Statistical comparison of properties of simulated and observed cumulus clouds in the vicinity of Houston during the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS)

Hongli Jiang,[1] Graham Feingold,[2] Haflidi H. Jonsson,[3] Miao-Ling Lu,[4] Patrick Y.

Chuang,[5] Richard C. Flagan,[4] John H. Seinfeld[4]

———

H. Jiang CIRA/NOAA Earth System Research Laboratory, Boulder, Colorado, 80305.
(Hongli.Jiang@noaa.gov)

G. Feingold NOAA Earth System Research Laboratory, 325 Broadway, Boulder, Colorado 80305.
(Graham.Feingold@noaa.gov)

[1]Cooperative Institute for Research in the

Atmosphere/NOAA Earth System Research

Laboratory, Boulder, CO 80305.

[2]NOAA Earth System Research Laboratory,

Boulder, CO 80305.

[3]Center for Interdisciplinary

Remotely-Piloted Aircraft Studies, Naval

Postgraduate School, Monterey, CA.

[4]California Institute of Technology,

Pasadena, CA.

Abstract.   We present statistical comparisons of properties of clouds generated by Large Eddy Simulations (LES) with aircraft observations of non-precipitating, warm cumulus clouds made in the vicinity of Houston, TX during the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS), carried out in the summer of 2006. Aircraft data were sampled with the Center for Interdisciplinary Remotely-Piloted Aircraft Studies (CIRPAS) Twin Otter airplane. Five flights (days) that are most suitable for studying aerosol-cloud interactions are selected from the 22 flights. The model simulations are initiated with observed environmental profiles. The simulations are used to generate an ensemble of thousands of cumulus clouds for statistically meaningful evaluations. Statistical comparisons focus on the properties of a set of dynamical and thermodynamical variables, sampled either in the cloud or the cloud updraft core. The set of variables includes cloud liquid water content (LWC), number mixing ratio of cloud droplets (N_d), cloud effective radius (r_e), updraft velocity (w), and the distribution of cloud sizes. In general, good agreement between the simulated and observed clouds is achieved in the normalized frequency distribution functions, the profiles averaged over the cloudy regions, the cross-cloud averages, and the cloud size distributions, despite big differences in sample size between the model output and the aircraft data. Some unresolved differences in frequency distributions of w and possible differences in cloud fraction are noted.

[5]University of California, Santa Cruz, CA.

These comparisons suggest that the LES is able to successfully generate the cumulus cloud populations that were present during GoMACCS. The extent to which this is true will depend on the specific application.

1. Introduction

Shallow, warm cumulus clouds have been the focus of much study because of their role in transporting heat and moisture to the free troposphere. They are also becoming recognized as important to climate because they modify the planetary albedo. This is particularly true over oceans where their albedo contrast with the underlying (dark) ocean is large, but their prevalence over land makes them of more general importance. In polluted urban areas these clouds are susceptible to modification by aerosol with significant implications for their albedo [*Twomey*, 1974] and ability to generate precipitation [*Warner*, 1968]. Over land and ocean they have a frequency of occurrence of 14 % and 33 %, respectively (global average); cloud cover ranges from 15 % in the trade wind regime to 20 % over land [*Warren et al.*, 1986, 1988].

Using aircraft, surface-based, and satellite data, studies have addressed characteristics of cumulus cloud populations such as size distribution, aspect ratio, and distance between nearest neighbors for fair weather cumulus over the southern part of the US [e.g. *Plank*, 1979; *Wielicki and Welch*, 1986], and over ocean [e.g. *Cahalan and Joseph*, 1989; *Joseph and Cahalan*, 1990; *Benner and Curry*, 1998; *Zhao and Di Girolamo*, 2007; and others]. Earlier cumulus cloud modeling efforts addressed single clouds [e.g., *Clark*, 1973] but in recent years Large Eddy Simulations (LES) have been used to simulate fields of cumulus clouds for idealized cases (e.g., the Global Energy and Water Experiment (GEWEX) Cloud System Study (GCSS) inter-comparison projects) both over land [*Brown et al.*, 2002] and ocean [e.g. *Siebesma et al.*, 2003]. *Neggers et al.* [2003a] and *Xue and Feingold* [2006] studied the statistics of the sizes of cumulus cloud populations generated by Large Eddy Simulations (LES) of boundary layer cases, but did not compare with observations. *Neggers et al.* [2003b] performed LES of shallow cumulus

convection and directly evaluated their results against in-cloud aircraft measurements during the Small Cumulus Microphysics Study. Their goal was to evaluate turbulent processes, the distribution of the bulk entrainment rate, and to develop parameterizations of shallow cumulus cloud processes. Recently, *Abel and Shipway* [2007] performed cloud resolving simulations of trade wind cumulus and compared selected model results with aircraft data collected in the south west Atlantic trade wind cumulus regime, with emphasis on evaluating some of the microphysical schemes in their model. Although these individual studies had different objectives, the common goal is to establish that LES is a credible tool to understand, evaluate, and parameterize processes related to cloud development. The field of statistical comparison and evaluation of LES model output against observations is still in its infancy and there is much need for refining and exploring new methodologies.

The question arises: how realistic are cumulus cloud populations generated by LES compared to observations? Direct comparison between LES-generated clouds and observations is limited due to a dearth of data at similar spatial and temporal scales. Moreover, methods of comparing observations to modeled cloud fields are not very well established. Satellite remote sensing from platforms such as the Geostationary Operational Environmental Satellite (GOES) has the advantage of providing reasonable temporal resolution of a cloud field (15 min); however, at a size resolution of 1 km, does not detect many of the prevalent small clouds. Cloud measurements from instruments such as the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Aqua and Terra polar-orbiting platforms provide higher resolution (250 m) images of clouds on a global scale, but are temporally limited to snapshots once or twice a day. The very high spatial resolution (15 m) Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images suffer from similar temporal sampling limitations. The close

match between typical temporal/spatial scales of aircraft measurements (1 s; 50 - 100 m) and LES (1 s; 50-100 m) warrant closer investigation of comparisons between the characteristics of observed and simulated cloud fields.

During the summer of 2006, the Texas Air Quality Study/Gulf of Mexico Atmospheric Composition and Climate Study (henceforth GoMACCS) was carried out in the vicinity of Houston, Texas. An instrumented aircraft, the Center for Interdisciplinary Remotely-Piloted Aircraft Studies (CIRPAS) Twin Otter, outfitted with a suite of aerosol size and composition instruments, cloud drop probes, aerosol optical measurements and irradiance measurements, performed a total of 22 flights between August 21 and September 15. During the mid-morning, shallow cumulus clouds were prevalent in the boundary layer with occasional development of deep convective storms in the early afternoon. The aircraft focused on non-precipitating clouds and avoided deep convection. The region is characterized by strong pollution from various industries and urban sources. This type of boundary layer provides an excellent environment for studying aerosol (pollution) - cloud interactions and for comparison between model results and aircraft data.

