and Whitney 1947). Figure 3.3 shows cloud droplet size in microns (μm) on the x–axis with aerosol number concentration (cm^{-3}) on the y–axis, with low pollution data in green and high pollution data in gold. Density curves are given to show how the data is distributed for

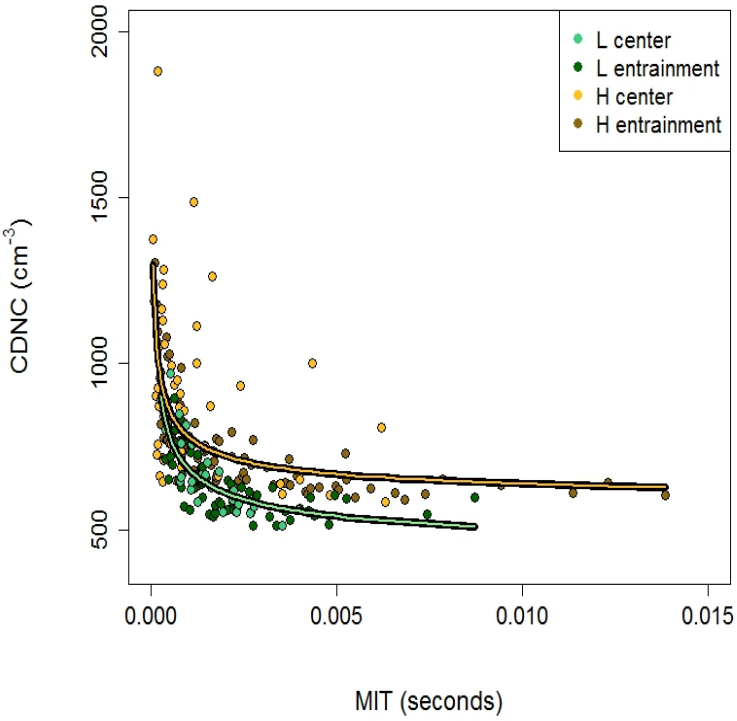

Figure 3.4: Gives MIT (sec) on the x–axis compared with CDNC (cm^{-3}) on the y–axis, with low pollution data in green (light green for cloud center and dark green for cloud edge) and high pollution data in gold (bright gold for cloud center and dark gold for cloud edge).

the respected axis. The p–value for the cloud droplet size (μm) data is $3.99 \cdot 10^{-10}$, while the p–value for the aerosol number concentration (cm^{-3}) is $2.22 \cdot 10^{-16}$. Both populations of data have p-values less than 0.05, making the difference in droplet size and aerosol number concentration significant for the two populations of data. Specifically looking at aerosol number concentration, the maximum probability density value for the low pollution case is 1554 cm^{-3} (the aerosol number concentration that corresponds to the maximum density value) while for the high pollution case it is 3129 cm^{-3}. This can be seen in the two density curves on the y–axis, where the density curve for low pollution data peaks quickly and drops off before 5000 cm^{-3}, whereas the peak for high pollution data is less prominent and decreases slowly towards 15000 cm^{-3}. The droplet sizes for the two populations of data are as expected. Droplets, on average, are larger for the low pollution clouds (13.4 μm) and smaller for the high pollution clouds (10.7 μm). This again, can be seen in analyzing the density curves on the x–axis, where the maximum probability density value for the low pollution case is 15.48 μm while for the high pollution case it is 11.35 μm. The linear best

fit trend lines show that as aerosol number concentration increases, the cloud droplet size decreases (with R–squared values of 0.24 and 0.07 for low and high pollution, respectively). Having two statistically different data populations is ideal for comparing PCF values for low and high pollution clouds. If clustering does not change between low and high pollution clouds, an argument cannot be made for the statistical similarities in the data sets.

In Section 2.3.2 we saw that the PCF is computed by using the MIT of droplets. Figure 3.4 gives the MIT (seconds) on the x–axis and the CDNC (cm^{-3}) on the y–axis. The key finding is that the MIT between the two datasets is statistically different, with a p–value of $9.16 \cdot 10^{-4}$. This finding is important due to the fact that the MIT is directly related to the computation of the PCF, strengthening the argument that similarities or differences in clustering cannot be accounted for due to similarities in the data populations. Another significant aspect of Figure 3.4 is the fact that the MIT has never been directly measured and related to a variable such as CDNC (to the best of the author's knowledge). Deductive reasoning tells one that as the number of droplets in a given volume increases, the space between said droplets will have to decrease. Having a mathematical formula to be able to calculate the MIT given that one has a measure of the CDNC can be advantageous not only for turbulence work, but to the consequences that droplet spacing has for collision–coalescence (increasing the amount of drops in a set volume increases the probability that a collision will occur between two random droplets). The formula used to develop the best fit lines in Figure 3.4 is CDNC = a + b $*$ MIT^d. Although beyond the scope of this work, more data will need to be collected to develop more distinct values of "a", "b", and "d" for one to be able to confidently predict the MIT given a CDNC value.

Note that it would be inappropriate to assume that all cloud measurements were taken from flights directly through the middle of clouds (in terms of vertical and horizontal extent). Although efforts were taken by the pilots, some in–cloud passes could have been made in favor of the top or bottom of a cloud (vertical extent), or passes may have been in favor for an edge of a cloud (horizontal extent), or any of the combinations just mentioned are possible (cloud top edge, cloud bottom edge, cloud top middle, cloud bottom middle, true cloud middle). There is no way to know for certain which section of a cloud is being measured. The results will be analyzed by assuming the aircraft passed through true cloud center, but it is important to keep in mind these possible scenarios which can affect the results.

3.2.3 Environmental Characteristics

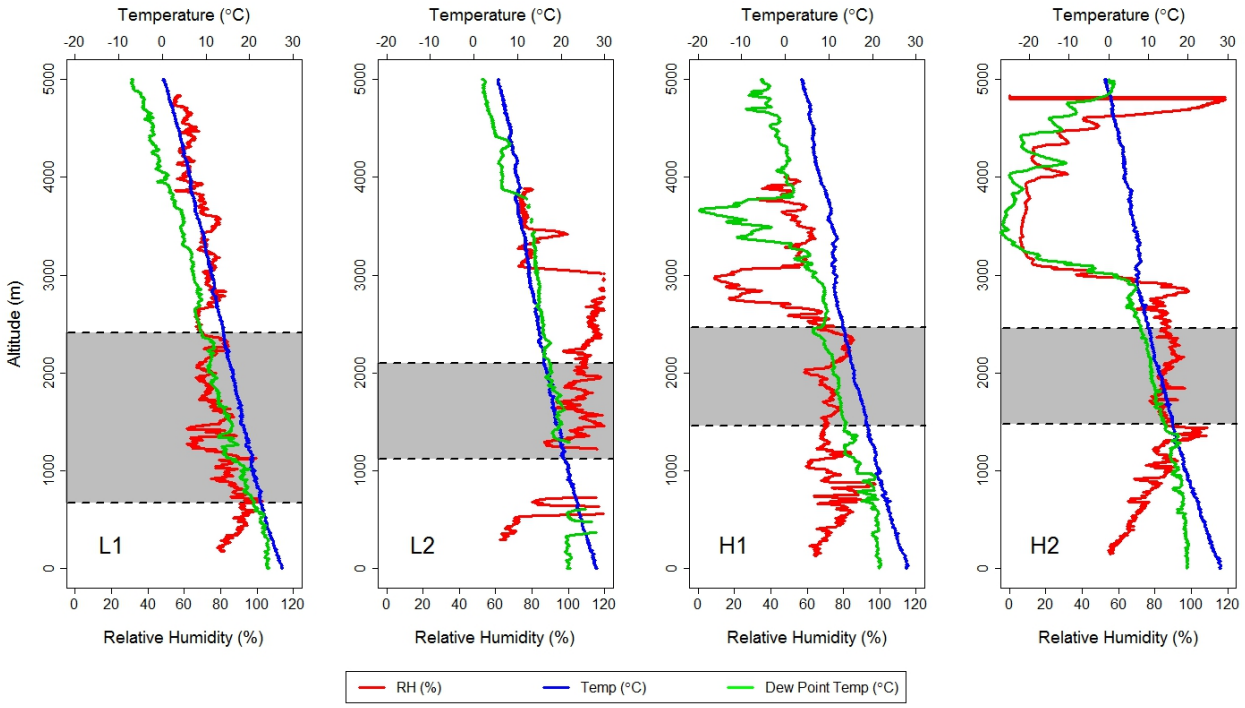

Figure 3.5: Shows vertical profiles of temperature in blue, dew point temperature in green, and relative humidity in red for L1, L2, H1, and H2 moving from left to right, respectively. The gray envelopes in each profile represent the range in altitude that clouds were measured.

Vertical profiles of temperature (blue), dew point temperature (green), and relative humidity (RH) (red) are presented in Figure 3.5. The data were collected during out–of–cloud spirals before or after cloud sampling on the four days that were selected for analyses. The gray envelopes represent the altitude range in which clouds were sampled for each flight. Cloud base and cloud top heights are defined as the lowest and highest level at which clouds were sampled for each day. Figure 3.5 is a recreation of Figure 1 from Small et al. (2013), only using different flight days. During any particular flight cloud base remained fairly constant, as described by CIRPAS pilots and the on–board mission scientists, with a cloud base height varying on average by $\pm$ 60 m (Small et al. 2013). From analyzing Figure 3.5, the vertical cloud range for H1 and H2 is virtually the same, with a thickness of 1007 and 976 m, respectively. L2 has a thickness of 981 m, while L1 has the largest thickness of 1740 m. The environmental lapse rate for each of the four flights, L1, L2, H1, and H2 is 5.4, 4.5, 4.8,

and 5.7 °Ckm^{-1}, respectively. The average RH, taken for the altitude range at which clouds were measured is 77, 105, 74, and 86 percent for L1, L2, H1, and H2, respectively. Note that RH measurement problems occurred during portions of L2 for values that were greater than 100 percent. Further analysis on the implications of the high RH encountered in the L2 flight will be discussed in section 3.6.1.

3.3 Entrainment vs. Center Clustering

3.3.1 Flight L1

It was hypothesized in Section 2.7 that droplet clustering will be enhanced in cloud entrainment zones due to the production of TKE from shear evaporative cooling at cloud edge. If droplet clustering acts as a function of turbulence (vorticity), then one would speculate that clustering would be enhanced in more turbulent areas. Figure 3.6 shows temporal lag on the x–axis (sec) and the PCF value (unitless) on the y–axis. Blue represents entrainment zone data while red represents cloud center data. The two envelopes represent the 85^{th} percent quantile (top of each envelope) and 15^{th} percent quantile (bottom of each envelope) of the data. The center lines in each envelope represent the entrainment mean clustering and center mean clustering (color dependent). From looking at Figure 3.6, one will notice that the mean lines have two thicknesses. A bold mean line represents entrainment and center data that is statistically significant from one another (p–value less than 0.05), while a thin mean line represents data that is statistically similar, determined using the Wilcoxon–rank–sum test. Temporal lag was used for the x–axis due to the simplicity of working with time data. In converting the temporal lag to spatial lag (by using the speed of the aircraft), on average, a temporal lag of $5 \cdot 10^{-3}$ sec corresponds to 30 cm. It is this scale, the cm scale, which we are concerned with. At larger, greater than a meter scale, the droplet spacing should be homogeneous. It is at the smaller scales (smaller scales of turbulence) where droplet clustering becomes dominant. This can be seen from analyzing Figure 3.6. The PCF function for both the center and entrainment zone is at a maximum at lower temporal ($< 5 \cdot 10^{-3}$ sec) scales, while it then decreases towards zero at larger temporal scales for all four flights. The main takeaway from Figure 3.6 is the fact that clustering is more dominant at smaller spatial/temporal scales, and the fact that clustering is enhanced for the entrainment zone of the cloud as compared to the center zone. Analyzing L1 (top left panel), the mean PCF value (taking the first 8 PCF values from the first 8 time lags, covering a temporal range of 0.0001 to 0.0021 seconds (6 to 13 cm)) for the center data is

0.18, while the mean PCF value for the entrainment data is 0.54. There is a statistically significant difference between the center and entrainment data, with the first 21 PCF values (temporal lag ranging from 0.001 to 0.006 (6 to 36 cm)) having an average p–value of 0.006, with 80.9 percent of the data being statistically significant. The first 21 PCF values were chosen to analyze the statistical significance because it is at these scales (centimeter scales and below) that we are concerned with. As can be seen for L1 in Figure 3.6, a small portion of the mean line is not significant near a temporal lag of 0.005 seconds. Another important aspect for L1 is the range of the quantile values. As can be seen, the 85^{th} percent quantile value for the entrainment zone is much larger than that of the center zone, which doesnt even reach the mean PCF value of the entrainment zone. The range of the quantile values for the center data is 0.053 to 0.25, whereas for the entrainment data the range is 0.094 to 0.835.

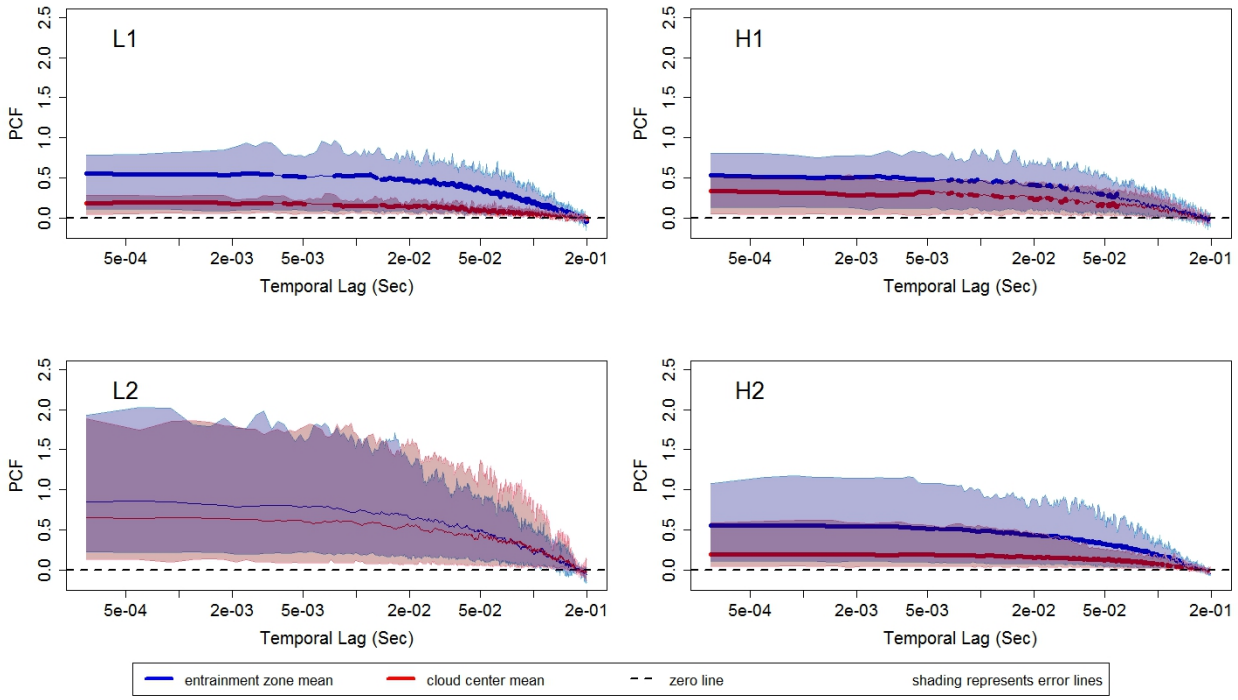

Figure 3.6: L1 (top right), L2 (bottom left), H1 (top right), H2 (bottom right). Shows temporal lag (sec) on the x–axis and the PCF value (unitless) on the y–axis, with entrainment zone data in blue and cloud center data in red. The envelopes represent the 85^{th} percent quantile value (top) and 15^{th} percent quantile value (bottom) of the data. The middle line in each envelope represents the mean PCF value, where a bold mean line represents entrainment and center differences that are statistically significant.