We have performed 3-dimensional large-eddy simulations of 5 cases chosen based on prevalence of warm convective clouds and data availability during GoMACCS, and compared the model output with aircraft data. By generating ensembles of boundary layer cumulus clouds for multiple cases, we provide a sense of the degree of variability between different cases and enable comparison with aircraft data on individual days as well as in the ensemble-mean sense. Cumulus cloud fields tend to be characterized by a large range of cloud depths, liquid water contents and updraft velocities, which requires that model/observation comparisons be performed in a statistical manner rather than on a single cloud basis. Aircraft sampling strategy was designed with this in mind, thus facilitating the comparisons to be shown below.

The goal of this study is to address the following questions:

(i) How well does LES simulate fields of non-precipitating cumulus clouds observed over a highly polluted urban area?

(ii) What methodology and statistical analyses can be used to compare the model results to aircraft measurements?

(iii) How do aerosol concentrations change the cloud microphysics?

We will show that the LES is able to generate cumulus clouds that are comparable to those observed under a wide range of aerosol concentrations. Aerosol effects on cloud microphysics will be shown to be of similar order to those observed. The methods of analysis will provide some insight into the challenges of these comparisons.

2. Model Description

The model is a large eddy simulation based on the Regional Atmospheric Modeling System [RAMS, version 6.0, *Cotton et al.*, 2003] coupled to a microphysical model described by *Feingold et al.*, [1996], and a surface model [*Walko et al.*, 2000]. A brief description of each module is given below.

2.1. Bin Microphysical Model

The model includes a size-resolved representation of cloud drops. The size-distribution is divided into 33 size-bins, covering the drop range 1.56 μm; 2.54 mm (radius) with mass-doubling from one size bin to the next. Warm cloud processes, including activation, condensation/evaporation, collision-coalescence, and sedimentation, are solved using the method of moments based on *Tzivion et al.* [1987]. To limit numerical diffusion, drop mass and drop number are accounted for in each drop bin. Droplet activation is based on the calculated super-

saturation field, and an assumed aerosol size distribution and aerosol composition (ammonium sulfate). Differences from the observed composition (typically 80 % soluble; 20 % insoluble) are not significant enough to warrant increased complexity. The initial aerosol size distribution is assumed lognormal based on fits to the observed aerosol probe data. (See section 3.) The vertical distribution is based on measurements by these same instruments at multiple aircraft levels. The aerosol number mixing ratio is a prognostic model variable but the size distribution is assumed constant. In these simulations we do not track the aerosol mass dissolved inside the droplets [*Feingold et al.*, 1996]. Thus we do not address questions associated with cloud processing of aerosol.

2.2. Surface Model

Because these clouds are driven by surface forcing, we include a surface model comprising a number of components. The Land Ecosystem-Atmosphere Feedback (LEAF) model represents the storage and exchange of energy (heat and moisture) fluxes between the surface and atmosphere. Four processes are considered when evaluating the latent heat fluxes: the transpiration through the plant stomata, evaporation from the soil, and evaporation and condensation of moisture on the vegetation. A version of the TOP-MODEL [*Band*, 1993], a land hydrology model, is coupled to the LEAF model to represent the sub-grid-scale run-off. In the LEAF model, vegetation may be multi-layered in terms of leaf area index, but is represented by a single prognostic temperature and surface moisture. There are 12 soil types and 18 vegetation types from which to select. Each individual grid column can be assigned to either a single type, or a mosaic of different types. A sandy clay loam for the soil texture, and crop/mixed farming and grass (vegetation height is about $30 \sim 60$ cm) for the vegetation are chosen for this study, and applied over the entire domain. There are 8 soil layers with a root depth of 1.0 m. The leaf area index is 6.

The initial volumetric soil moisture content used in the model is 0.189 $m^3 \, m^{-3}$ corresponding to a relative wetness of 45 % at the saturation content of 0.42 $m^3 \, m^{-3}$. The sensitivity of the simulations to the initial volumetric soil moisture is tested, and a summary of these results discussed in Section 6.2 and Table 4.

Longwave radiation is emitted, absorbed, and reflected by soil and vegetation, while downward solar (shortwave) radiation is absorbed by soil and vegetation. Changes in temperature and heat fluxes due to absorption and reflection of radiation are calculated in the LEAF model.

3. Instrumentation, Flight strategy, and Statistical sampling

The CIRPAS Twin Otter included a number of aerosol and cloud instruments, the most pertinent of which are described below. All instruments were carefully calibrated during the course of the experiment. A forward scattering spectrometer probe (FSSP; Particle Measurement Systems, Inc., CO) was used to measure cloud droplet size distributions over the diameter range 2 μm $\leq D < 45 \, \mu$m, and corrected for coincidence errors [*Lu et al.*, 2007]. A cloud imaging probe (CIP; Droplet Measurement Technologies, CO) measured the drizzle drops (25 μm $\leq D < 1550$ μm). Simultaneously, drop size distributions over the range $4 < D < 200 \, \mu$m were measured by the Phase Doppler Interferometer [PDI; *Chuang et al.*, 2007]. A PVM-100 [*Gerber et al.*, 1994] measured bulk liquid water content (LWC), surface area, and hence drop effective radius ($r_e = \langle r^3 \rangle / \langle r^2 \rangle$, where $\langle r^n \rangle$ represents the nth moment of the drop size distribution). A hot wire probe provided an additional measurement of LWC. Aerosol size distributions were measured by an optical particle counter (PCASP; Particle Measurement Systems, Inc., CO) for 0.15 μm $\leq D \leq 3 \, \mu$m. The heaters on the PCASP were turned on during flight providing some degree of drying of the aerosol particles but in the humid environment in Texas, it is uncertain how dry the particles were. Additional size distribution information was measured using differential

mobility analyzers (DMA; TSI, Inc., MN) over the diameter range 10 nm to 806 nm. One of these measured the particles in their dry state and the other at a fixed relative humidity (RH) of 65 %. Aerosol composition information was available from an Aerosol Mass Spectrometer (AMS; online at http://cires.colorado.edu/∼jjose/ams.html) and a Particle into Liquid Sampler (PILS; Brechtel Manufacturing, Inc.).

Although many field experiments tend to focus on individual clouds, much of the sampling during GoMACCS consisted of multiple level legs below, in, and above cloud fields. Subcloud legs focused on aerosol size and composition measurements; in-cloud legs (multiple) were such that after a given cloud was sampled, the aircraft would proceed at constant altitude and sample the next cloud along the sample line, with small deviations from that line to increase sampling statistics. Above-cloud legs provided measurements of free-tropospheric air and irradiance using the Solar Spectral Flux Radiometer [SSFR; *Pilewskie et al.*, 2003] The flight patterns for the five flights are shown in Figure 1 and color-coded by the total CN concentration. Extremely polluted conditions are evident, particularly near the Houston city-center and local sources. On occasion, single clouds were penetrated at multiple levels for extended periods of time but these are not the focus of this study.

4. Model Initial Conditions and Experiment Design

4.1. Initial soundings

We have chosen five days in September 2006 (6, 7, 8, 11, and 15) from a total of 22 flight days during the campaign, to run model simulations and to compare with aircraft data. The selected 5 days were deemed suitable for studying aerosol-cloud interactions based on prevalence and persistence of clouds, and instrument performance. All cases can be categorized as

non-precipitating warm, cumulus mediocris clouds. Given the domain-size discussed below,

simulating five cases is not a trivial task, requiring the use of a large multiprocessor computer.

Simulations are based on soundings at the University of Houston for the five days in question.

Simulations were usually started at 12 UTC (07:00 local time) and run for 12 h (720 min). On

Sept 15 the simulation was started at 15 UTC because the 12 UTC sounding was saturated from

the surface up to 2.5 km. Turbulence is initiated by imposing instantaneous, spatially-random

perturbations of $\pm$ 0.1 K in potential temperature at each grid point in the lowest 200 m above

the surface. The domain size is 12.8 km $\times$ 12.8 km $\times$ 5 km with $\Delta x = \Delta y = 100$ m and $\Delta z =$

50 m. The time step is 2 s. The days chosen for the three-dimensional simulations along with

other pertinent information are listed in Table 1.

Figure 2 shows a composite sounding for the 5 days, including mean and standard deviation

at 7 am local time (top panels). The potential temperature (θ) profile at 7 am shows a typical

early-morning stable boundary layer. The 1 pm soundings (bottom panels, along with the 7 am

soundings) reflect the deepening continental boundary layer as the day progresses. The largest

variability in the potential temperature profile is about 2.4 K between 2 km and 3 km, while the

moisture profile shows variability as high as 4 g kg^{-1} in the lower levels from the surface up 2

km. The wind profiles (not shown) ranged from predominantly northerly flow on September 6,

2006 to mostly southerly flow on the other 4 days.

4.2. Initial aerosol profiles and scaling

The mean and standard deviation of the initial aerosol profiles based on the PCASP mea-

surement are plotted in Figure 3 for the 5 days. (All in-cloud data are excluded in this plot.)

The heaviest aerosol loading is located in the lowest 1 km of the PBL. There is a fair degree of

variability in the maximum aerosol number mixing ratio, N_a among the days, ranging from 500

mg^{-1} to 1400 mg^{-1}. [Mixing ratio units of number per milligram or mg^{-1} are of similar magnitude to the commonly used cm^{-3}. (The difference between the two units is the air density.) They are preferred because in the absence of microphysical processes that might modify N_a, the variables are invariant with height. Mixing ratio units will become even more important when considering cloud drop number mixing ratio N_d in the context of cloud processes.]

The DMA, which samples smaller aerosol particles, has mixing ratios ranging from 2 to 10 times greater than those of the PCASP for the below cloud flight leg for the 5 days in question. We have assumed that the DMA data have similar (scaled) vertical profiles to the PCASP on any given day. The initial aerosol size distribution is assumed to be lognormal (single or multimodal) based on fits to the observed PCASP and DMA data. Note however that this paper does not attempt a careful closure on drop number mixing ratio N_d based on observed aerosol size/composition and updraft velocity w. The latter will be addressed with cloud parcel models that resolve more of the details of drop activation in a later publication. Our primary concern here is a broader comparison of the statistics of N_d.

5. Comparison between the model results and observations

The typical lifetime of cumulus clouds being studied ranges from 20 to 40 min [*Jiang et al.*, 2006] so that during the course of a couple of hours of observations or model simulations, many cumulus cells form and dissipate. It is unrealistic to expect that model simulations will produce clouds with properties and characteristics that match the observations on an individual cloud basis. Instead we have focused on statistical characteristics of cumulus cloud populations when comparing between model output and observations.

It is assumed that the atmospheric conditions in which clouds form in the vicinity of the sounding site and in the area covered by the aircraft are similar for any given day, and that the statistics of the simulated clouds represent the characteristics of the observed clouds.

5.1. Definitions of cloud, cloud updraft core, and conditional sampling methodology

5.1.1. Cloud

A grid point in the model or along the flight path is defined as cloudy when the cloud LWC is greater than 0.05 g m^{-3}. The aircraft data are sampled along the flight path at a frequency of 1 Hz, which for a ground speed of ~ 50 m s^{-1}, corresponds to ~ 50 m. To qualify as a distinct cloud in the aircraft data, we require at least 6 consecutive cloudy points along the flight-leg, i.e., a horizontal length scale of at least 300 m. A total of 92 clouds were sampled in this manner (Table 1). (When the 6 consecutive cloudy point requirement is relaxed to 4 points, the total number of clouds increases to 141.) The 0.05 g m^{-3} LWC threshold used in this analysis filters out cloud remnants. (This value corresponds to ~ 25 m of vertical displacement above cloud base for an adiabatic LWC profile.)

It should be noted that the modeled cumulus clouds are three-dimensional, as opposed to the two-dimensional slices sampled by the aircraft. The modeled cloud size (or diameter) is calculated as the square-root of the total number of horizontal grid boxes meeting the cloud criteria, at a given height, for that individual cloud. Any cloud with size smaller than 300 m is excluded in the analysis.

For the simulations, we have sampled the entire domain for 3 h (1400-1700 local time), with output at 1 min intervals, for each day. We have also tested the sensitivity of the results to randomly sampling only 5 % of the model output or reducing sample frequency to every 2 min, or 5 min to gain further understanding of the simulated cloud structure, and to test the soundness

of the analysis methodology (section 6.1). Applying the minimum cloud size of 300 m, the total number of clouds sampled from the model at a sampling frequency of 1 min is 2828 (vs. 92 sampled from the aircraft data - a factor of about 30). The simulated clouds clearly have far better statistics. In the following, all the comparison plots are between 5 day composites for both aircraft data and model output, unless otherwise mentioned.

5.1.2. Cloud Updraft Core

The concept of an updraft core is commonly used in cumulus cloud studies [e.g. *Siebesma and Cuijpers*, 1995]. It is defined using various criteria but all convey the sense that the updraft core is the region of actively growing cloud, relatively unaffected by entrainment and mixing. A cloud updraft core is defined here as follows. The cloud dimension (or flight-leg) must be at least 500 m [*Anderson et al.*, 2005]. Over this 500 m scale, the LWC must exceed 0.05 g m^{-3}, and the vertical velocity (w) must be greater than 1 m s^{-1} at all points. The 500 m requirement is applied to the updraft core region and the actual cloud size may be larger. Note that variables within the updraft core, as defined here, are not necessarily adiabatic. This definition of cloud updraft core is applied to both observed and modeled clouds. As will be shown later, this excludes a large portion of the cloud population which have sizes of 500 m or smaller (see section 5.4).

The modeled cloud updraft core size (or diameter) is calculated as the square-root of the total number of horizontal grid boxes meeting the core criteria, at a given height, for that individual cloud. As in the case of comparison between total number of clouds (modeled vs. observed) the total number of cloud updraft cores is much larger than the observed number (918 modeled vs. 10 observed, see Table 1). In section 6.1 we will address this issue by sub-sampling the model output to test the robustness of the comparisons.