3.3.2 Flight L2

From analyzing L2 (bottom left panel) in Figure 3.6, there are significant differences from the L1 case. One feature of the figure remains the same however, the fact that the amount of clustering in the entrainment zone is enhanced as compared to the center zone. The mean entrainment clustering (PCF value) is 0.81 while the mean for the center is 0.62. However, the clustering differences are not statistically significant, with zero percent of the data having a p–value below 0.05. The average p–value for the first 21 PCF values is 0.30. Another difference that can be recognized are the envelope ranges. The quantile range for the center data is 0.099 to 1.79, while for the entrainment data the range is 0.21 to 1.81. The two 85^{th} percent quantile values are basically the same, as compared to the vast differences for the values in L1. More information as to the uniqueness of the L2 flight will be provided in Section 3.5, leading into a more significant discussion of the L2 flight in Section 3.6 as to why L2 displays the observed clustering characteristics.

3.3.3 Flight H1

H1 (top right panel) in Figure 3.6 shows clustering characteristics which are similar to the results which were presented in Section 3.3.1 for flight L1. Again, the entrainment zone clustering is greater than the clustering occurring in the center of the cloud. The mean entrainment clustering is 0.50 while the mean center clustering is 0.30, with 90.5 percent of the data having a p–value below 0.05 (for the first 21 PCF values). As can be seen, the statistical significance of the data breaks down at larger temporal lags, specifically when the lag is larger than 0.05 seconds. This is expected, since the two datasets converge to zero as the clustering disappears and droplet spacing shifts from nonhomogeneous to homogenous at larger temporal lags. Similar to L1, the 85^{th} percent quantile value for the entrainment data is larger than that of the center data. The quantile range for the center data is 0.03 to 0.51, while for the entrainment data the range is from 0.11 to 0.84.

3.3.4 Flight H2

Much of what has already been shown for L1 and H1 will be seen again for H2 (bottom right panel) in Figure 3.6. Again, the amount of clustering is greater for the entrainment data than that of the center data, with average PCF values of 0.54 and 0.19, respectively. One hundred percent of the data is significant for the first 21 PCF values, with an average p-value of $2.3 \cdot 10^{-4}$. Again, one can see that the 85^{th} percent quantile value for the entrainment

data is much larger than that for the center data, with ranges of 0.098 to 1.13 and 0.04 to 0.58, respectively. Note that individual graphs for each flight can be found in Appendix A.

3.3.5 Entrainment vs. Center Discussion

From analyzing the clustering for flights L1, L2, H1, and H2 in Figure 3.6, one can see that there are similarities that all four figures share, including:

- The mean PCF value for the entrainment data is always greater than the mean PCF value for the center data.
- The 15^{th} percent quantile value for the center data is always smaller than the 15^{th} percent quantile value for the entrainment data.

Excluding L2 from the analysis, the figures share even more in common, including:

- There is a statistical significance between the clustering occurring in the entrainment zone and the center zone of the cloud.
- The 85^{th} percent quantile value for the entrainment data is always larger than the 85^{th} percent quantile value for the center data.

The conclusion can be made that droplet clustering does change as a function of cloud center vs. cloud entrainment, with the entrainment zone having a larger amount of clustering than the center of the cloud, which is shown to be statistically significant using the Wilcoxon–rank–sum test. The enhanced clustering in the entrainment zone lets one "see" the turbulence of the cloud. These results agree with the statement in Shaw (2003) that TKE is generated by enhanced evaporation at cloud top and cloud edge due to evaporational cooling. Adding on, these results also agree with the findings by Smith and Jonas (1995), which is one of a few studies that looks at in–cloud fluxes as well as the horizontal domain-averaged transport properties of TKE. Smith and Jonas (1995) found that the dominant TKE source was at cloud top and cloud edge and was interpreted as evidence for the cloud–top entrainment instability process which produced observed strong downdrafts at cloud top and enhanced the overall turbulence in clouds. The evaporation cooled the surrounding air, causing negatively buoyant parcels which increased the turbulence not only in the overall cloud, but specifically cloud top and cloud edge where the TKE was being produced. Kitchen and Caughey (1981) used turbulence probes to measure dissipation rates in vigorous clouds. It was found that

dissipation rates were twice as large at cloud top (as compared to cloud base), which implied that there was strong TKE generation at cloud top. This concept can be inferred for cloud edge as well. Note that cloud top entrainment will be examined in Section 3.7.

The finding that clustering is more dominant at smaller spatial/temporal scales agrees with the findings from Kostinski and Shaw (2001), who concluded that statistically significant clustering was present at centimeter scales and below. The enhanced clustering on millimeter and centimeter scales is consistent with the inertial clustering hypothesis, which would suggest that the clustering continues to increase at scales below 1 mm (Shaw 2003). The result that clustering is enhanced at centimeter scales and below is also in accordance with finding from Kostinski and Jameson (2000); Larsen (2007) and Shaw et al. (2002).

3.4 Clustering Significance

Although it is clear from Section 3.3 that there is enhanced clustering in the entrainment zone as compared to the center zone, one needs to understand how to define if the overall clustering (including both the entrainment and center data) is significant. We begin by looking at the range of possible values the PCF can reasonably take on due to the finite nature of the data only. If the measured dataset falls outside of this range, one can conclude that they are seeing real, physical variability and the clustering in question is not perfectly homogeneous. For the data, 1000 realizations of a Poisson simulation were generated using the same intensity and duration as the original dataset. For example, L1 on average had 1290 droplets that spanned two seconds in duration. From looking at Figure 3.7, the largest and smallest value for each time lag from the Poisson simulations form the light–colored envelope while the 100^{th} largest and 100^{th} smallest (similar to an 80 percent confidence interval) form the darker envelope. This test is outlined in Larsen and Kostinski (2005).

Note that only the L1 case will be analyzed here, as to avoid redundancy since the L2, H1, and H2 flights show similar results with slightly different outcomes for the maximum and minimum values of the Poisson envelopes. The Figures for L2, H1, and H2 can be found in Appendix A. Looking at Figure 3.7, the far left panel shows the L1 graph from Figure 3.6. Panel (a) shows the center clustering from the L1 flight, with the maximum and minimum values for the Poisson simulations in light grey, and the 100^{th} largest and smallest values from the Poisson simulations (80 percent confidence interval) in dark gray. Panel (b) gives the same information as panel (a), except for entrainment data. Looking closely at the range of the Poisson envelopes, it can be seen that the envelope for the entrainment data is more

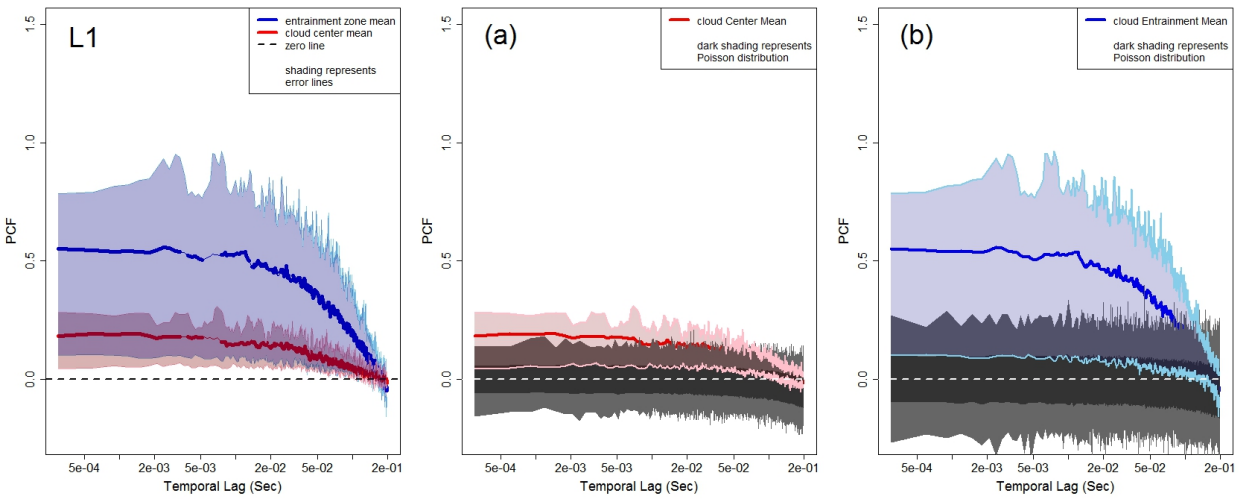

Figure 3.7: Shows the L1 panel from Figure 3.6 on the far left. Panel (a) shows the center clustering from the L1 flight, with the maximum and minimum values from the Poisson simulations in light grey, and the 100^{th} largest and smallest values from the Poisson simulations (80 percent confidence interval) in dark gray. Panel (b): as in Panel (a), except for entrainment data.

variable (covers a larger range) than the Poisson range for the center data. This is due to the significance of the PCF function relating to data limitations. Thinking back to Figure 2.2, more variation for Poisson simulations was clear below 2000 drops used in the PCF. The mean number of drops used in the Poisson simulation for the center data was 1788 drops over two seconds, while only 1042 drops over two seconds were used for the entrainment Poisson simulations. The smaller amount of drops in the entrainment zone is responsible for the larger Poissonian simulation range as compared to the center data.

PCF values that lie within the Poissonian simulation envelope were recorded by using the average PCF value (again, the average PCF value for the first 8 time lags). From looking at Figure 3.8, you can see that the total percentage of PCF values that are not considered significant is 44.4 percent for the entrainment zone of L1, while 50 percent of the PCF values for the center zone of L1 are not significant. Looking at the other three fights, the amount of nonsignificant clustering for L2 entrainment and center is 19.2 and 15.4 percent, respectively. Nonsignificant clustering for H1 entrainment and center is 39.7 and 44.8 percent, while for H2 entrainment and center it is 4.8 and 14.3 percent. Except for the L2 case, which does not have a statistical difference between entrainment and center clustering, the center clustering for L1, H1, and H2 contains a higher percentage of PCF values that are nonsignificant, as compared to the entrainment data. This is as expected, considering the PCF values for the

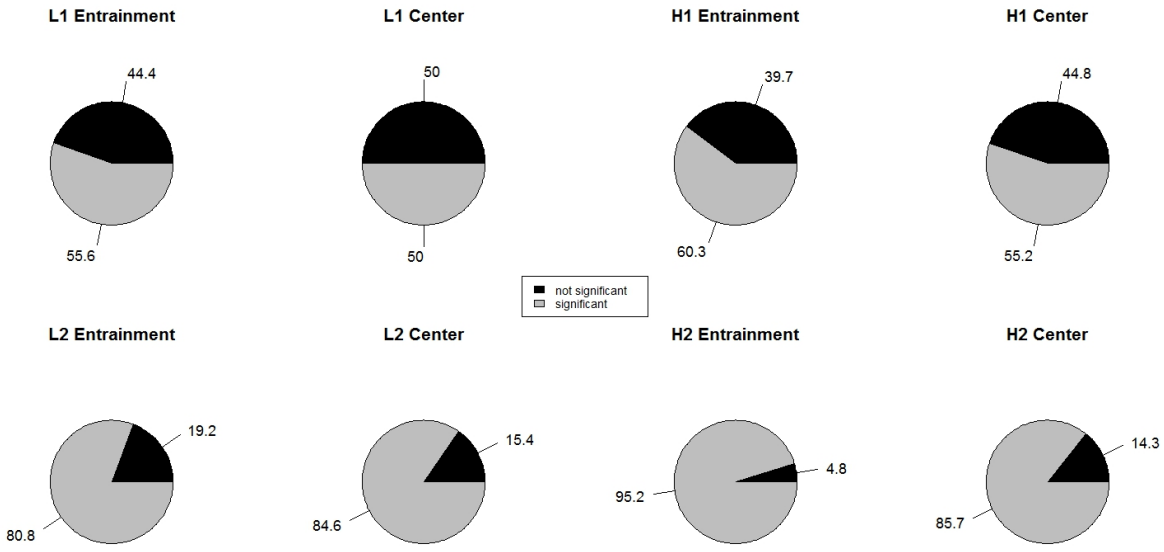

Figure 3.8: Pie charts of nonsignificant vs. significant PCF values for the low pollution clouds (L1 and L2) on the left, and high pollution clouds (H1 and H2) on the right. Black represents nonsignificant values, while gray represents significant values.

center data are significantly lower than those for the entrainment data.

Extending the analysis, from looking at Figure 3.9 (a), one can see that the total percentage of PCF values that are not significant for low pollution entrainment and center (averaging together L1 and L2) data is 33.9 and 35.5 percent, respectively. For high pollution entrainment and center (averaging together H1 and H2), the percentage of nonsignificant PCF values are 25 and 32 percent, respectively. Looking at the overall significant amount of clustering (total low pollution and total high pollution) in Figure 3.9 (b), the total amount of nonsignificant clustering for low pollution clouds is 34.4 percent, while for high pollution clouds it is 27.3 percent. The conclusion from Figure 3.9 (b) is that low pollution clouds have a higher amount of nonsignificant clustering than the high pollution clouds. Whether this is the case for all low vs. high pollution clouds remains to be seen however. More analysis will need to be conducted (more low pollution cases vs. high pollution cases) to get added confidence in the results.

The amount of significant vs. nonsignificant clustering appears to be dependent on a case by case basis. Where H1 and H2 have drastically different results (39.7 percent vs. 4.8 percent for nonsignificant entrainment clustering, respectively), along with L1 and L2, where nonsignificant entrainment clustering is 44.4 vs. 19.2 percent, respectively. A more

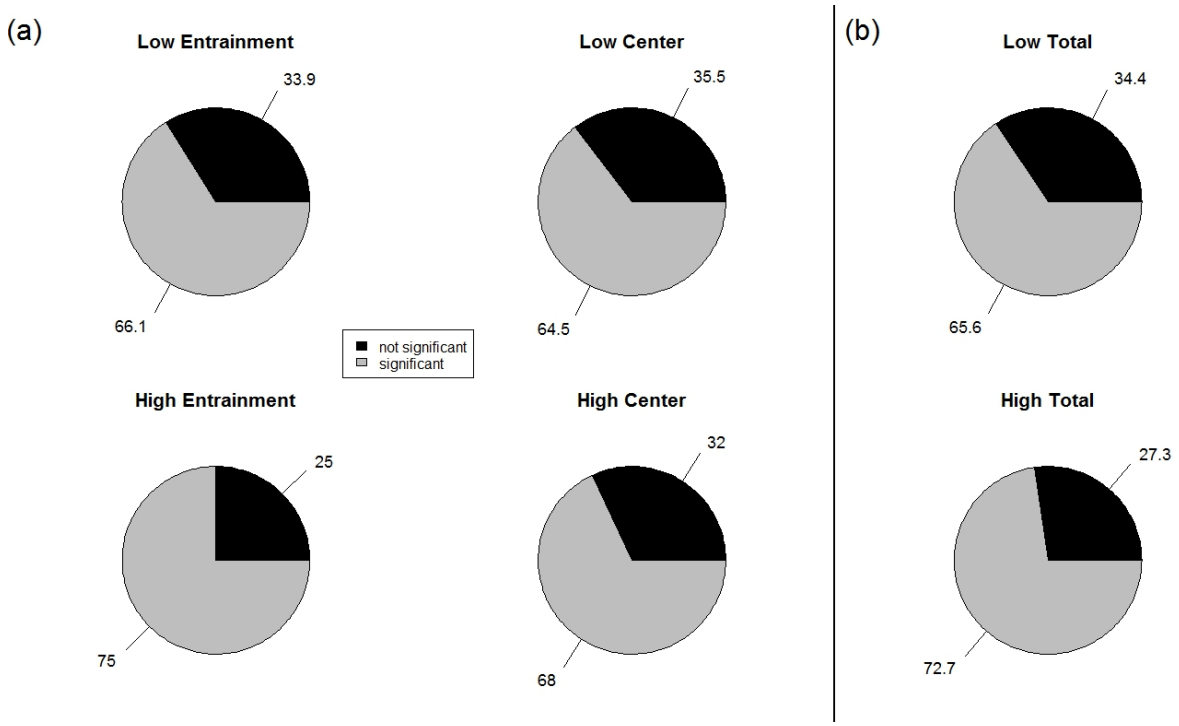

Figure 3.9: On the right, (a) shows pie charts of the percent of significant and nonsignificant clustering for low entrainment and low center on the top, and high entrainment and high center on the bottom. Black represents nonsignificant PCF values, and grey represents significant PCF values. On the left, (b) gives the same information as (a), except for low total (averaging center and entrainment) and high total (again, averaging center and entrainment).

scientifically direct measurement may give one an idea on the variables that control clustering significance. From looking at Figure 3.10, there does appear to be a relationship between the number of drops used in the PCF and the percentage of significant PCF values. As the number of drops used in the PCF decreases (5368, 1661, 1472, 1290 drops for H1, L2, H2, and L1, respectively), the percentage of nonsignificant PCF values exponentially increases (7.9, 17.9, 41.4, 46.2 percent for H1, L2, H2, and L1, respectively), as is modeled by the best fit exponential line in red (R–squared value of 0.79). Excluding H2 from the analysis (H2 drives the exponential fit), the blue line gives the linear best fit line for L1, H1, and L2, although the fit is not exponential, the percent of nonsignificant PCF values still decreases as the Number of drops used in the PCF increases. This relationship appears to be logical, as the more drops that are used in the PCF, the smaller the Poisson simulation envelope range will be, resulting in less PCF values being within that range. Whereas when less data is used, a larger envelope results, leading to a higher probability of PCF values falling within that

range. Again, more research and data needs to be collected to understand the relationship between significant PCF values and whether it truly is a function of the number of data points (in this case droplets) used, or whether the aerosol number concentration may also play a role in determining the percent of significant PCF values. Based on these results, it appears that the number of drops used to calculate the PCF plays a larger role than aerosol number concentration. Since a larger number of drops were used for high pollution data, this could be the reason as to why high pollution clouds have a smaller amount of nonsignificant PCF values (27.3 percent) as compared to the low pollution clouds which used a smaller amount of data and had a larger amount of nonsignificant PCF values (34.4 percent). Keep in mind that the number of drops used in the PCF depends directly on the aerosol number concentration, since this is what determines the size and quantity of droplets contained within a cloud. Besides the fact of what clustering significance is dependent on, all cases show evidence of non-Poissonian clustering. Meaning real, physical variability (clustering) is occurring for each case.