5.2. Time Series

To illustrate the temporal evolution of various cloud fields derived from model output, time series of LWP (averaged only over columns that have LWP greater than 20 g m^{-2}), cloud fraction (the fraction of vertical columns that have LWC greater than 0.05 g m^{-3}), cloud-top maximum ($z_{t,max}$) and cloud-base minimum ($z_{b,min}$), and the mean cloud base and top are plotted in Figure 4 for September 6, 2006. ($z_{t,max}$ and $z_{b,min}$ are respectively the highest and lowest cloud top and base in the domain.) Figure 4 shows that clouds start to form between 1 and 2 pm (local time) and are sustained for 3 to 4 hours. Because the modeled clouds are mainly initiated by surface energy fluxes, they may form at different times than those observed, especially when mid-level cirrus clouds reduce the net surface radiation for some of the days. (The limited model-top does not allow for mid- and high-level clouds.) The lowest cloud base is steady, and located between 1300 and 1500 m over the time period shown, while the highest cloud top fluctuates a fair amount, reflecting the variability in the strength of convection over the domain. The difference between $z_{t,max}$ and $z_{b,min}$ should not be interpreted as a proxy for cloud depth; the time series of mean z_t and z_b show that cloud depth is on the order of a few 100 meters.

The other four days are not shown as the temporal evolution of the boundary layers is qualitatively similar. The model output of various fields for each of the 5 days of simulations are time-averaged over a 3 h time period and summarized in Table 2. The extent of day-to-day variability in various fields is noted. Most of the days have fairly similar general characteristics, with the exception of September 7, which has lower LWP and cloud fraction, and a higher cloud base. September 11 has the lowest cloud base, while September 8 has the strongest surface total heat (sensible plus latent) fluxes, and the highest LWP and cloud top. Cloud fractions range from about 5.3 % (Sept. 7) to 17.4 % (Sept. 11), with the mean for all 5 days at 10.3 %. For

perspective, the cloud fraction for the cumulus clouds in the ten year composite between 1971–

1981 for September for the Houston area is about 7 % [*Warren et al.*, 1986]. It should be noted

that for fair comparison the 3 h (1400-1700 h local time) is applied to all 5 days even though

the most vigorous clouds may have occurred outside this 3 h window on some of the days.

5.3. Vertical Normalization of Clouds

As the time series of $z_{t,max}$ and $z_{b,min}$ show, cloud base and top can occur over a wide range

of heights, a characteristic of this type of clouds. Therefore, two different types of sampling

method are used to present the results. One is to conditionally sample and average over the

cloud region (as defined in Section 5.1.1) as a function of altitude and the other is to average

individual clouds to a normalized grid such that cloud base is represented by 0 and cloud top by

1.

The philosophy behind vertical normalization is that clouds have certain well-defined pro-

cesses related to their vertical development; cloud base is the level at which condensation

and droplet activation occur; mid-levels are regions of condensation and collision-coalescence;

cloud top represents some distance above the limit of convection where parcels become nega-

tively buoyant and drop evaporation is active. The fact that these regions are fairly well-defined

suggests that vertical normalization is justified. Clouds that are sampled at different stages

of their evolution will exhibit profiles that represent the dominance of different sets of pro-

cesses. For example, actively growing clouds are expected to exhibit different profiles to decay-

ing clouds. These differences will in fact become evident in the ensuing discussion when we

compare cloudy versus core samples.

The normalization applied is a linear one, much the same as the normalization applied in

boundary layer studies (surface = 0, inversion = 1). Thus all clouds are linearly scaled to a

common grid. The linear scaling is primarily in the interest of simplicity, and is appropriate

for fields that tend to increase linearly with height (e.g., LWC in adiabatic cores), or N_d, which

in the cores of non-precipitating clouds, is constant. Although these are simple, and somewhat

restrictive cases, they do provide some support for linear scaling.

Results shown below will include both altitude and normalized cloud-depth averaging for the

modeled results. Ideally one would prefer to analyze and process the aircraft data in a similar

manner. Because of the large variability in observed cloud top, the difficulty of identifying

cloud top, and the limited number of vertical levels sampled, the data (LWC, N_d, r_e, and w) are

binned into 500 m altitude intervals. The best data coverage is at the lower levels, while the

contribution from the higher altitudes is significantly smaller.

5.3.1. Normalized Frequency Distributions

For the frequency distribution calculations, only the cloud region (as defined above) is sam-

pled to filter out cloud remnants. Data have been sampled at all altitudes to improve the sample

statistics. The observations include vertical bars at $\pm$ 1 standard deviation. For clarity, these

are not plotted for the model output, which tends to exhibit less variability. The frequency of

occurrence of observed LWC compared to the model output show broad agreement (Figure 5a).

The model maximum frequency occurs at a slightly lower value of LWC. There is a great deal

of day-to-day variability in the aircraft data as shown in the standard deviations. The inset of

Figure 5a shows the tail of the distribution on a logarithmic scale to focus on the differences. At

these low frequencies of occurrence, the observations suffer from poor sampling statistics and

differences are expected. The modeled LWC shows the occurrence of values as high as 3 g m^{-3}

(inset), while maximum observed values are slightly less than 2 g m^{-3}.

The frequency distributions of the modeled N_d (Figure 5b) show a clear bimodal structure with one peak at the lowest N_d bin and a second peak at $\sim 500 \, \text{mg}^{-1}$, while the aircraft data show a very flat second peak from $\sim 300 \, \text{mg}^{-1}$ to $\sim 700 \, \text{mg}^{-1}$. As in the frequency distribution of LWC, the modeled N_d shows the occurrence of values as high as $1100 \, \text{mg}^{-1}$, while maximum observed values are at about $800 \, \text{mg}^{-1}$ (inset). Differences are at least partially due to the small sampling statistics of these rare events but we cannot rule out instrumental uncertainties. Even though the FSSP has been corrected for coincidence events, we cannot be certain that the probe does not undersample N_d. Preliminary results from the PDI cloud probe suggest that this is indeed the case.

Figure 5c shows a similar comparison for the w field and indicates that the modeled w distribution is narrower and more peaked around zero than the observed distribution of w. The distributions also have different skewness. The sensitivity of the distribution of w to the surface energy fluxes will be investigated in section 6.2.

The frequency of occurrence of LWC with respect to altitude is plotted in Figure 6 for both aircraft data and model. In the case of the model, the output is presented in normalized height space. As discussed above, this is not feasible for the aircraft data so sampling is instead done in 500 m layers. N_d and w show weaker dependence on altitude (particularly N_d) and are therefore not shown. Both the aircraft data and the model results show the expected increase in the maximum LWC with increasing altitude (documented, e.g., by *Brenguier et al.,* 2003 using similar analysis of stratocumulus clouds). The observed five day average values show much more variability than the model results due to limited sampling statistics. (Close examination of the data shows that the very large standard deviations are due to a very low number of sample points.)

5.3.2. Cloud Profiles

Figure 7 shows the profiles of LWC, N_d, r_e, w, and cloud width, conditionally sampled and averaged over the cloud region. The cloud width is the cross-section of any individual cloud $\geq$ 300 m (as discussed in section 5.1.1). For the aircraft data, the error bars (standard deviations) are calculated from the means of the 5 individual days and reflect the day-to-day variability. As in Figure 6 we have vertically binned the aircraft data into 500 m intervals to improve sample statistics. The largest standard deviations reflect poor sample statistics. For the model output, both individual-day and the 5-day composites are plotted. The model results have much finer vertical resolution and therefore for clarity, every other point is plotted.

Note that the vertical range of clouds should not be interpreted as the typical vertical dimension of the clouds. Moreover, these plots do not represent mean profiles of observed and modeled clouds. At any given height the plot simply indicates the range of cloud sizes sampled by the aircraft or in the model. Because of the much larger model sample, more deep clouds are sampled, as indicated in the plot.

The number of clouds sampled at the (five) aircraft sample altitudes and (six) model output levels are indicated on the right margin of Figures 7a and 7f, respectively. (These numbers are slightly different from those listed in Table 1 due to the averaging process.) For the purpose of direct comparison, the modeled cloud number is averaged among the 5 days and binned to the 500 m altitude as in the aircraft data. The distribution shows that the majority of clouds are located below 3 km in the model output. Only 0.6 % (15) of these clouds reach the highest level. The number of clouds is similarly distributed in the aircraft data.

Several features are noteworthy:

(i) In the model results, the lowest cloud base is $\sim$1.0 km, and the highest cloud top is $\sim$ 4.0 km (similar to Figure 4c). In the aircraft data, the lowest flight leg is at a similar height to where the lowest modeled cloud base is located. This provides confidence in the model's ability to simulate cloud base since the pilots were careful to perform their lowest in-cloud leg just above cloud base. Without additional information on the cloud top height measurement, we venture that the aircraft data was unable to sample a sufficient number of the largest clouds during the course of the flights, at least to some extent because of the prescribed flight plans that called for statistical sampling at a number of levels. On other flight days, individual clouds that grew to higher altitudes (e.g. $\sim$ 3.85 km in the September 6 flight) were targeted so we know that such clouds did develop.

(ii) Over the altitude range between 1.0 km and 3.0 km where the aircraft data and model output overlap, LWC increases with height. Trends between the aircraft data and the model output are similar, but with quantitative differences. In the model output, LWC increases with height from 0.1 g m^{-3} at the lowest level up to 0.6 g m^{-3} at 3.0 km, but is consistently lower than the observed. The maximum modeled N_d of 450 mg^{-1} occurs near the cloud base, and decreases with height. This decreasing trend derives from two sources: First, N_a decreases with increasing altitude in the initial profiles so that droplet activation in clouds with higher cloud bases will result in lower N_d; Second, the clouds that have penetrated deeper into the free troposphere are more susceptible to entrainment and dilution of N_d. [The dashed line in Figure 7a indicates the adiabatic liquid water content and deviations from this line, a measure of entrainment, become increasingly larger with increases in height; see also *Lu et al.*, 2007.] The observed profile of N_d shows no distinct trend. A variety of factors may play a role. Drop number mixing ratio, which is determined largely by N_a, varies greatly from day to day, unlike

a parameter such as LWC, which is largely determined by thermodynamics. The fact that all 5 days of observed N_d data have been composited, together with the relatively small sample size is the most likely reason for the lack of a trend.

(iii) The modeled effective radii r_e are very consistent and range from 5 μm low in the clouds to 11 μm at the highest levels. They compare well with aircraft observations over the altitude range where the aircraft data and model output overlap. However, it is clear that this agreement derives from the tendency for modeled LWC and N_d to both be lower than the observed, and since r_e depends, to first order, on the ratio between LWC and N_d, this agreement is fortuitous.

(iv) The modeled w is fairly constant at ~ 1.3 m s^{-1} for the majority of the altitude range with lower values at the lowest altitudes and high values at the highest altitudes. These high values at altitudes of ~ 4 km reflect the fact that clouds that penetrated as high as 4 km tended to be more energetic. The aircraft data show a similar value of 1 m s^{-1} at the lowest level, decreasing slightly with height. The highest w occurs at the highest altitude similar to the model output. Because of poor sampling statistics (only 2 clouds at the highest level), it is hard to gauge whether this is a trend.

(v) The modeled cloud width decreases from 600 m at the cloud base to 400 m at the highest level, while the aircraft data show two distinct groups of cloud width: 600 m at the lowest level, 900 m at intermediate altitudes, decreasing back to 400m at the highest altitude.

(vi) On 4 of the days, there is remarkable consistency in the shape of the various modeled profiles. A distinct exception is September 8, most notably in the updraft profile. This is consistent with the average values listed in Table 2 for September 8 which has the strongest surface forcing, and the highest $z_{t,max}$, suggesting more vigorous convection and stronger updrafts. On

the other hand, September 8 has below average cloud fraction, which on closer examination

results from the existence of a relatively small number of deeper cells.

5.3.3. Vertically Normalized Cloud and Updraft Core Profiles

Results shown in this sub-section including Figure 8 and 9 are based on model output only.

Ideally one would prefer to analyze and process the aircraft data in a similar manner, but because

of the large variability in observed cloud-top, and the difficulty of identifying cloud-top, vertical

normalization is not performed for the aircraft data.

Figure 8 shows the profiles of LWC, N_d, r_e, w, and cloud width similar to Figure 7 but now

all variables are vertically normalized from cloud base (0) to cloud top (1). These profiles

show how each variable is distributed in a cloud-centric sense, making it easier to consider the

relationship to cloud microphysical and dynamical processes.

Normalization is performed on an individual cloud basis and subsequently all individual

clouds are averaged. The method used to vertically normalize a cloud is as follows. Cloud

base is identified first and the cloud width has to exceed 300 m in order to be considered. Sub-

sequent layers from cloud base and up are continuously counted if cloud width is greater than

300 m until the cloud definition is no longer met. At this point the cloud depth is recorded. Only

clouds with both width and depth exceeding 300 m are included. After some experimentation

with different binning procedures, the cloud depth was binned into 6 vertical bins, based on its

robust representation of results. A higher number of bins (e.g. 11) introduces some sampling

noise since a cloud depth of 300 m will include 6 vertical grid points ($\Delta z = 50$ m), with the

result that some bins are always filled (e.g. at the cloud top and base) and others are not.

The LWC increases from the lowest value at cloud base to a maximum value at normalized

height $z' = 0.8$, and then decreases thereafter towards cloud top. The increase in LWC is in

accord with theory and the decrease, indicative of substantial entrainment drying. Note that

the lowest values of LWC at the cloud base are about 0.16 g m^{-3}, i.e., higher than 0.05 g m^{-3}

because of the averaging to fairly broad vertical bins.

N_d decreases gradually from the cloud base (436 mg^{-1}) to cloud top (328 mg^{-1}), where

entrainment drying has diluted the clouds. The mixing ratio units are useful here since any

deviations from a constant value are indicative of cloud processes, rather than the reduction in

air density with increasing height.

The effective radii increase from 5 μm at the cloud base to a maximum of 8 μm at $z' = 0.8$,

and remain constant at 8 μm thereafter, reflecting the concomitant decreases in LWC and N_d.

The normalized updraft profile tends to peak somewhere between cloud base and top, al-

though with a fair amount of day-to-day variability. This is consistent with theory: latent heat

release above cloud base accelerates the updraft, but as cloud top is approached there is a reduc-

tion in w associated with a reduction in buoyancy as the clouds reach their limit of convection.

These normalized cloud profiles resemble the analysis in Figure 7 with small quantitative dif-

ferences.