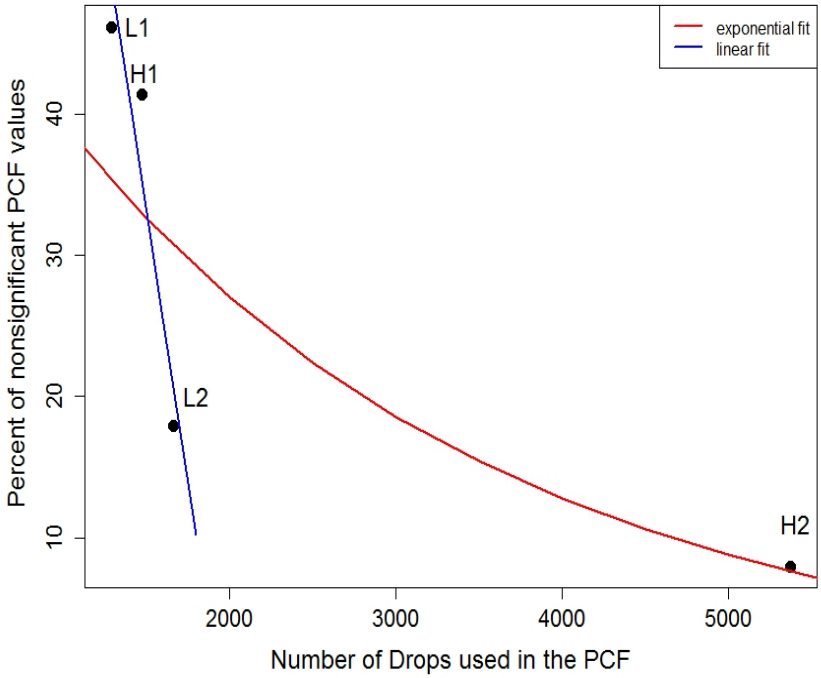

Figure 3.10: Gives the number of drops used in the PCF on the x–axis, with the percent of nonsignificant PCF values on the y–axis. The data points are shown in black with labels based on their respective flight. The best fit exponential curve is in red while the best fit linear line is in blue.

Results were also given for PCF values that lie within the 80 percent confidence interval of the Poisson simulations, although not shown here, to give the reader a quick idea, the results that were seen in Figure 3.9 also relate to the 80 percent confidence interval analysis. Looking at just total low and high pollution, the percent of nonsignificant PCF values was 12.9 and 8.7 percent, respectively.

3.5 Low vs. High Pollution Clustering

The work presented so far has determined that the clustering amount does change as a function of cloud location (cloud edge vs. cloud center), and the clustering measured is real, physical variability (non-Poissonian). As was mentioned in Chapter 1, clouds remain one of the greatest sources of uncertainty in estimating climate sensitivity, specifically due to cloud aerosol interactions. This Section aims to determine whether clustering changes as a function of aerosol number concentration. Figure 3.11 gives the same information as Figure 3.6 from Section 3.3.1, except for the PCF values for total low pollution (average of L1 and L2) in the left panel and total high pollution (average of H1 and H2) in the right panel. As can be seen, the characteristics of the two images are similar to what was analyzed in Figure 3.6. The average PCF values for the low pollution entrainment and center data are 0.69 and 0.42, respectively. For the high pollution entrainment and center data, the average values of the PCF are 0.53 and 0.25, respectively. One hundred percent of the first 21 time lags are statistically significant for both the low and high pollution cases. A close look at the average PCF values for the low and high pollution cases show that the mean PCF values for both the entrainment and center data are higher for the low pollution case as compared to the high pollution case (0.69 vs. 0.53 for the entrainment data and 0.42 vs. 0.25 for the center data). The larger average clustering amount for the low pollution clouds can be seen well in Figure 3.12. Panel (a) in Figure 3.12 shows low pollution data in green while high pollution data is presented in gold. The boundaries of each green and gold envelope are created by the average center clustering (bottom of each envelope) and the average entrainment clustering (top of each envelope). As can be seen, low pollution clouds are clearly offset to higher clustering amount for both average center and entrainment clustering. Panel (b) of Figure 3.12 shows the overall average of all of the PCF values for low pollution clouds (green) and high pollution clouds (gold). The overall average PCF value for low pollution clouds (average of entrainment and center clustering for both L1 and L2 clouds) is 0.57, while the overall average PCF value for high pollution clouds is 0.44. Although it appears that low pollution

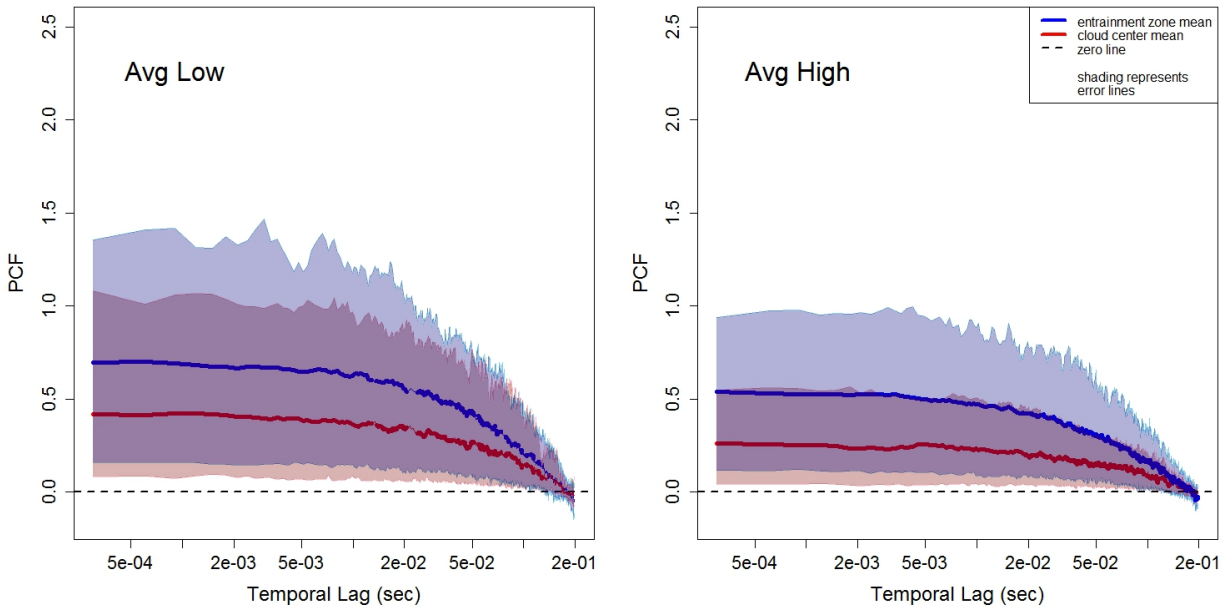

Figure 3.11: As in Figure 3.6 except for average low pollution PCF values (L1, L2) on the left and average high pollution PCF values (H1, H2) on the right.

clouds experience more clustering as compared to high pollution clouds, the difference is statistically similar (the p–value is greater than 0.05) using the Wilcoxon–rank–sum test. The average p–value is 0.10 for the first 21 time lags, while the average p–value for the entire graph is 0.28. In order to calculate the p–value using the Wilcoxon–rank–sum test, every PCF curve generated for low pollution clouds were grouped and every PCF curve generated for high pollution clouds were grouped. A p–value was then generated for each time–lag on the x–axis by calculating the Wilcoxon–rank–sum test between the two sets of data given for the specific x–axis location.

Although it appears that low pollution clouds have a non–statistically significant higher amount of clustering than high pollution clouds, further analysis shows that the higher amount of clustering in the low pollution case is due entirely to the L2 flight. Figure 3.13 gives the same information as Figure 3.12, except for the individual flights (L1, L2, H1, and H2) are shown instead of total low pollution and high pollution. From looking at panel (a), one can see that the average center and entrainment clustering for L2 (light green envelope) is beyond the range of the other three flights. The total average PCF for the clouds in L1, L2, H1, and H2 is shown in panel (b) of Figure 3.13. It can be seen that L2 has an average PCF value of 0.77, which is roughly twice the average PCF value (and statistically

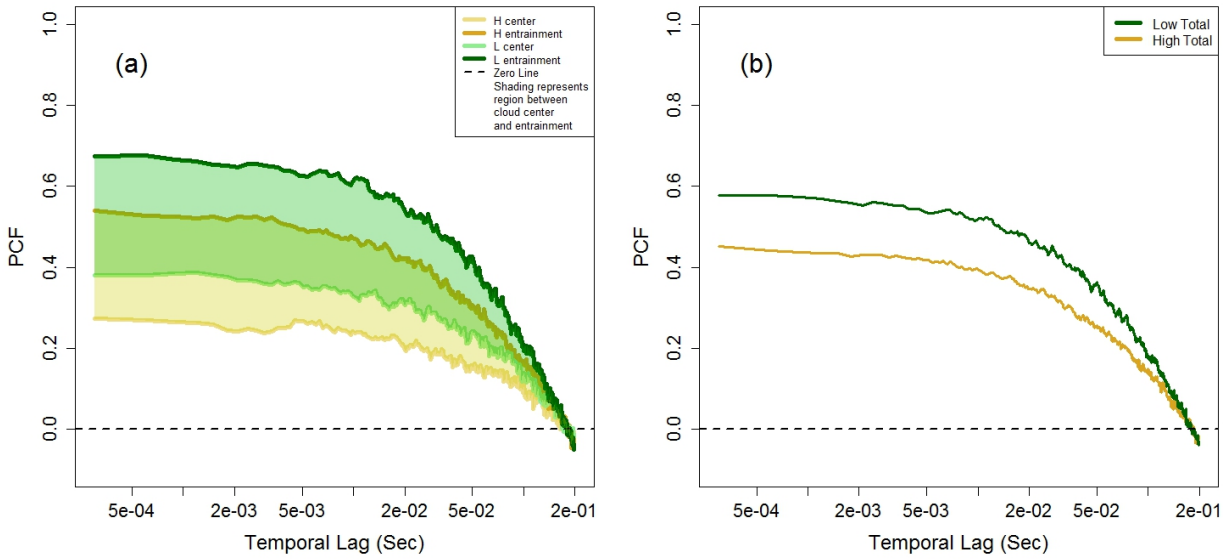

Figure 3.12: Shows low pollution data in green and high pollution data in gold. Panel (a) shows envelopes that span the average center PCF value (lower limit of the envelopes) to the average entrainment PCF value (upper limit of the envelopes). Panel (b) gives the overall average PCF value for the low pollution clouds and high pollution clouds.

significant) of the other three flights, where L1 (dark green), H1 (dark gold), and H2 (light gold) have average PCF values of 0.42, 0.44, and 0.43, respectively. The question of whether or not clustering depends on a function of aerosol number concentration cannot confidently be answered. Although Figure 3.12 shows that low pollution clouds have a larger amount of clustering, statistically speaking the clustering between low and high pollution clouds is the same. Further analysis shows that L1, H1, and H2 all have statistically similar clustering values, with average clustering amounts that are almost identical. Flight L2 has statistically significant clustering compared to the other three cases, and is solely responsible for causing the low pollution clouds to have a higher average PCF value than that of the high pollution clouds.

It was hypothesized in Section 2.7 that one would think if you have a higher number of drops (larger amount of aerosols), that the clustering would be greater due to the simple fact that there are more droplets available for clustering. However, the definition of the PCF from Equation 1.27 is: given the detection of a droplet, what is the probability of detecting another droplet in a given time window, divided by the mean number of drops per bin. The fact that you divide the probability by the mean number of drops "normalizes" the probability, making the PCF function non–dependent on the droplet concentration.

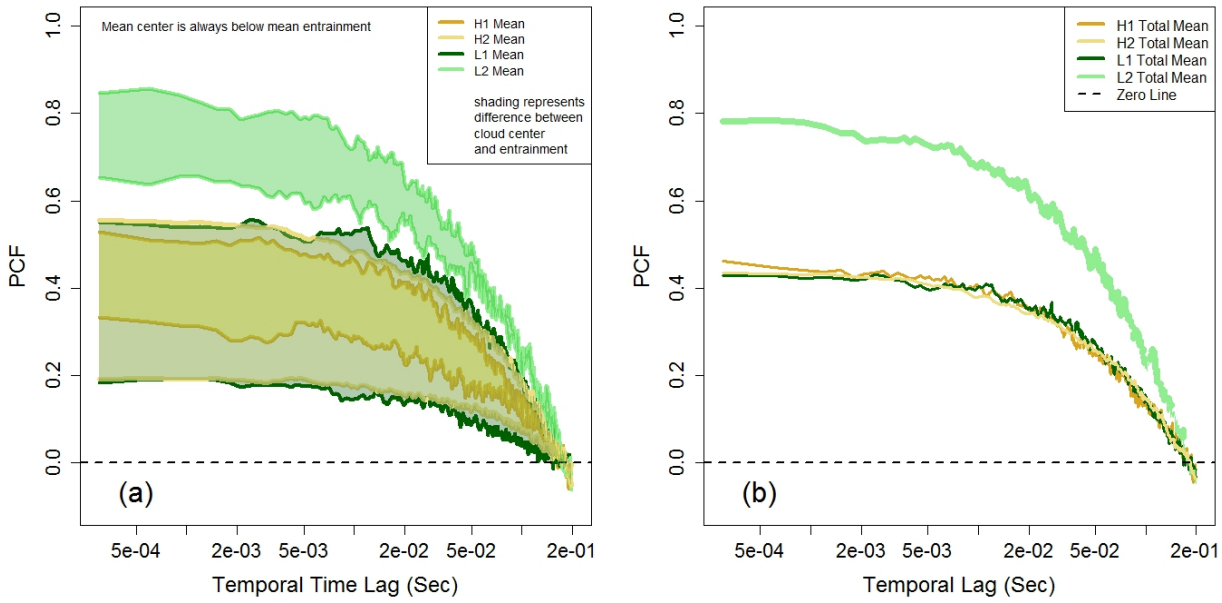

Figure 3.13: As in Figure 3.12 except for the individual flights of L1 (light green), L2 (dark green), H1 (light gold) and H2 (dark gold).