Cloud width has a maximum value of 650 m at cloud base, and decreases to a minimum of

330 m at cloud top, indicative of a progressively increasing degree of entrainment and mixing

of drier air into clouds. The modeled day-to-day variability is the smallest in r_e and the largest

in w due to a more vigorous profile from September 8.

We consider the effect of aerosol on cloud microphysics by plotting the mean N_d for each of

the profiles in Figure 8b as a function of the cloud-base N_a in Figure 8f. The best fit to these

points is given by $N_d = 36.3 \, N_a^{0.35}$ On days that LWC values are similar, $r_e \propto N_a^{-0.35/3}$ (not

shown), as expected from theory. A similar analysis based on aircraft observations for 14 days

in Houston yields $N_d = 16.2\,N_a^{0.42}$ [*Lu et al.*, 2007], i.e., there was a somewhat higher sensitivity

of N_d to N_a in the data. Given the very different approaches to the modeling and observational

analyses, this difference is neither unusual, nor unexpected. What is clear is that the slopes

of these fits are much lower than typically observed in clean stratocumulus; for example, a

power-law fit to drop number concentration and PCASP aerosol concentration data from *Twohy*

et al.[2005] yields a power of 0.81.

Figure 9 is similar to Figure 8 except that the analysis examines the cloud updraft core regions.

The number of core samples is significantly lower (see Table 1). The methodology is similar to

the manner in which clouds are identified and normalized except that only cloud core regions, as

defined in section 5.1.2, with width exceeding 500 m and depth exceeding 300m, are included.

In general, the distributions are qualitatively similar to those shown in Figure 8. The differ-

ences illustrate the relative importance of microphysical and dynamical processes associated

with both growing and decaying clouds (Figure 8) and only actively growing clouds (Figure 9).

In Figure 9, the profiles are much smoother, and the maximum values are about 40 % higher

in LWC, and 50 % higher in w compared to the normalized cloud profiles (Figures 8a and 8d),

while N_d is nearly constant and decreases with height only slightly. The maximum values of N_d

are about 10 % higher at the normalized cloud base and about 26 % higher at the normalized

cloud top than those shown in Figure 8b. This relatively constant profile of N_d is indicative of

the fact that indeed only relatively unmixed cores are being sampled. The r_e profile in Figure

9c is similar to that in Figure 8c because it is dominated by the ratio of LWC and N_d, both of

which are larger in the updraft cores. (The effect of drop-size-distribution breadth is expected

to be of secondary importance.) The implication is that LWC and N_d are increasing with height

to a similar degree as in the cloudy case. Updraft-core cloud widths are larger (based on the 500

m criterion) and decrease much more slowly with height than the cloud widths, again indicating

less entrainment dilution than in the cloudy profiles.

5.3.4. Comparison of Horizontally Normalized Clouds

To further understand the morphology of cumulus clouds, we extend the concept of cloud

normalization by applying it horizontally across individual cloud cells for both aircraft data and

model output (Figure 10). For the model output, these are averages at normalized cloud height

$z' = 0.6$. Analyses near cloud base and cloud top are qualitatively similar and are therefore not

shown; they are, however, quantitatively different as reflected in the cloud profiles of Figure

8. All data that qualify as clouds ($\geq$ 300 m in width and LWC $>$ 0.05 g m^{-3}) are binned to 6

horizontal bins.

For the aircraft data, binning is performed along each in-cloud level leg. To eliminate any

potential bias due to flight directions or wind direction and shear, data is sampled twice, and

averaged along the in-cloud level leg, once forward in time, and once in reverse order. In the

case of the model analysis, clouds are identified, and binning is performed in all directions

across the center of the cloud and then averaged from east to west, west to east, south to north,

and north to south in the domain. No attempt was made to sample the model output in the same

direction as for the aircraft.

Most of the variables presented here exhibit maxima at the center of the cloud and lower

values at both ends. In the aircraft data, LWC increases from 0.31 to 0.60 g m^{-3} (48 %) from

boundary to center, N_d increases from 287 to 406 mg^{-1} (29 %), and r_e increases from 7.0 to 8.4

μm (16.6 %). For the aircraft data, the updraft does not have a clear maximum at the center, but

is highest at the left edge and lowest at the right edge. The reason for this is not clear but may

represent some remaining cross-flight directional bias that cannot be accounted for by simply averaging forward and reverse directions.

The model output exhibits similar relative changes from cloud boundary to center to the aircraft data, consistent with the normalized cloud profile at $z' = 0.6$ of the cloud depth. The agreement between the mean aircraft data and the model output is remarkable. The standard deviation reflects the differences shown in the cloud profiles (Figure 7).

It should be noted that the standard deviation is either on the same order of magnitude as or greater than the means, except for N_d. We performed a Student's T-test on the model output and aircraft data to evaluate the statistical significance of the agreement between the two populations that have different variances. The T-test examines the hypothesis that two populations have significantly different means. In this case, variables N_d, LWC, and w have probabilities of 31 %, 65 %, and 36 %, i.e., values much larger than the 5 % significance level, indicating that we can reject the hypothesis that the populations have significantly different means.

The T-test is also performed to test the robustness of the difference between the populations at the center and edge of the clouds for the 5 days (model and aircraft). For the model output, the LWC, N_d and w, differences between center and edge of cloud are again statistically significant. In the case of the observations, LWC and N_d differences between center and edge of cloud are statistically significant but the differences for observed w are not.

Figure 10 confirms what is commonly observed from satellite remote sensing [e.g., *Wielicki and Welch*, 1986], namely that when viewed from above, clouds tend to have strongly reflective centers and that reflectance gradually decrease as one moves radially outward from cloud center. The 0.05 g m^{-3} threshold on LWC used in this analysis does not address the zone around the periphery of clouds that comprises hydrated aerosol and cloud fragments addressed by *Koren*

et al. [2007]. Had we used a lower threshold, this decrease would have been much more pronounced.

5.4. Cloud Size Distributions

The cloud size distribution, or cloud size density, of a cumulus population is defined as the probability of occurrence of clouds at a range of cloud sizes. Many previous studies have tried to extract a functional relation for cloud size density from observations [e.g *Plank*, 1969; *Lopez*, 1977; *Benner and Curry*, 1998; *Cahalan and Joseph*, 1989; *Neggers et al.*, 2003a,b]. Common functional forms include exponential, lognormal, power-law, and double power-law. For simplicity, we use a power-law fit:

$$N(\ell) = a\ell^b. \tag{1}$$

The cloud size ℓ is identified and measured in the aircraft data as follows. As the aircraft penetrates a cloud, the cloud-size index i is registered if LWC > 0.05 g m^{-3} (cloudy point definition); adjacent cloudy points continue to be counted and accumulated until LWC becomes less than 0.05 g m^{-3}. Since we are using 1 Hz aircraft data, this gives a distance of ~ 50 m for every sample. We count clouds along the flight path and only count clouds that have at least 6 consecutive cloudy points, i.e., clouds with horizontal cross-sections less than 300 m are ignored.

For the modeled clouds, we adopt a similar methodology to *Neggers et al.* [2003a], but with some modifications. LWP is calculated, and each cloud projects onto a two-dimensional LWP footprint. The cloud size is calculated as the square-root of the projected area with the condition that LWP is greater than 20 g m^{-2}. As in the case of the observations, clouds that are smaller than 300 m across are excluded.

Figure 11 shows the normalized cloud size density and power-law fits (based on least absolute deviation) for both the model and aircraft data including all five days of data. After eliminating the smaller clouds (< 300 m), the aircraft data only covers the range from 300 m to 1000 m, while the modeled output includes a wider range of cloud sizes. Therefore, the fits are only applied to the subset of both the aircraft and the modeled clouds in the range 300 m to 800 m. As can be seen in Figure 11, there appears to be a scale break in the power law at ∼ 800 m in the modeled clouds. This compares well with the scale break derived by *Neggers et al.* [2003a,b] using LES of another cumulus-over-land case. For the aircraft data there are too few data points to warrant attempts to detect a scale break.

The slopes of the power law fit (*b* in Equation 1) are -2.3 and -1.9 for the aircraft data and the model results, respectively. The slope for the aircraft data is slightly higher, largely because of the limited number of clouds and the fact that smaller clouds are more prevalent and therefore more likely to be sampled. The slopes in other studies are -1.7 [*Neggers et al.*, 2003a,b] and -2.0 [*Benner and Curry*, 1989], respectively

The good agreement between the model results and aircraft data provides further support that modeled cumulus populations represent the characteristics of the observed clouds during GoMACCS reasonably well.

6. Discussion

6.1. Sensitivity To Sample Size and Frequency

As can be seen in Table 1, the number of clouds sampled from the model is far greater than that sampled by the aircraft. In fact at an interval of 1 min it is likely that the cloud fields are not decorrelated. Our decision to sample model output at high frequency was based on a desire to capture as many clouds as possible, including the more rare large clouds, and to document

these clouds over the course of their lifetimes. To assess whether there is any potential bias in the results due to the different sample sizes, we randomly sampled only 5 % of the modeled clouds over the cloud region. Five percent of the modeled clouds reduces the sub-sampled cloud numbers to ~ 141 (Table 1, column 6), i.e. much closer to the aircraft data (92, Table 1, column 4). The cloud profiles from the randomly sampled results (not shown) are almost identical to those shown in Figure 7 providing confidence that the oversampling does not induce a bias.

To further address the question of decorrelation in the cloud fields, we performed an autocorrelation on the domain-averaged liquid water fields and found that the fields were decorrelated after 4 min. We therefore reduced the sampling frequency to 2 min and 5 min, instead of every 1 min as shown in most of the figures. Again, the cloud profiles from the reduced frequency samples (not shown) are, for the most part, very close to those shown in Figure 7. Time-averaged variables similar to those listed in Table 2 are evaluated as a function of sample frequency, and summarized in Table 3. The only variable that is sensitive to sample frequency is the maximum cloud top height. The reason is that the deepest clouds have a much lower frequency of occurrence (Figure 11) and are therefore missed in the less frequent sampling.

6.2. Sensitivity To Initial Soil Moisture

Recalling the difference in the frequency distribution of w between aircraft observations and the model, we investigate one of the possible causes for this disparity, namely the surface energy fluxes, which are the primary driving force for convection. The sensitivity of the simulations and clouds to the initial soil moisture is tested by running additional 3D simulations using a simpler version of microphysics (a saturation adjustment scheme) to save simulation time. In a polluted environment such as Houston, this simplification is justified especially since there is no significant precipitation produced in any of the simulations. The onset of cumulus clouds

and the cloud base height depend on the Bowen ratio (the ratio of sensible to latent heat fluxes), which in turn can be adjusted by changing the initial soil moisture in the surface model. Sensitivity test results are summarized in Table 4. As the initial soil moisture decreases from 45 % (used in all simulations described above) to 35 % , the sensible heat flux increases, resulting in higher cloud fraction and more vigorous clouds, as expressed by an increase in turbulence, σ_w. In the absence of measurements of surface fluxes and soil moisture, and their spatial variability in the Houston region, our experience suggests that the range of initial soil moistures used is reasonable.

The frequency distributions of LWC and w (Figure 12) are somewhat different from those already presented for September 6. These differences reflect a difference due to cloud microphysics alone. However, the frequency distributions show very little dependence on the initial soil moisture. The same is true for cloud size distributions (not shown); the simulations with a saturation adjustment scheme tend to produce larger clouds than in the case of the bin microphysics, but the cloud size distributions are very similar amongst simulations initialized with different levels of soil moisture. This sensitivity to cloud microphysics, rather than surface forcing, is unexpected and suggests that the instantaneous condensation/evaporation timescale associated with the saturation adjustment scheme appears to have some unexpected feedback to the cloud dynamics. This will be explored in future studies.

7. Summary

We have presented a comparison of the microphysical and dynamical properties of 5 different large-eddy simulations (LES) of cumulus clouds and compared them with aircraft data observed in the vicinity of Houston during the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS). The comparison is facilitated by the fact that model spatial and temporal

resolution are comparable to those of the aircraft. The five simulations have generated an en-semble of over two thousand individual clouds (Table 1), at least 300 m in diameter and depth, exhibiting a wide range of sizes and at different stages of their lifecycles. Aircraft observations yield poorer statistics and have therefore been composited for the days in question. The goal has been to compare statistically the properties of the aircraft data and model output for warm continental clouds, as has been done previously for observed stratocumulus [e.g. *Brenguier et al.*, 2003), shallow cumulus [*Neggers et al.*, 2003b], and for trade cumulus [*Abel and Shipway*, 2007]. This study offers a number of additional tools that can be used in such analyses. The major results of this study may be summarized as follows:

(i) We have demonstrated that when viewed in a statistical sense, the large-eddy simulations reproduce the observed populations of non-precipitating warm boundary layer cumulus clouds over land reasonably well. The focus of the comparisons has been on frequency distributions, and profiles averaged over the cloud region and cloud updraft core region (specified *ad hoc* in section 5.1).

(ii) Only a limited number of microphysical and dynamical variables have been chosen for comparison in order to maintain a focused study. The variables examined are LWC, number mixing ratio of cloud droplets (N_d), effective radius (r_e), updraft velocity (w), and cloud width, available in both the model output and the aircraft data. Good agreement is achieved in the normalized frequency distributions (Figures 5 and 6), the profiles averaged over cloudy regions (Figure 7), the cross-cloud averages (Figure 10), and the cloud size distributions (Figure 11). In addition, the observed cloud base is well-reproduced by the model (Figure 7). Cloud top height has not been compared because this is a highly variable parameter, and because the aircraft does not provide a measure of cloud-top height for a field of clouds. The big difference in the number

of clouds sampled between the model output and the aircraft data does not induce significant bias. Infrequent sampling of the cloud fields (every 5 min) or a random sub-sampling of 5 % of the model output showed that these results are robust.