Therefore, whatever clustering that is occurring would need to be from a physical response due to the larger or smaller aerosol number concentration, and not based solely on the fact that there are more (less) droplets available for clustering, therefor there should be more (less) clustering. Thinking back to the Stokes Equation (Equation 1.19), the Stokes number depends on (1) turbulence parameters and (2) droplet size. Saw et al. (2008) found that droplet clustering increases for an increasing Stokes numbers (for Stokes numbers ranging from 0.01 to 1.5). Although the droplet sizes for the low and high pollution clouds analyzed here are statistically different, they are still too similar in size (mean sizes of 13.4 and 10.7 μm for low and high pollution clouds, respectively) to produce significant differences in the Stokes number. Remember, larger droplets are going to produce a larger Stokes number, resulting in slower droplet response rates to the flow, while the opposite is true for smaller droplets. When considering droplet sizes, perhaps the aerosol number concentration can affect the amount of clustering that is occurring if there are significant changes in the sizes of the droplet populations. Such a case could be comparing a high polluted cloud in Houston to a clean cloud over the ocean. More research will need to be performed however to confirm this theory.

3.6 Analysis of Flight L2

It has been made clear from Sections 3.4 and 3.5 that the clustering properties of L2 are abnormal when comparing them to L1, H1, and H2. First, the clustering is statistically similar between the entrainment and center zones of the clouds and second, the total clustering (the average PCF value) of L2 is almost double that of the other three flights. What is responsible for the abnormalness of the L2 flight? Here, several theories and arguments will be made as to why L2 is different, with the main theory being related to cloud age (growing, mature, dissipating stages). Another possible reason could be due to the location of the sampling within the clouds during the L2 flight.

3.6.1 Cloud Lifetime Theory

A study by Schmeissner et al. (2015) found that dissipating clouds have five main characteristics, including: a negative buoyancy, a negative vertical velocity, lower LWC and CDNC as compared to actively growing clouds, and a larger RH shell around the cumulus cloud. Cooper and Lawson (1984) also found that the LWC decreases due to entrainment as cumulus clouds deteriorate. Figure 3.14 shows a box plot with vertical velocity (left) and LWC (right) on the y–axis and L1, L2, H1, and H2 represented in that order on the x–axis. Red median lines represent datasets that are statistically significant when compared to L2. Note that the boxplot has a lower quartile, defining the lower limit of the box, a median line, represented by the heavy line inside each box, and an upper quartile, defining the upper limit of each box. The lines that fall outside of the box represent the limits of the nominal range of the data inferred from the upper and lower quartiles, while any points that fall outside of this range are shown as open circles (Marsh and Elliott 2008). The idea is that the interquartile distance, or the difference between the upper and lower quartiles, corresponding to the width of the central box, is used to determine the nominal range of data variation. The upper (lower) end of the nominal data range is defined as the upper quartile plus (minus) 1.5 times the interquartile distance (Marsh and Elliott 2008). As can be seen from looking at the left panel in Figure 3.14, L2 has the lowest vertical velocity and is statistically significant (p–values can be found in Table 3.2) when compared to the other three flights. The median value for vertical velocity is 0.25 m/s, as compared to the median values of L1, H1, and H2 of 0.94, 1.21, and 0.74 m/s, respectively. The fact that L2 has the lowest vertical velocity reflects the fact that the clouds in L2 are either dominated by downdrafts, or have week updrafts, making growth unlikely. The right hand panel in Figure 3.14

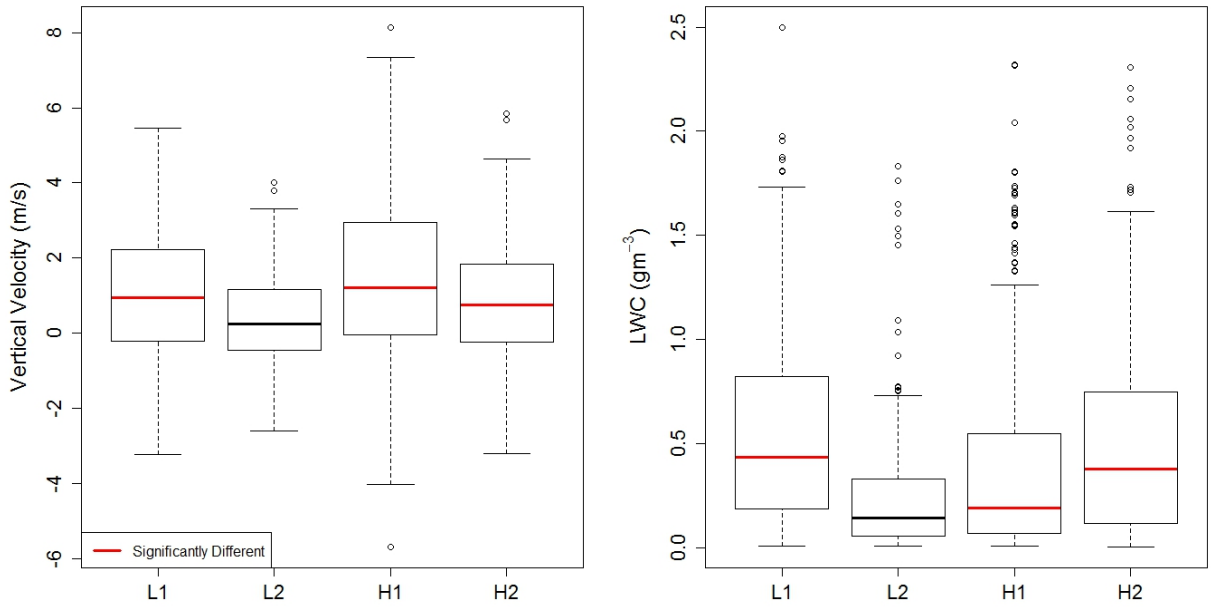

Figure 3.14: Box plots of L1, L2, H1, and H2, represented in that order on the x–axis, with vertical velocity (m/s) on the left and liquid water content (gm^{-3}) on the right. Red median lines represent datasets that are statistically significant compared to the L2 dataset.

shows that L2 has the lowest LWC, with a median of 0.14 gm^{-3}, and is statistically different (p–values can be found in Table 3.3) than the other three flights, which have median LWC values of 0.43, 0.19, and 0.38 gm^{-3}, for L1, H1, and H2, respectively. As was mentioned earlier, low LWC values signify that entrainment of dry air has been occurring, resulting in the evaporation of liquid water droplets, reducing the LWC (Pruppacher and Klett 1997).

Figure 3.15 shows a box plot of in–cloud RH for L1, L2, H1, and H2. The median in cloud RH for L2 is 101.2 percent, while the median in–cloud RH for L1, H1, and H2 is 101.3, 108.1, and 105.1 percent, respectively. Although the in–cloud RH for L2 is the lowest (by 0.1 percent), it is statistically similar to that of L1 with a p–value of 0.81. The red dots in Figure 3.15 represent the median out–of–cloud RH (100 m before and after cloud edge). As can be seen, the L2 case is the only flight where the RH increases out of the cloud. The out–of–cloud RH decreases for L1, H1, and H2. The median out–of–cloud RH is 97.8, 101.9, 106.8, and 99.4 percent for L1, L2, H1, and H2, respectively. According to Schmeissner et al. (2015), the width of the humid shell around clouds is larger for dissolving clouds as compared to actively growing clouds. The fact that the RH is larger, on average, outside of the clouds in flight L2 as compared to inside of the clouds for L2 could be a sign of a large humid shell that is surrounding the individual clouds within L2. The humid shell results

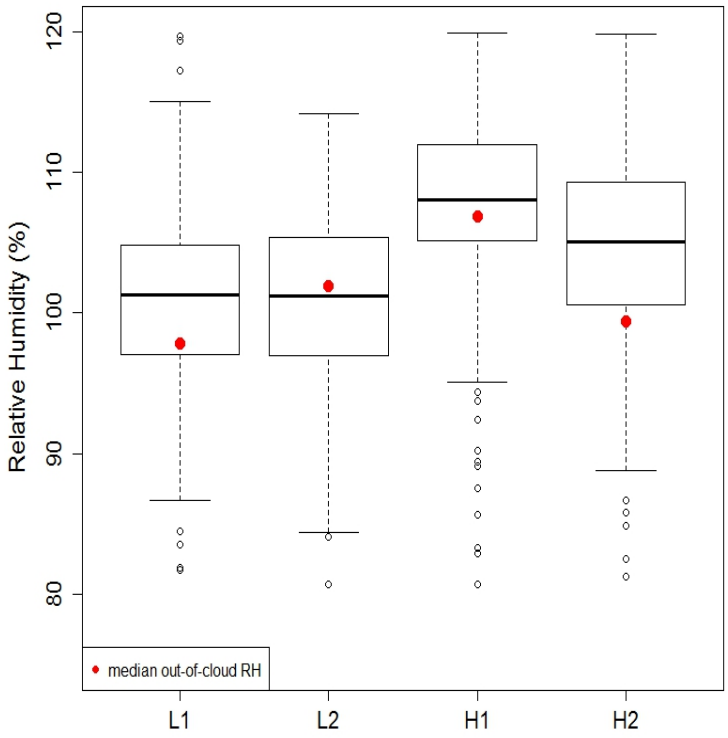

Figure 3.15: Gives a box plot of the in–cloud RH values for L1, L2, H1, and H2. The red dot on each box plot represents the median value of the out–of–cloud RH (100 m before and after the cloud edges).

from entrainment of dry air into the cloud, while moist air is detrained out of the cloud into the cloud free environment, resulting in a lower LWC in the cloud and a larger RH outside of the cloud. More evidence for the large humid shell can be gathered from the vertical profiles of RH from Figure 3.5, where the average RH for the altitude range of cloud measurements was 105.2 percent, while for the other flights the RH was considerably lower (see Section 3.2.3).

Figure 3.16 shows box plots of CDNC (cm^{-3}) on the left and cloud width (m) on the right. As can be seen from analyzing the CDNC panel, L2 has the lowest CDNC and is statistically significant (p–values can be found in Table 3.3) when compared to the other three flights. The median CDNC value for L2 is 172 cm^{-3}, as compared to the median values of L1, H1, and H2 of 261, 399, and 406 cm^{-3}, respectively. The fact that L2 has the lowest CDNC reflects the results from Schmeissner et al. (2015) that the clouds in L2 are characteristic of dissipating clouds. Although the study by Schmeissner et al. (2015) doesn't discuss cloud width, it can be seen from looking at the right hand side of Figure 3.16 that L2

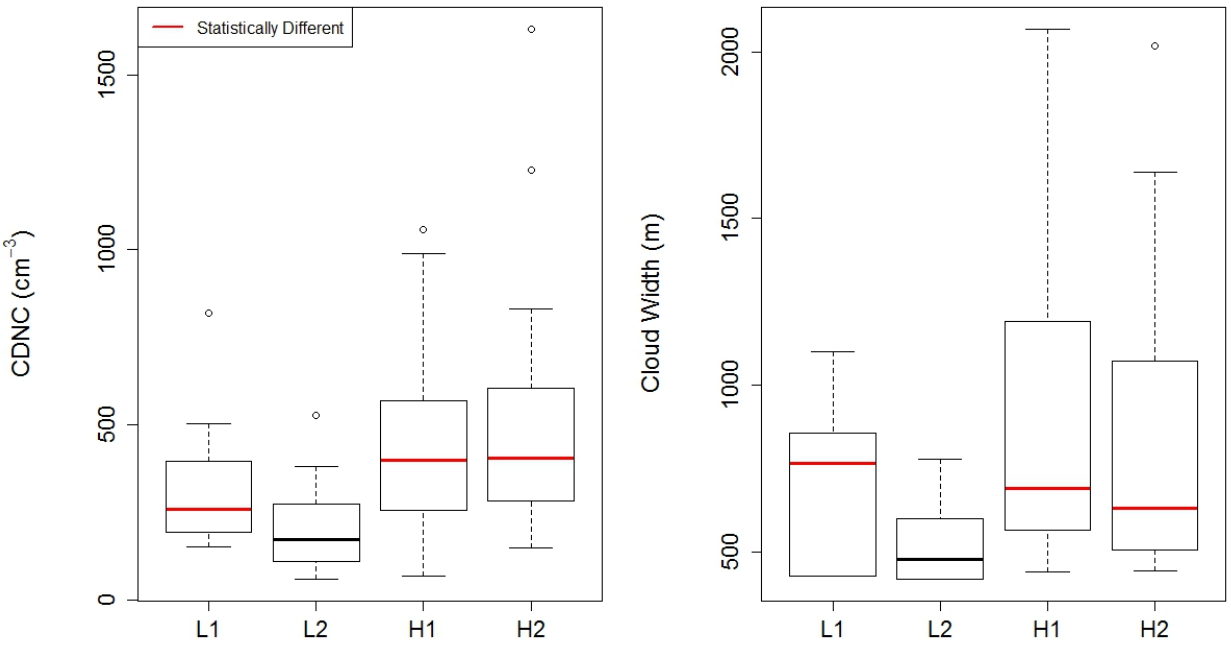

Figure 3.16: As in Figure 3.14, except for CDNC (cm^{-3}) on the left and cloud width (m) on the right.

has the lowest cloud width, with a median of 480 m, and is statistically significant (p–values can be found in Table 3.3) when compared to the other three flights, which have median cloud width values of 765, 690, and 631 m, for L1, H1, and H2, respectively. Clouds that are dissipating would be expected to have a smaller horizontal extent due to entrainment of dry air leading to evaporation of cloud edge droplets as compared to a developing or mature cloud. The results shown in Figure 3.16 are in agreement with the average cloud width and CDNC values given in Table 3.1, where it is shown that L2 has the lowest average CDNC and cloud width (see Section 3.2.2). Comparing cloud width on different days can become complicated due to the environmental factors that control cloud size. As is discussed in Hill (1973), the dominant factor governing the size of cumulus clouds is the size and strength of the sub–cloud circulations. There is no way to know what the sub–cloud circulation velocity was for the given days. Only vertical velocity (m/s) is available, which, as we saw from Figure 3.14, was smallest for the in–cloud portions of the L2 flight. Whether or not the L2 clouds were smaller as compared to the other flights is due to dissipation or environmental characteristics is unknown.

Figure 3.17 shows the in–cloud buoyancy (ms^{-2}) values for each of the four flights, with

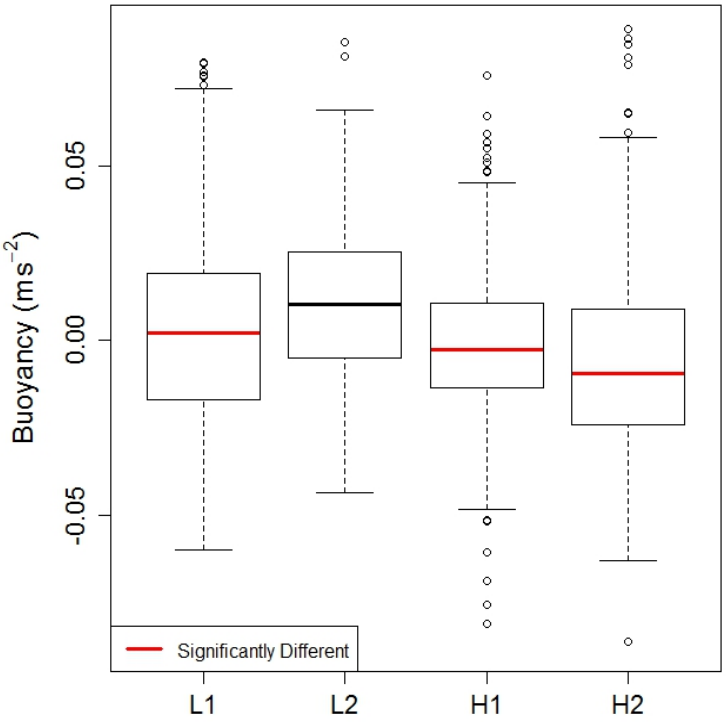

Figure 3.17: As in Figure 3.14 except for buoyancy (ms^{-3}).

the red median lines in the box plot representing values that are statistically different than the L2 flight (p–values can be found in Table 3.3). It can be seen that L2 has the largest median buoyancy of 0.0081 ms^{-2}, which is statistically significant as compared to the other median buoyancy values of -0.00097, -0.0031, and -0.010 ms^{-2} for L1, H1, and H2, respectively. The buoyancy was calculated by calculating the in–cloud virtual potential temperature and out of cloud (100 m before and after cloud edge) virtual potential temperature.

The clouds in the L2 flight have four out of the five characteristics mentioned in Schmeissner et al. (2015) for decaying clouds, including (1) lowest vertical velocity, (2) lowest LWC, (3) largest RH shell around the cumuli (a larger RH recorded outside of the cumulus cloud as compared to within the cloud), (4) lowest CDNC. Although not mentioned in Schmeissner et al. (2015), the L2 clouds have (5) lowest average cloud width. All evidence points to the clouds in the L2 flight to be decaying on average, and therefore to be more turbulent as dry air is mixed into the clouds causing dissipation. The statistically similar values between center and entrainment clustering add to the cloud lifetime theory, as it is not only the entrainment zone in the L2 cumuli that is experiencing mixing, but the entire horizontal extent of the clouds (both the entrainment and center zones) that are experiencing mixing and

Table 3.3: P–values calculated using the Wilcoxon–rank–sum test between L2 and L1, L2 and H1, and L2 and H2 for in–cloud vertical velocity (m/s), LWC (gm^{-3}), RH (percent), CDNC (cm^{-3}), cloud width (m), and in–cloud buoyancy (ms^{-2}). Statistically significant values are presented in bold.

Comparison	Vertical Velocity (m/s)	LWC gm^{-3}	RH (Percent)	CDNC (cm^{-3})	Width (m)	Buoyancy (ms^{-2})
L2–L1	$\mathbf{6.0 \cdot 10^{-4}}$	$\mathbf{2.1 \cdot 10^{-12}}$	0.82	**0.03**	**0.008**	**0.011**
L2–H1	$\mathbf{2.9 \cdot 10^{-7}}$	**0.009**	NA	**0.002**	$\mathbf{7.8 \cdot 10^{-4}}$	$\mathbf{3.9 \cdot 10^{-5}}$
L2–H2	**0.011**	$\mathbf{2.2 \cdot 10^{-7}}$	NA	$\mathbf{1.9 \cdot 10^{-4}}$	**0.003**	$\mathbf{5.5 \cdot 10^{-11}}$

dissipation. However, one would expect the buoyancy of turbulent, dissipating clouds to be negatively buoyant, not positively buoyant as is shown in Figure 3.17. Although the clouds in the L2 flight show most of the characteristics of decaying clouds, more research needs to be conducted to give a definite answer as to the cause of the higher clustering amount experienced in the L2 flight. Unfortunately, there is no way to know for certain whether or not the clouds were decaying, as no visual observations are available for the clouds at the time of measurement. Observations of the cloud edge could tell one whether the cloud is growing (sharp cloud edge) or dissipating (fuzzy cloud edge).

3.6.2 Measurement Location for L2

The flight path through the cumulus clouds could have favored cloud edge or cloud top instead of true cloud center. Favoring cloud edge (or top, see Section 3.7) would result in measuring areas of the cloud that favor a higher amount of clustering, and could result in the overall larger amount of average clustering experienced in the L2 flight. Measuring just cloud edge would result in: a lower LWC, a lower vertical velocity, a lower CDNC due to evaporation from entrainment, and a shorter cloud width (from not traversing the maximum diameter of the cloud). However, just as with the cloud lifetime theory, the buoyancy is expected to be negative, not positive, in the entrainment zone and cloud top due to evaporational cooling of the air. Both the dominant flight traverse through the clouds, the cloud lifetime theory, or a combination of both could explain the larger amount of clustering in L2. However, more research is required to resolve the buoyancy, which causes conflict for both explanations.

3.7 Case Study (PCF vs. Altitude)

We can confidently say that we see a physical change in the amount of clustering that is occurring between the center of a cloud and the edge of a cloud, although this clustering does not have a statistical change as a function of aerosol number concentration. This section analyzes multiple passes that occurred through a single, vertically developed cumulus cloud (non–precipitating), allowing one to analyze how droplet clustering changes as a function of cloud height. Figure 3.18 shows flight altitude as a function of time (blue) in the top panel, with droplet counts per second represented in black. Sixty–nine cloud traverses were made over an altitude range of 806 to 3381 m. As can be seen from the bottom panel of Figure 3.18, the cloud location was to the North/Northwest of Houston. Note that the location of the cloud is not in one single location, but with each passing cloud traverse, the location of the cloud has slightly shifted. This is due to the fact that the cumulus cloud is embedded within the wind field and therefore moves with the dominant winds of the atmosphere at the time of measurement. Cloud and flight properties can be found in Table 3.4, with standard deviations presented in parentheses for select variables. Important variables to look at include: the mean CDNC was 472 cm^{-3}, the mean drops per second encountered was 1679, and the aerosol number concentration was 1396 cm^{-3}. These characteristics make the Case cloud similar in nature to the low pollution clouds in L1 and L2 analyzed in Sections 3.2–3.6. A comparison of the overall mean PCF for the Case to the mean PCF values for flights L1, L2, H1, and H2 would be inappropriate due to the different measurement methods. For the low and high pollution flights, singe flight traverses were made through individual cumulus clouds, whereas for the Case flight, multiple flight traverses were made through a single cumulus cloud at different altitudes. The purpose of this Case flight is solely to determine how the PCF changes as a function of cloud height.

Figure 3.19 shows how the cloud droplet number distribution (cm^{-3}) changes as a function of normalized cloud altitude in panel (a). Panel (b) shows the median droplet diameter in black with the 5^{th} and 95^{th} percent quantile for droplet diameters represented by thin red lines. Panel (c) shows LWC (gm^{-3}) in blue, RH (percent) in red, and the PCF in black. Panel (d) gives the average vertical velocity (m/s). Each variable is presented as a function of normalized cloud height. It is expected, that the median droplet size should increase as you go up in cloud altitude due to the fact that the droplets have had more time to grow either through condensation or collision–coalescence. The median droplet size increases from 11.88 μm at cloud base to 17.65 μm at cloud top, as seen in panel (b) in Figure 3.19. Klazura (1971) found that the higher a cloud builds, the broader the size spectrum becomes, and the

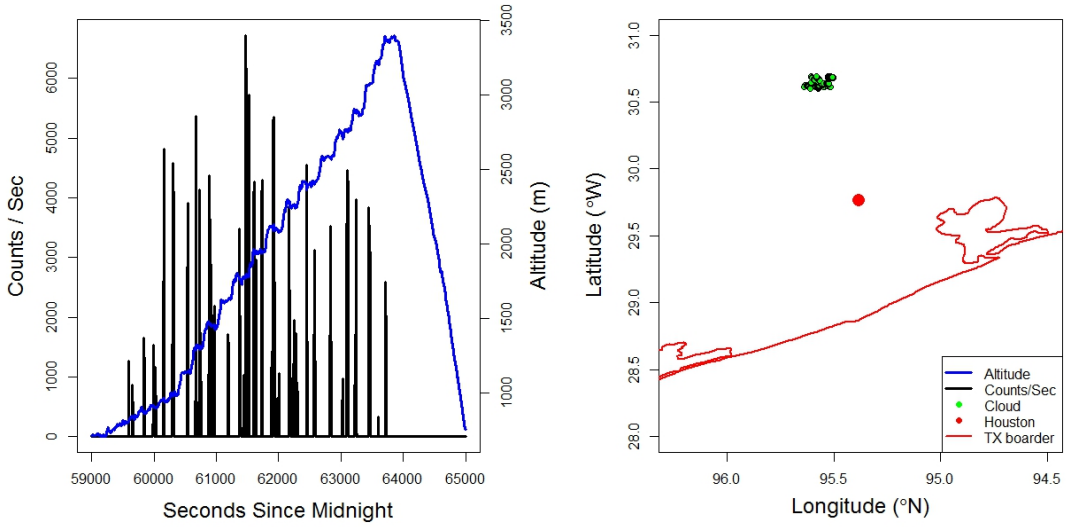

Figure 3.18: As in Figure 3.1 except for the Case Flight. Note flight path has been removed.

larger the concentration of large drops becomes. This can be understood from looking at panel (a), where the drop size distribution can be seen to shift to larger sizes slowly as you go up in normalized cloud altitude. The longer a droplet stays suspended within a cloud, the more likely the said droplet is to grow via condensation due to supersaturation values associated with strong vertical updrafts. For this reason, the median cloud size is typically largest at cloud top, which agrees with the results in Figure 3.19. From the observations of Zaitsev (1950); Mei-Yuan (1963) and Warner (1955), one notes that the distribution of cloud LWC has several characteristics, including: (1) LWC at any given level varies over short distances in a manner that is closely related to the variation of the vertical velocity (stronger vertical velocity leads to larger supersaturations and droplet growth via condensation). (2) Cloud LWC increases with height above cloud base and assumes a maximum somewhere in the upper half of the cloud, and then decreases again toward cloud top. (3) The distribution of LWC parallels the distribution of the drop size; this statement is in disagreement with the finding in Small et al. (2013) who found that LWC variability as a function of altitude is driven primarily by variability in drop concentration rather than drop size. All of these characteristics can be seen by analyzing panels (b) and (c) of Figure 3.19. The LWC increases

Table 3.4: As in Table 3.1 except for the Case flight.

Variables	Case Flight
Date	2006–Sept–10
Flight Number	RF 18
UTC for Cloud Sampling	1633–1742
Clouds > 300 m in width	1
Min cloud base height (m)	806
Max cloud top height (m)	3381
Cloud thickness (m)	2575
Cloud width (m)	943 (472)
Mean true air speed (m/s)	61.4 (2.2)
Mean CDNC (cm^{-3})	472 (404)
Max CDNC (cm^{-3})	2342
Mean drops (s^{-1})	1679 (1441)
Mean vertical velocity (ms^{-1})	0.35 (1.89)
$N_{aerosol}$ (cm^{-3})	1396 (918)
N_{acc} (cm^{-3})	943 (472)

from cloud base (0.073 gm^{-3}) to a normalized height of 0.7 (0.98 gm^{-3}), and then decreases to 0.23 gm^{-3} at cloud top. As the LWC increases, so too does the median drop size (comparing panels (a) and (b)). Once the LWC starts to decrease, median cloud droplet size still continues to increase. Warner (1969) observed that the tendency of a size distribution to be bimodal increased with height above cloud base and with decreasing stability in the cloud environment. Based on these observations, it is suggested that bimodal drop size distributions (two local maximum peaks in the size distribution) are the result of a mixing process between the cloud and the environment (entrainment). Since the drop size spectra was found by Warner (1969) and Small et al. (2013) to be fairly uniform for a given level across a cloud and the bimodality was not confined to the cloud edges, it was proposed that the mixing process producing the bimodality is due mostly to entrainment of drier air at the growing cloud top, and to a lesser degree to entrainment at the cloud edges. As can be seen in panel (a) of Figure 3.19, bimodal droplet size spectra are present at a normalized cloud altitude of 0.7, 0.9, and 1. It is therefore proposed that the median droplet diameter increases with normalized cloud height because (1) The droplets have a longer period of development through condensational and collision–coalescence growth mechanisms and (2) Entrainment mixing leads to a bimodal distribution of the droplet spectra. This agrees with the findings

from Warner (1969) that entrainment must be occurring at cloud top. Panel (c) in Figure 3.19 shows that at a normalized cloud altitude of 0.7 the RH sharply decreases along with a steady decline in the LWC. This is characteristic of cloud top entrainment (mixing of dry air into the cloud reducing LWC and the RH).

As can be seen in panel (c), the PCF (which doesn't have any physical change throughout the lower portion of the cloud, ranging from 0.73 at cloud base to 0.26 at a normalized height of 0.8) has a sharp increase at cloud top, increasing from 0.26 to 1.81 between the normalized heights of 0.8 and 1.0 (cloud top), indicating enhanced droplet clustering at cloud top where cloud top entrainment is occurring. Small et al. (2013) found that turbulence increases with altitude within a cloud. Although there is no clear response in the PCF increasing with cloud height, there is a clear increase in the PCF at the cloud top entrainment zone where TKE is being produced through evaporational cooling of liquid water. The entrainment can be seen to cause a negative vertical velocity in the upper portion of the cloud (panel (d) in Figure 3.19), where the average vertical velocity is increasingly negative above a normalized cloud height of 0.5, suggesting that penetrative downdrafts extend into the middle section of the cloud. This supports the findings from Smith and Jonas (1995) who found that the dominant TKE source was at cloud top and was interpreted as evidence for the cloud–top entrainment instability process which produced observed strong downdrafts at cloud top and enhanced the overall turbulence of the cloud. The secondary maximum peak in the PCF of 0.84 at a normalized cloud height of 0.4 could be indicative of the convergence of the cloud base updraft and the cloud top downdraft at a normalized height of 0.5, although this is purely speculative and more research will need to be conducted to determine if any patterns emerge. The larger average PCF value of 0.76 for the normalized cloud height between 0.0 and 0.1 (cloud base) as compared to the average lower PCF value of 0.31 before the cloud top entrainment zone for the normalized cloud heights between 0.7 and 0.8 could be indicative of TKE being produced due to condensational heating at cloud base (Shaw 2003). Again, more data will need to be collected to identify consistent patterns in the vertical variations of the PCF with cloud height.

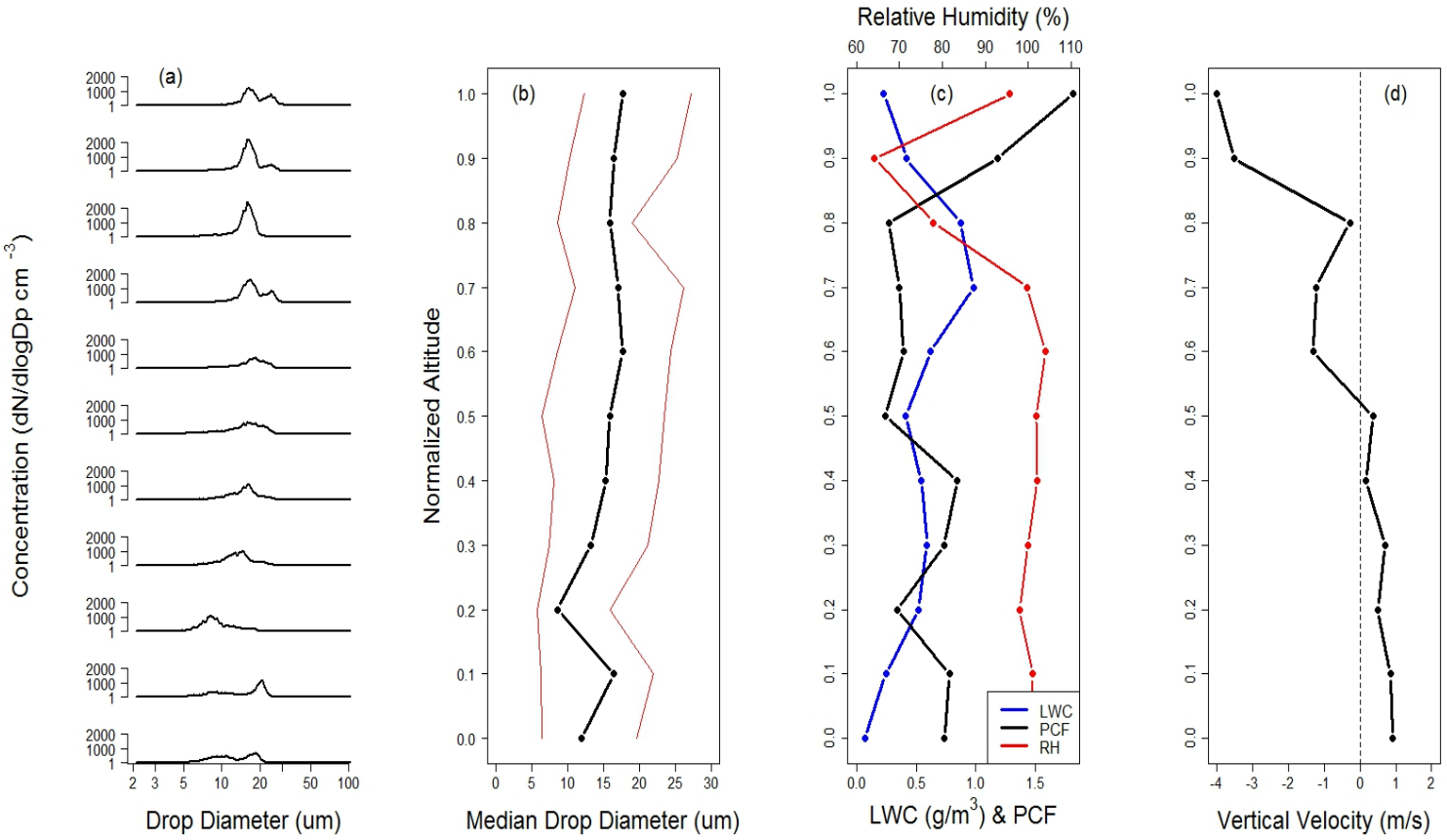

Figure 3.19: Shows the cloud droplet size distribution in panel (a). Panel (b) gives the median drop diameter in black with 5^{th} and 95^{th} percent quantile values in red. Panel (c) gives LWC (blue), RH (red), and the PCF (black). Panel (d) shows vertical velocity. All variables are represented as a function of cloud normalized altitude.

3.8 Issues with Presented Results

3.8.1 PCF vs. Number of Drops

It was mentioned in Section 2.3.2 that the longer the data series used to calculate the PCF is, the more statistically significant the calculation will become. Figure 2.2 showed, particularly when less than 2000 drops were used, that the Poisson simulated PCF values began to variate. This becomes concerning for the results seen throughout Chapter 3 because the number of droplets used to calculate the PCF varies for each PCF value that was calculated. Figure 2.2 also showed that the variation was symmetric around the x–axis, so as long as multiple PCF calculations were computed, comparisons could still be made in confidence. In checking the results to see how the number of drops used in the PCF affected the overall PCF value, one can refer to Figure 3.20. The PCF value is shown on the y–axis with the number of drops used in the PCF on the x–axis, with entrainment data in blue and center data in red. Each point represents each two–second zone (entrainment and center) from each cloud measured. The data is fit with a nonlinear least squares (NLS) fit line. A NLS fit determines the nonlinear (weighted) least–squares estimates of the parameters of a nonlinear model. The R–squared value of the fit isn't of importance. The fact that the fit line increases below 2000 drops used is of importance. It is clear that larger PCF values were calculated below 2000 drops used as compared to when a larger numbers of drops were used. Seventy percent of the entrainment data used lies below 2000 drops, as compared to only 46 percent of center data. Could this be the reason why entrainment data has larger PCF values than center data, due to the amount of droplets used in the entrainment zone vs. the center zone? The answer appears to be no. The ratio of the average PCF value for the entrainment to center data below 2000 drops used is 1.34, while for above 2000 drops used it is 2.15. These ratio values show that the PCF is actually larger for entrainment data vs. center data when more than 2000 drops were used as compared to when less than 2000 drops were used. When extending the ratio of droplets used out to above and below 4000, the average PCF value ratio for entrainment vs. center data is 1.64 and 1.68 for above and below 4000 drops used, respectively. These results mean that, although more variation occurs below 2000 drops used, the physical amount of clustering for entrainment and center data is still recorded. Recall that the variation that occurs below 2000 drops is symmetric from Figure 2.2, and this is shown in computing the ratio values. The average number of drops used to calculate the PCF for L1, L2, H1, and H2 was 1290, 1661, 1472, and 5383 droplets, respectively. In accordance with the number of drops used, L2 had the second low-

est percentage of data below 2000 droplets at 69 percent, while the percentage of data below 2000 for L1, H1, and H2 was 79, 74, and 23 percent, respectively. Recall that the average PCF values for L1, H1, and H2 were all statistically similar, even though the average number of drops used ranges from 1290 for L1 to 5383 for H2. L2 has the largest average PCF value, but has the second largest number of drops used, behind only H2. If the number of drops was the sole determining factor in the output of the PCF, H2 would have the smallest amount of clustering, followed by L2, while L1 would have the highest amount of clustering. The fact that this is not the case means that, although the PCF may have variance based on the number of drops used, the physical amount of clustering is still computed as long as multiple calculation are made. Physically, the PCF would not be a statistical tool if the outcome was based solely on the amount of data used and not the physical amount of clustering that is occurring within the data.

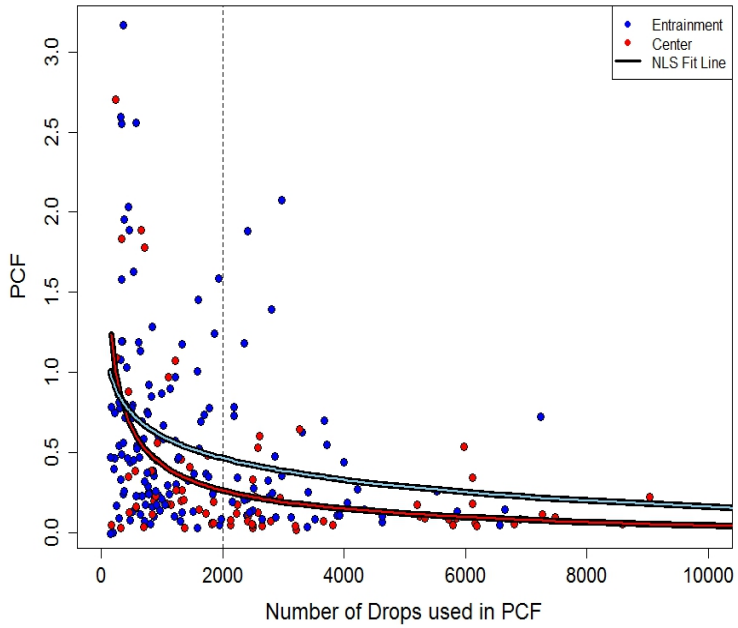

Figure 3.20: Shows the PCF on the y–axis and the number of drops used in the PCF on the x–axis, with entrainment data in blue and center data in red. Each dataset is fit with a NLS fit line which matches the color of the dataset that they correspond too. A dashed line is present at 2000 drops used for reference.

3.8.2 Relationship Between LWC and Number of Drops

It was shown in Section 3.3 that entrainment clustering has a statistically significant larger amount of clustering than the center of the cloud. Pruppacher and Klett (1997) states that entrainment of relatively dry environmental air into convective clouds lowers the LWC substantially below that expected on the basis of a closed saturated ascent. It is therefore expected that the lower the LWC, the higher the clustering should be due to entrainment mixing (recall Figure 3.15, where L2 has the largest average PCF value and the lowest average LWC). Figure 3.21 shows the PCF as a function of LWC, with entrainment data in blue and center data in red, with a reference line at 0.3 gm^{-3}. From analyzing Figure 3.21, it can be seen that as the LWC decreases, the PCF increases (as is seen from the NLS fit lines). Particularly below a LWC value of 0.3 gm^{-3}, the PCF takes on large values as compared to the PCF values above a LWC of 0.3 gm^{-3}. L1, L2, H1, and H2 have 42.6, 82.1, 71.2, and 55.5 percent of their data that lies below a LWC of 0.3 gm^{-3}, with average PCF values of 0.79, 0.88, 0.52, and 0.63, respectively. In comparing center and entrainment data, 43.2 and 72.2 percent lie below 0.3 gm^{-3}, respectively, with average PCF values of 0.71 and 0.52 as compared to the average PCF values of 0.26 and 0.14 for data that lies above 0.3 gm^{-3}. These values show for both entrainment and center data, that you get larger PCF values when the LWC is lower as compared to larger LWC values. LWC is a variable that can be used to determine whether entrainment is occurring, but the actual clustering amount depends on the physical strength of the turbulence that is being generated by the entrainment, not the value of the LWC, since LWC values are relative to individual clouds, and change from case to case.

Although the value of the LWC acts as an indicator for entrainment, and LWC values that are smaller result in larger clustering values for both center and entrainment zones as compared to higher LWC values, one must be careful due to the correlation of LWC and the number of drops used in the PCF. Figure 3.22 shows the LWC as a function of the number of drops used in the PCF, with L1 in blue, L2 in light blue, H1 in red, and H2 in light red. As can be seen in Figure 3.22, as the number of drops used increases, so too does the LWC. This is as expected, since the LWC is calculated by using the number of droplets present (Wallace and Hobbs 2006). The R–squared values for the linear best fit lines are 0.34, 0.41, 0.13, and 0.27 for L1, L2, H1, and H2, respectively. However, this creates a conundrum of whether the PCF increases due to the lower LWC or the lower number of drops used. It is pretty clear that PCF values vary at a lower number of drops used from Figure 3.22. The larger PCF values calculated for the number of droplets used in the PCF that are less than 2000 as seen

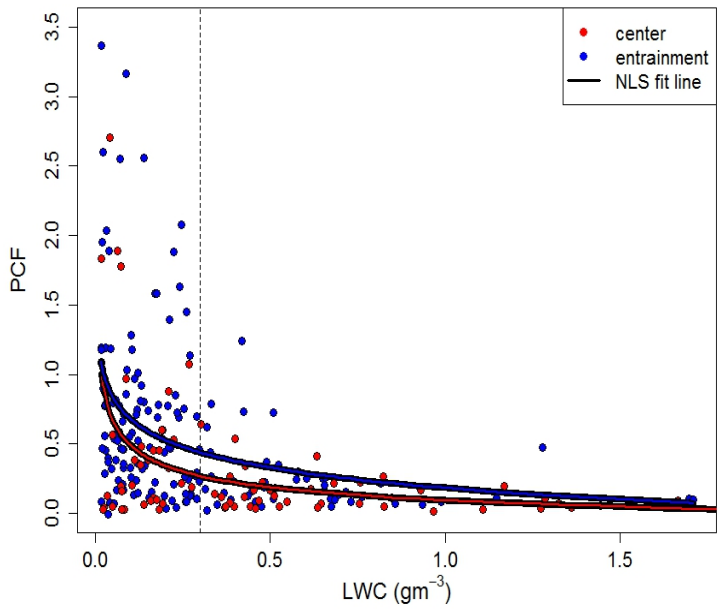

Figure 3.21: Shows LWC (gm^{-3}) on the x–axis and the PCF on the y–axis, with center data in red and entrainment data in blue. Each dataset is fit with a NLS fit line which matches the color of the dataset that they correspond too. A dashed line is present at 0.3 gm^{-3} for reference.

in Figure 3.19 was determined to be variation that still captured the overall physical amount of clustering. We can now see that the larger PCF values are either due to variation, or the fact that the lower LWC is a result of entrainment mixing, which is expected to have larger values of clustering due to the production of TKE. Although there is no way to tell whether the larger PCF values are due to the number of drops used or the LWC, the fact that the LWC and the number of drops used are correlated help to add confidence to the results that the clustering calculated is not a result of the length of the data, but a result of the physical clustering that is occurring within the data.

Note that the different slopes for each best fit line in Figure 3.22 is a result of the unique cloud characteristics for each flight, where each flight has different average LWCs and number of drops used. For example, L1 has the largest slope because it has the largest average LWC and lowest average number of drops.

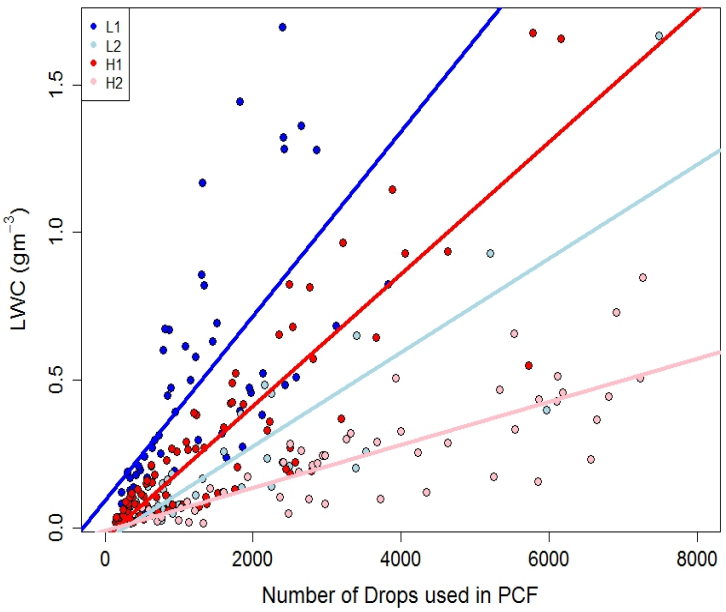

Figure 3.22: Shows number of drops used in the PCF on the x–axis and the LWC (gm^{-3}) on the y–axis, with L1 in blue, L2 in light blue, H1 in red, and H2 in light red. The best fit linear lines match the color of the flight that they correspond too.

CHAPTER 4
SUMMARY

4.1 Conclusions

Flight data obtained from the CIRPAS Twin otter aircraft flown during the GoMACCS campaign near Houston, TX from 2006 were used to investigate 81 non–precipitating cumulus clouds to better understand how droplet clustering changes as a function of cloud location (through both vertical and horizontal extent) and aerosol number concentration. Out of the 22 flights flown during the GoMACCS campaign, two low pollution (L1, L2) and two high pollution (H1, H2) flights were selected to analyze how droplet clustering changed as a function of aerosol number concentration. A Case flight was chosen based on a single, vertically developed cloud that would allow for analysis based on clustering changes as a function of altitude. The **first hypothesis** stated in Section 2.7 that clustering changes as a function of cloud center and cloud edge was validated by taking two second cloud chunks for cloud edge (entrainment zone) and cloud center. The **second hypothesis** that clustering changed as a function of cloud aerosol number concentration was investigated by analyzing low and high pollution clouds that had mean aerosol number concentrations of 1814 and 4491 cm^{-3}, respectively.

It was discussed in Chapter 1 that evidence of inertial clustering was already abundant, and the purpose of this study was not to prove inertial clustering, but to understand how it changes within a cumulus cloud and with different cloud characteristics. The main findings reported in this study can be summarized as follows:

- Droplet clustering is enhanced at smaller temporal lags, which is in agreement with findings from Kostinski and Shaw (2001); Shaw (2003); Kostinski and Jameson (2000); Larsen (2007) and Shaw et al. (2002).

- Droplet clustering is enhanced for cloud entrainment zones as compared to cloud center zones, with a statistically significant difference at the 5 percent level, confirming the hypothesis that droplet clustering is enhanced in the entrainment zone due to the production of TKE through evaporational cooling (Shaw 2003). Although this result is hypothesized, direct calculations have never been made before.

- There is no statistical difference at the 5 percent level for droplet clustering between low pollution and high pollution clouds. Although it was found that low pollution clouds do, on average, have a larger amount of clustering in both the center and entrainment zones, this is due entirely to the L2 flight. The hypothesis that clustering changes as a function of aerosol number concentration cannot be confirmed.

- L1, H1, and H2 have a statistically similar amount of clustering at the 5 percent level, while L2 has a larger, statistically significant amount of clustering. It is proposed that cloud age plays an important role in the amount of clustering that is occurring, with decaying clouds demonstrating a higher amount of clustering as compared to developing clouds, as confirmed through LWC, RH, CDNC, vertical velocity, and cloud width which are in agreement with the characteristics that are shared with decaying clouds from Schmeissner et al. (2015). Buoyancy is in disagreement with the characteristics of decaying clouds, suggesting that more research must be done.

- Droplet clustering sees no changes as a function of height, even though it is shown in Small et al. (2013) that turbulence increases with height. There is significantly enhanced droplet clustering in the cloud top entrainment zone (indicated by decreased RH and LWC values) due to the generation of TKE though evaporational cooling, as suggested in Shaw (2003). It can therefore be stated that turbulence does change as a function of cloud height, but only in the top portion of the cloud where cloud top entrainment is occurring.

4.2 Future Work

The results presented here have a vast potential to be built upon, particularly in the discussion of the effects that turbulence and droplet clustering have on collision–coalescence. According to models, collision between equal sized particles may occur as a result of turbulent shear (Pruppacher and Klett 1997). The overall goal is to generate a model that can accurately demonstrate collision–coalescence using the results presented here, however, more work must first be done to better understand the results presented. Other goals include understanding how aerosol-cloud interactions such as the cloud lifetime and albedo effects are affected by droplet clustering. Although droplet clustering doesn't change as a function of aerosol number concentration, understanding how the clustering can change the albedo and lifetime of a cloud through modeling are of importance to better understand

cloud microphysics and lead to better implementation of cloud microphysical parameters in GCMs.

Future work that must be done includes:

- Gathering more data, either from the GoMACCS campaign, or other campaigns from the past/future to replicate these results and see if a trend can be found in low pollution and high pollution clouds. More replication will hopefully lead to an understanding of what the most important determining factor is in the amount of clustering that is present in a cumulus cloud. Understanding whether the determining factor is cloud age can only be found with more data collection.

- More Case flights need to be analyzed to find whether or not patterns can be found in the vertical profile of clustering in cumulus clouds. There is confidence in the increased clustering amount relating to cloud top entrainment. Understanding of other processes such as the convergence of the cloud base updrafts and cloud top downdraft, and TKE production at cloud base from condensational heating will take more research.

- Measurements need to be taken throughout the middle of single shallow cumulus clouds from the cloud genesis to the cloud termination time. Although this type of measurement comes with complications, the right tools (such as attaching a PDI to an industrial grade drone) could make it possible.

Although the work presented here gives one a strong base for understanding how clustering changes throughout the horizontal and vertical profile of a cloud, and why that clustering might change given individual cloud characteristics such as aerosol number concentration and age, to build on these results will take replication. Future observations and replication may provide key information to further illuminate the results presented in this study.

APPENDIX A
ADDED FIGURES

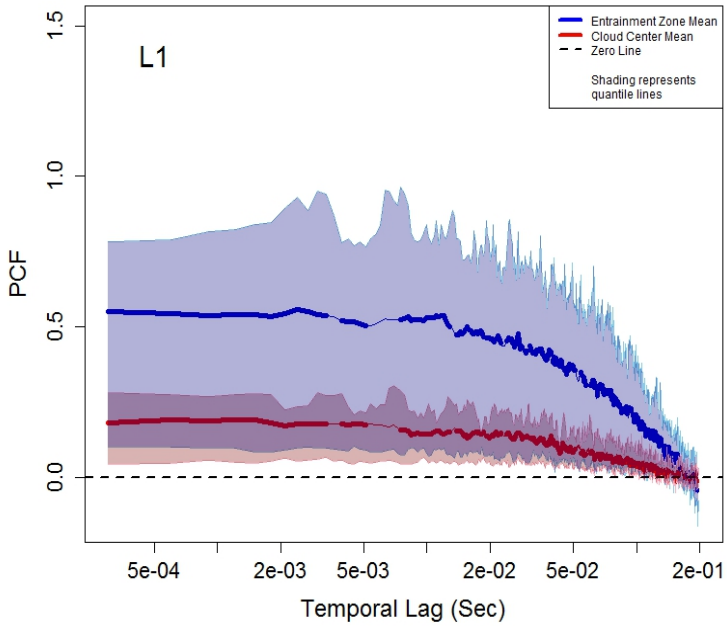

Figure A.1: Shows temporal lag (sec) on the x–axis and the PCF value (unitless) on the y–axis for the L1 flight, with entrainment zone data in blue and cloud center data in red. The envelopes represent the 85^{th} percent quantile value (top) and 15^{th} percent quantile value (bottom) of the data. The middle line in each envelope represents the mean PCF value, where a bold mean line represents that entrainment and center differences are statistically significant.

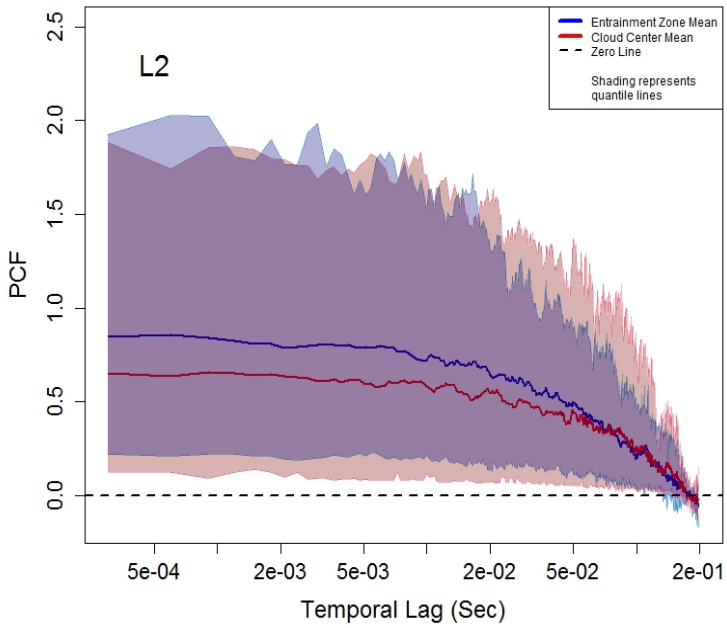

Figure A.2: As in Figure A.1, except for the L2 flight.

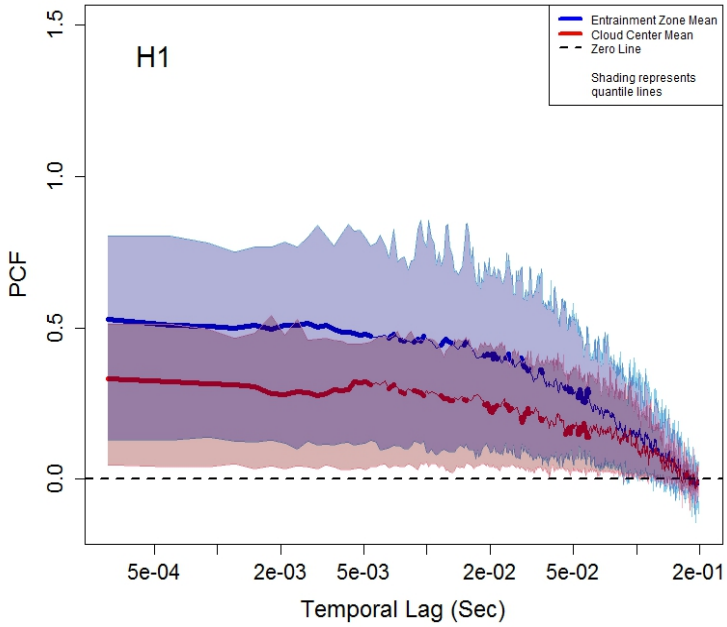

Figure A.3: As in Figure A.1, except for the H1 flight.

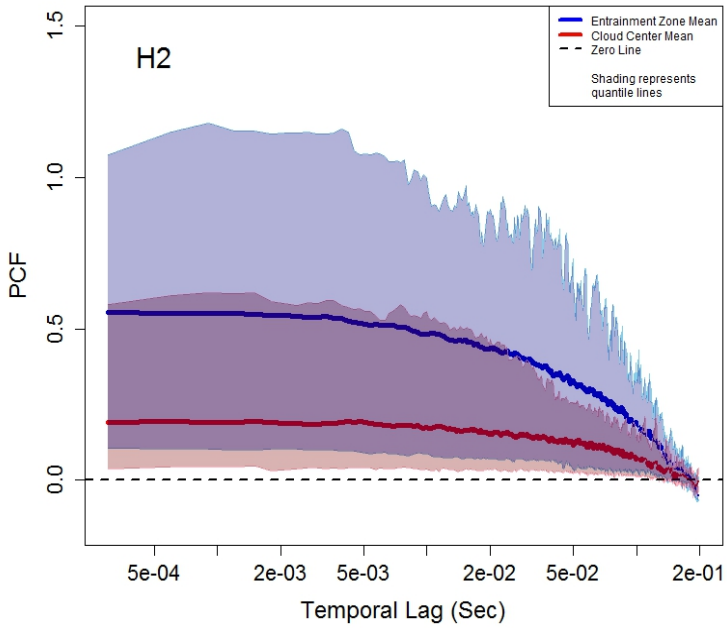

Figure A.4: As in Figure A.1, except for the H2 flight.

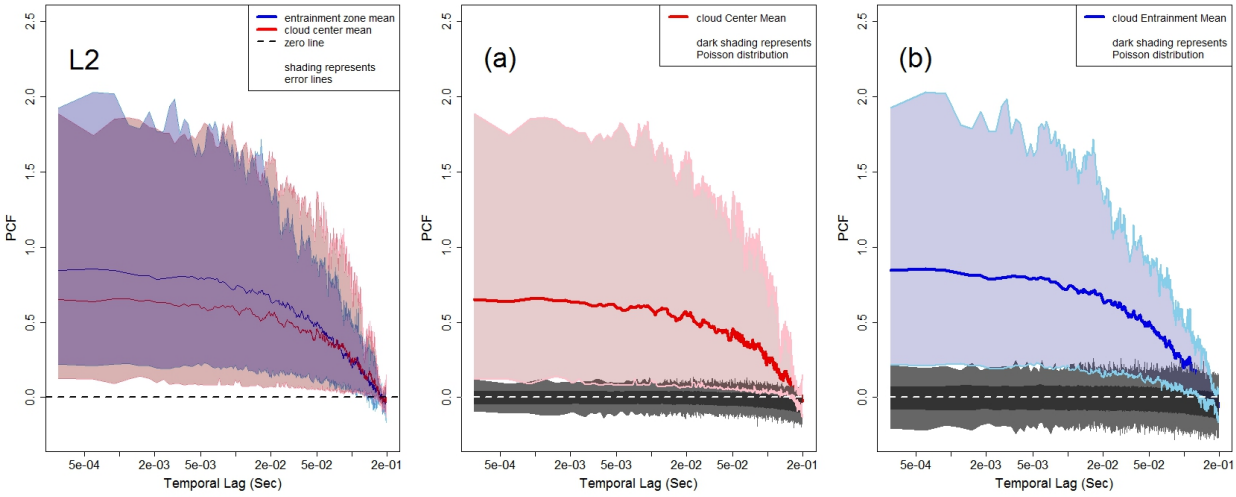

Figure A.5: Shows the L2 panel from Figure A.2 on the far left. Panel (a) shows the center clustering from the L2 flight, with the maximum and minimum values from the Poisson simulations in light grey, and the 100^{th} largest and smallest values from the Poisson simulations (80 percent confidence interval) in dark gray. Panel (b) shows the same thing as Panel (a), except for the entrainment data.

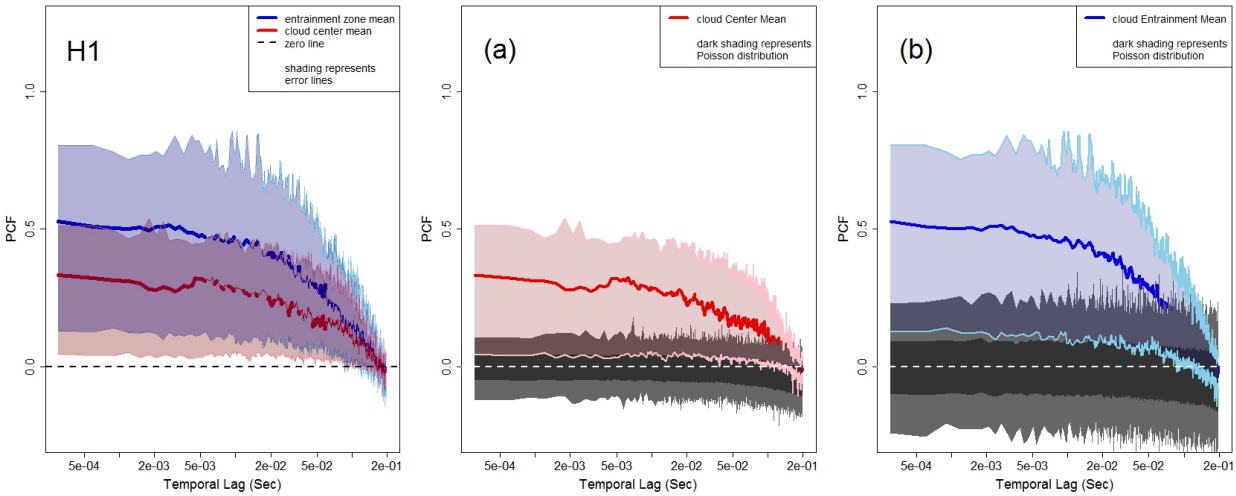

Figure A.6: As in Figure A.5, except for the H1 flight.

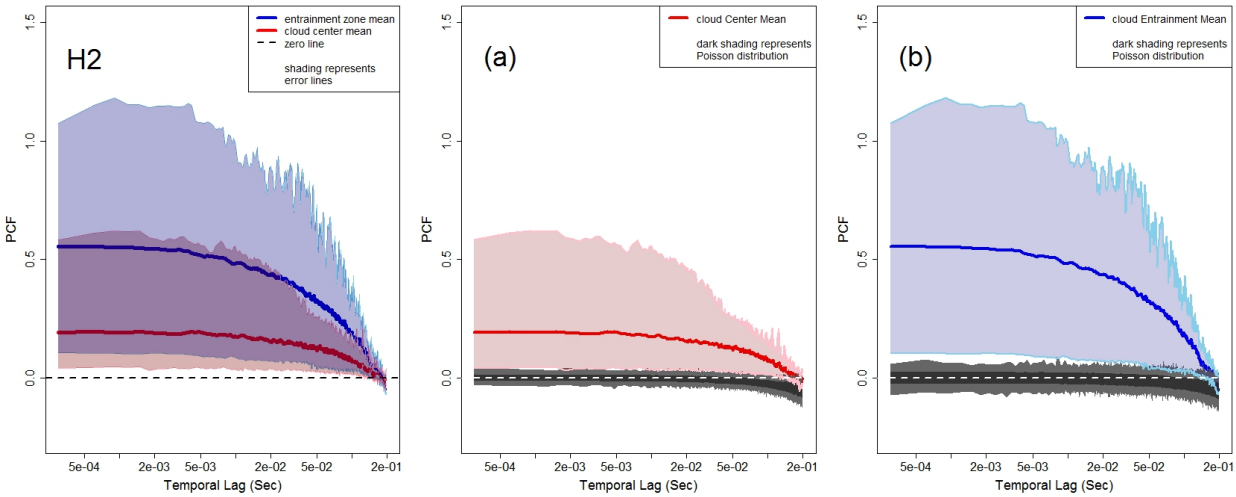

Figure A.7: As in Figure A.7, except for the H2 flight.

BIBLIOGRAPHY

Ahrens, C. D., 2008: *Essentials of Meteorology: An invitation to the atmosphere*. 5th ed., Thonson Brooks/Cole, 485 pp.

Albrecht, H. E., M. Borys, N. Damaschke, and C. Tropea, 2002: *Laser Doppler and Phase Doppler Measurement Techniques*. Springer–Verlag, 660 pp.

Anderson, A. and A. Kostinski, 2010: Reversible record breaking and variability: Temperature distributions across the globe. *J. Appl. Meteor. Climatol.*, **49**, 1681–1691.

Bachalo, W. D. and M. J. Houser, 1984: Phase doppler spray analyzer for simultaneous measurements of drop size and velocity distributions. *Optical Engineering*, **5**.

Baker, B., 1992: Turbulent entrainment and mixing in clouds: A new observation approach. *J. Atmos. Sci.*, **49**, 387–404.

Baker, B. A. and R. P. Lawson, 2010: Analysis of tools used to quantify droplet clustering in clouds. *J. Atmos. Sci.*, doi:10.1175/2010JAS3409.1.

Baker, M., R. E. Choularton, and J. Latham, 1984: The effects of turbulent mixing in clouds. *J. Atmos. Sci*, **106**, 581–598.

Blyth, A., 1993: Entrainment in cumulus clouds. *J. Appl. Meteor.*, **32**, 626–640.

Borovoi, A., 2002: On the extinction of radiation by a homogeneous but spatially correlated random medium: comment. *J. Optical Soc. of America A*, **19**, 2517–2520.

Chorin, A. J., 1968: Numerical solution of the navier-stokes equations. *Mathematics of Computation American mathematical Society*, **22**, 745–762.

Chuang, P. Y., E. W. Saw, J. D. Small, R. A. Shaw, C. M. Sopperley, G. A. Payne, and W. D. Bachalo, 2008: Airborne phase doppler interferometry for cloud microphysical measurements. *Aerosol Sci. Tech.*, **42**, 685–703.

Conant, W., et al., 2004: Aerosol-cloud drop concentration closure in warm cumulus. *J. Geophy. Res.*, **109**.

Cooper, W. A. and R. P. Lawson, 1984: Physical interpretation of results from hiplex–1 experiment. *J. Climate Appl. Meteor.*, **23**, 532–540.

Dimotakis, P. E., 2004: Turbulent mixing. *Annu. Rev. Fluid Mech.*, **37**, 329–356.

Eaton, J. K. and J. R. Fessler, 1994: Preferential concentration of particles by turbulence. *Int. J. Multiphase Flow*, **20**, 169–209.

Falkovich, G., A. Fouxon, and M. G. Stepanov, 2002: Acceleration of rain initiation by cloud turbulence. *Nature*, **419**, 151–154.

Feingold, G., A. S. Frisch, B. Stevens, and W. Cotton, 1999: On the relationship among cloud turbulence, droplet formation and drizzle as viewedd by doppler radar, microwave radiometer, and lidar. *J. Geophys. Res.*, **104**, 22 195–22 203.

Flugge, W., 1972: *Tensor analysis and continuum mechanics*. Springer, 201 pp.

Frisch, U., 1995: *Turbulence*. Cambridge University Press, 289 pp.

Gerber, H., B. G. Arends, and A. S. Ackerman, 1994: New microphysics sensor for aircraft use. *Atmos. Res.*, **31**, 235–252.

Ghosh, S., J. Davila, J. C. Hunt, A. Srdic, H. J. Fernando, and P. R. Jonas, 2005: How turbulence enhances coalescence of settling particles with applications to rain in clouds. *Proc. Roy. Soc. London*, **416**, 3059–3088.

Granger, R. A., 1995: *Fluid Mechanics*. Dover Publications, Inc, 896 pp.

Hill, G. E., 1973: Factors controlling the size and spacing of cumulus clouds as revealed by numerical experiments. *J. Atmo. Sci*, **31**, 646–673.

Holton, J. R. and G. J. Hakim, 2013: *An introduction to dynamic Meteorology*. 5th ed., World Scientific Publishing, 532 pp.

Jonas, P., 1996: Turbulence and cloud microphysics. *Atmos. Res.*, **40**, 283–306.

Khain, A., M. Ovtchinnikov, M. Pinsky, A. Pokrovsky, and H. Krugliak, 2000: Notes on the state–of–the–art numerical modeling of cloud microphysics. *Atmos. Res.*, **55**, 159–224.

Kitchen, M. and S. J. Caughey, 1981: Tethered–balloon observed observations of the structure of small cumulus clouds. *Q.J.R. Meteorol. Soc.*, **107**, 853–874.

Klazura, G. E., 1971: Measurements of precipitation particles in warm cumuli over southeast texas. *J. Appl. Meteor.*, **10**, 739–750.

Knight, C. A., J. Vivekanandan, and S. G. Lasher-Trapp, 2002: First radar echoes and the early zdr history of florida cumulus. *J. Atmos. Sci.*, **297**, 1667–1670.

Kogan, Y. L., 1993: Drop size separation in numerically simulated convective clouds and its effect on warm rain formation. *J. Atmos. Sci.*, **50**, 1238–1253.

Kolmogorov, A. N., 1941: The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. *Dokl. Akad. Nauk SSSR*, **30**, 301–305.

Koren, I., O. Altaratz, and G. Dagan, 2015: Aerosol effect on the mobility of cloud droplets. *Environ. Res. Lett.*, **10**, 1–10.

Kostinski, A., M. Larsen, and A. Jameson, 2006: The texture of rain: Exploring stochastic microstructure at small scales. *J. Hydrol.*, **328**, 38–45.

Kostinski, A. B., 2002: On the extinction of radiation by a homogeneous but spatially correlated random medium: reply to comment. *J. Optical Soc. of America A*, **19**, 2521–2525.

Kostinski, A. B. and R. A. Jameson, 2000: On the spatial distribution of cloud particles. *J. Atmos. Sci*, **57**, 901–915.

Kostinski, A. B. and R. A. Shaw, 2001: Towards quantifying droplet clustering in clouds. *Quarterly J. of the Meteor. Soc.*, **128**, 1043–1057.

Kostinski, A. B. and R. A. Shaw, 2005: Fluctuations and luck in droplet growth by coalescence. *Bull. Am. Meteor. Soc.*, **86**, 235–244.

Laird, N. F., H. T. Ochs III, R. M. Rauber, and L. J. Miller, 2000: Initial precipitation formation in warm florida cumulus. *J. Atmos. Sci.*, **57**, 3740–3751.

Langley, R., 1971: *Practical Statistics Simply Explained*. Dover Publisher, 399 pp.

Langmuir, I., 1948: *J. Atmos. Sci.*, **5**, 175–192.

Larsen, M., 2006: Studies of discrete fluctuations in atmospheric phenomena. Ph.D. thesis, Michigan Technical University, 220 pp.

Larsen, M., 2007: Spatial distributions of aerosol particles: Investigation of the poisson assumption. *Aerosol Science*, **38**, 807–822.

Larsen, M., 2012: Scale localization of cloud particle clustering statistics. *J. Atmos. Sci.*, **69**, 3277–3289.

Larsen, M. and A. Kostinski, 2005: Observations and analysis of uncorrelated rain. *A. Meteor. Soc.*, **62**, 4071–4083.

Lau, L. M. and H. T. Wu, 2003: Warm rain processes over the tropical oceans and climate implications. *Geophy. Res. Lett.*, **30**, doi:10.1029/2003GL018567.

Lu, M., H. Jonsson, P. Chuang, G. Feingold, and R. Flagan, 2008: Aerosol–cloud relationships in continental shallow cumulus. *J. Geophy. Res.*, **113**, D15 201.

Mann, H. B. and D. R. Whitney, 1947: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, **18**, 50–60.

Marsh, C. and J. Elliott, 2008: *Exploring Data: an introduction to data analysis for social scientists*. Polity Press, 299 pp.

Marshak, A., K. Y., M. L. Larsen, and W. J. Wiscomebe, 2004: Small-scale drop-size variability: Empirical models for drop-size dependent clustering in clouds. *J. of the Atmospheric Sciences*, **62**, 551–558.

Matsuda, K., R. Onishi, R. Kurose, and S. Komori, 2012: Turbulence effect on cloud radiation. *Phys. Rev. Lett.*, **108**, 1–4.

Maxey, M. R., 1987: The gravitational settling of aerosol particles in homogeneous turbulence and random flow fields. *J. Fluid Mech.*, **174**, 441–465.

Mazin, I. P., V. I. Silaeva, and M. A. Strunin, 1984: Turbulent fluctuations of horizontal and vertical wind velocity components in various cloud forms. *Izv., Atmos. Oceanic Phys.*, **20**, 6–12.

McIlveen, J. F. R., 1986: *Basic meteorology: a physical outline*. Acford Ltd., 457 pp.

Mei, C. C. and B. Vernescu, 2010: *Homogenization methods for multiscale mechanics*. Academic Press, 330 pp.

Mei-Yuan, H., 1963: Microstructure of cumulus clouds. *Izv. Akad. Nauk. SSR.*, **2**, 632–376.

Meischner, P. F., T. Jank, R. Baumann, and H. Holler, 2001: Eddy dissipation rates in thunderstorms estimated by doppler radar compared with aircraft in situ measurements. *J. Atmos. and Oceanic Tech.*, **18**, 1609–1627.

Picinbono, B. and C. Bendjaballah, 2005: Characterization of nonclassical optical fields by photodetection statistics. *Physical Rev. A*, **71**, 1–12.

Pilewskie, P., et al., 2003: Solar spectral radiative forcing during the southern african regional science initiative. *J. Geo phys. Res.*, **108**, 8486.

Pinsky, M. and A. Khain, 1996: Turbulence effect on the collision kernal i: Formation of velocity deviations of drops falling within a turbulent three dimensional flow. *Quart. J. Roy. Meteor. Soc.*, **123**, 1517–1542.

Pinsky, M., A. Khain, and M. Shapiro, 1999: Collision of small drops in a turbulent flow. part i: Collision efficiency. problem formulation and preliminary results. *J. Atmos. Sci.*, **56**, 2585–2600.

Pinsky, M. B., A. Khain, and M. Shapiro, 2006: Collisions of small drops in a turbulent flow. part iv: Droplet hydrodynamic interactions. *J. Atmos. Sci.*, **64**, 2462–2482.

Politovich, M. K., 1993: A study of the broadening of droplet size distributions in cumuli. *J. Atmos. Sci.*, **50**, 2230–2244.

Pope, S. B., 2000: *Turbulent Flows*. Cambridge Univ. Press, 771 pp.

Pruppacher, H. and J. Klett, 1997: *Microphysics of clouds and precipitation*. 2d ed., Kluwer Academic Publishers, 954 pp.

Pruppacher, J. R. and J. D. Klett, 1980: *Microphysics of clouds and precipitation*. D. Reidel Publishing, 714 pp.

Ramaswamy, V., et al., 2001: *Radiative forcing of climate change, in: Climate Change: 2001: The Scientific Basis. Contribution of working group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge Univ. Press, 476-483 pp.

Reade, W. and L. Collins, 2000: Effect of preferential concentration on turbulent collision rates. *Physics of Fluids*, **12**, 2530–2540.

Roberts, G. C. and A. Nenes, 2005: A continuous–flow streamwise thermal–gradient ccn chamber for atmospheric measurements. *Aerosol. Sci. Technol.*, **39**, 206–221.

Rogers, R. R. and M. K. Yau, 1989: *A short course in cloud microphysics*. 3d ed., Pergamon Press, 290 pp.

Rosenfeld, D., R. Lahav, K. A., and M. Pinsky, 2002: The role of sea spray in cleansing air pollution over ocean via cloud processes. *Science*, **297**, 1667–1670.

Rossow, W. B. and R. A. Schiffer, 1999: Advances in understanding clouds from isccp. *Bull. Am. Meteor. Soc.*, **80**, 2261–2287.

Saw, E. W., 2008: Studies of spatial clustering of inertial particles in turbulence. Ph.D. thesis, Michigan Technological University, 142 pp.

Saw, E. W., R. A. Shaw, S. Ayyalasomayajula, P. Y. Chuang, and A. Gylfason, 2008: Inertial clustering of particles in high-reynolds–numer turbulence. *Phys. Rev. Lett.*, **100**, 214 501.

Schmeissner, T., R. A. Shaw, J. Ditas, F. Strathmann, M. Wendisch, and H. Siebert, 2015: Turbulent mixing in shallow trade wind cumuli: dependence on cloud life cycle. *J. Atmos. Sci.*, **72**, 1447–1465.

Shaw, R. A., 2003: Particle–turbulence interactions in atmospheric clouds. *Annu. Rev. Fluid Mech.*, **35**, 183–227.

Shaw, R. A., A. B. Kostinski, and M. L. Larsen, 2002: Towards quantifying droplet clustering in clouds. *Quarterly J. of the Meteor. Soc.*, **128**, 1043–1057.

Shaw, R. A., W. C. Reade, L. R. Collins, and J. Verlinde, 1998: Preferential concentration of cloud droplets by turbulence: Effects on the early evolution of cumulus cloud droplet spectra. *J. of the Atmos. Sci.*, **55**, 1965–1976.

Siebert, H., S. Gerashchenko, A. Gylfason, K. Lehmann, L. R. Collins, R. A. Shaw, and Z. Warhaft, 2010: Toward understanding the role of turbulence on droplets in clouds: in situ and laboratory measurements. *Atmos. Res.*, **97**, 426–437.

Small, D. J., P. Y. Chuang, and H. H. Jonsson, 2013: Microphysical imprint of entrainment in warm cumulus. *Tellus B*, **65**, 1–18.

Small, J. D., P. Y. Chuang, G. Feingold, and H. Juang, 2009: Can aerosol decrease cloud lifetime? *Geophy. Res. Let.*, **36**, 1–5.

Smith, S. A. and P. R. Jonas, 1995: Observations of the turbulent fluxes in fields of cumulus clouds. *Z.J.R. Meteorol. Soc.*, **85**, 144–151.

Sorooshian, A., F. J. Brechtel, Y. L. Ma, R. J. Weber, R. C. Corless, and J. H. Seinfeld, 2006: Modeling and characterization of a particle–into–liquid sampler. *Aerosol Sci. Technol.*, **40**, 396–409.

Srivastava, R. C., 1989: Growth of cloud droplet by condensation: A criticism of currently accepted theory and a new approach. *J. Atmos. Sci.*, **46**, 869–887.

Stein, A. F., R. R. Draxler, G. D. Rolph, B. J. Stunder, M. D. Cohen, and F. Ngan, 2015: Hysplit atmospheric transport and dispersion modeling system. *Bull. Amer. Meteor. Soc.*, **96**, 2059–2077.

Stull, R. B., 1988: *An introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, 666 pp.

Sundaram, S. and L. R. Collins, 1997: Collision statistics in an isotropic particle-laden turbulent suspension. part i. direct numerical simulations. *J. Fluid. Mech.*, **335**, 75–109.

Szumowski, M. J., R. M. Rauber, H. T. Ochs III, and L. J. Miller, 1997: The microphysical structure and evolution of hawaiian rainband clouds. part i: Radar observations of rainbands containing high reflectivity cores. *J. Atmos. Sci.*, **48**, 112–121.

Tennekes, J. and J. L. Lumley, 1972: *A first course in Turbulence*. The MIT press, 320 pp.

Vaillancourt, P. A., M. K. Yau, and P. Bartello, 2002: Microscopic approach to cloud droplet growth by condensation. part ii: Turbulence, clustering, and condensational growth. *J. of the Atmos. Sci.*, **59**, 3421–3435.

Vali, G., 1996: *Ice nucleation: a review*. Pergamon Press, 271–279 pp.

Villermaux, E., C. Innocenti, and J. Duplat, 2001: Short circuits in the corrsin-ubukhov cascade. *Physics of Fluid*, **13**, 284–289.

Wallace, J. and P. Hobbs, 2006: *Atmospheric Science: An introductory survey*. 2d ed., Academic Press, 504 pp.

Wang, L. P., A. S. Wexler, and Y. Zhou, 1998: Statistical mechanical descriptions of turbulent coagulation. *Physics of Fluids*, **10**, 869–887.

Wang, L. P., Y. Xue, O. Ayala, and W. W. Grabowski, 2005: Effects of stochastic coalescence and air turbulence on the size distribution of cloud droplets. *Atmos. Res.*, **82**, 416–432.

Warner, J., 1955: The water content of cumuliform clouds. *Tellus.*, **7**, 449–457.

Warner, J., 1969: The microstructure of cumulus clouds. part i. general features of the droplet. *J. Atmos. Sci.*, **26**, 1049–1056.

Warren, S. G. and C. J. Hahn, 2002: *Cloud climatology, in: Encyclopedia of Atmospheric Science*. Oxford University Press, 476–483 pp.

Welch, P. D., 1967: The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions and Audio and Electracoustics*, **2**, 70–73.

Wendisch, M., et al., 1998: Drop size distribution and lwc in po valley fog. *Contr. Atmos. Phys.*, **71**.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, International Geophysics Series, Vol. 100. 3d ed., Academic Press, 676 pp.

Zaitsev, V. A., 1950: Liquid water content and distribution of drops in cumulus clouds. *Trudy Glavnoi Geofiz. Observ.*, **19**, 122–132.

Zhou, Z., A. S. Wexler, and L. P. Wang, 2001: Modeling turbulent collision of bidisperse inertial particles. *J. Fluid Mech.*, **433**, 77–104.