(iii) Model output generally shows remarkable qualitative consistency in profiles of the aforementioned dynamical and microphysical properties from one day to the next (Figures 7, 8, 9). Normalized profiles show the expected increase in LWC with height, decrease in N_d in height associated with entrainment and the initial aerosol profile, and the commensurate increase in r_e with height. Updraft profiles exhibit a mid-cloud maximum expected from theoretical considerations. Comparison of normalized profiles between cloudy regions and updraft core regions verify that the modeled core regions do indeed represent less dilute regions of the cloud. These profiles, together with statistics on the variability in cloud base and cloud top, may provide a useful way to parameterize the structure of these clouds in models that do not explicitly resolve small-scale convection.

(iv) The initial aerosol profiles from the five selected days cover a wide range of values from 400 mg^{-1} to 1500 mg^{-1} (Figure 3), representing less polluted to more polluted environments. The day-to-day differences in N_d are largely in response to the differences in the initial aerosol concentrations. Comparison of modeled (normalized) profiles of N_d for the 5 days show that $N_d \propto N_a^{0.35}$ [Figure 8f, see also *Lu et al.,* 2007 who derived $N_d \propto N_a^{0.42}$ from aircraft observations on 14 flights during GoMACCS]. These slopes are significantly lower than those derived in relatively clean stratocumulus.

(v) Sensitivity tests using different surface forcing, with a simple (saturation adjustment) microphysical model showed that the frequency distributions were quite sensitive to the level of microphysical complexity, but not to the surface forcing. Reasons for this are under investi-

gation. In the case of warm, precipitating clouds, the sensitivity to microphysics is likely to be even greater than shown here.

In spite of the good agreement between many of the variables, differences do exist, and some remain unresolved; an example is the difference in the normalized frequency distribution of w. In addition, modeled cloud fraction is sometimes lower than other estimates. This may be due to observed increases in the cloud fraction, associated with moisture advection, as the day progressed. However, cloud fractions are notoriously sensitive to cloudy vs. cloud-free thresholds, and to instrument resolution. Instruments with coarser resolution yield significantly higher cloud fraction than those with finer resolution [*Zhao and Di Girolamo,* 2006].

Both LWC and N_d are lower in the model output compared to the aircraft data shown in Figure 7. There are several possible reasons; aircraft sampling biases may exist because pilots tend to sample larger clouds, while model output sampling is unbiased. Differences in cloud sizes (Figure 11) may also be due to asymmetry in cloud shape and the limited information available to a pilot when choosing a path through a cloud based on a two-dimensional view prior to penetration. Owing to the human element, aircraft sampling biases are inherently difficult to evaluate. It is also likely that the model cannot capture all the details of real clouds. This is to some degree, but not only, a function of the model resolution. The question is what constitutes a "good comparison"? Clearly the answer depends on the application and the current comparisons should be viewed in this light. To explore this, a separate paper (in preparation) will examine comparisons of observed radiative fluxes and fluxes simulated based on the model output.

The challenges posed by comparison studies of this kind are many-fold. First, it is important that flight planning consider the manner in which the data will be compared to models. In this case, the concept of statistical comparisons was a goal of the study and flight plans followed

accordingly. Second, it is crucial to process and present both the model output and the aircraft data in as close as possible a manner to make the comparison meaningful. This is not a trivial task because aircraft data lacks the defined spatial grid and three-dimensional context provided by the model. Third, given the large differences in the sample points between the model output and the aircraft data, the decision of whether to composite observations must be considered. To the extent that the individual cases are similar, compositing data improves the statistical sampling and the robustness of the results. However, when day-to-day variability is significant, compositing should be avoided.

This study has purposefully compared a small number of microphysical and dynamical variables. As more studies of this kind are undertaken, the experience gained will allow flight strategies and methodologies for comparison to be refined, and further benefits to be derived.

References

Abel, S. J. and B. J. Shipway, A comparison of cloud resolving model simulations of trade wind cumulus with aircraft observations taken during RICO, *Q. J. Royal Meteorol. Soc., 133,*781-794, 2007.

Anderson, N. F., C. A. Grainger, and J. L. Smith, Characteristic of strong updrafts in precipitation systems over the central Tropical Pacific Ocean and in the Amazon, *J. Atmos. Sci., 44,* 731-738, 2005.

Band, L. E., Effect of land surface representation on forest water and carbon budgets. *J. Hydrol., 150,* 749–772, 1993.

Benner, C. T, and J. A. Curry, Characteristics of small tropical cumulus clouds and their impact on the environment, *J. Geophys. Res., 103,* 28753-28767, 1998.

Brenguier J.-L., H. Pawlowska, and L. Schuller, Cloud microphysical and radiative properties for parameterization and satellite monitoring of the indirect effect of aerosol on climate, *J. Geophys. Res., 108,* NO. D15, 8632, doi:10.1029/2002JD002682, 2003.

Brown, A. R., and co-authors, Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quart. J. Roy. Meteor. Soc., 128,* 1075–1094, 2002.

Cahalan, R. F. and J. H. Joseph, Fractal statistics of cloud fields. *Mon. Wea. Rev., 117,* 261-272, 1989.

Clark, T. L, Numerical modeling of the dynamics and microphysics of warm cumulus convection. *J. Atmos. Sci., 30,* 857-878, 1973.

Chuang, P. Y., J. D. Small, E. W. Saw, R. A. Shaw, C.M. Sipperley, G. A. Payne, W. D. Bachalo, H. Siebert, F. Burnet, and J.L. Brenguier, Airborne Phase Doppler Interferometry for Cloud Microphysical Measurements. *Aerosol Sci. Tech*, submitted, 2007.

Cotton, W. R., R. A. Pielke Sr., R. L. Walko, G.E. Liston, C.J. Tremback, H. Jiang, R.L. McAnelly, J.Y. Harrington, M.E. Nicholls, G.G. Carrio, and J. P. McFadden, RAMS 2001: Current status and future directions. *Meteorol. Atmos. Phys.*, doi: 10.1007/s00703-001-0584-9, 2003.

Feingold, G., S. M. Kreidenweis, B. Stevens, and W. R. Cotton, Numerical simulation of stratocumulus processing of cloud condensation nuclei through collision-coalescence. *J. Geophys. Res., 101,* 21,391–21,402, 1996.

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

Jiang, H., H. Xue, A. Teller, G. Feingold, and Z. Levin, Aerosol Effects on the Lifetime of Shallow Cumulus. *Geophys. Res. Lett.*, 33. L14806, doi:10.1029/2006GL026024, 2006

Joseph, J. H. and R. F. Cahalan, Nearest neighbor spacing of fair weather cumulus clouds. *J. Appl. Meteoro., 29*, 793-805, 1990.

Koren, I, L. A. Remer, Y. J. Kaufman, Y. Rudich, 2007: On the twilight zone between clouds and aerosols. *Geophys. Res. Lett.*, 34. L08805, doi:10.1029/2007GL029253.

Lopez, R. E., The lognormal distribution and cumulus cloud populations. *Mon. Wea. Rev, 105,* 865–872, 1977.

Lu, M-L, H. H. Jonsson, P. Y. Chuang, G. Feingold, R. C. Flagan, J. H. Seinfeld, Aerosol-cloud relationships in continental shallow cumulus, *J. Geophys. Res., submitted*, 2007.

Neggers, R., A., J., H. J. J. Jonker, and A. P. Siebesma, Size statistics of cumulus cloud populations in Large-Eddy simulations, *J. Atmos. Sci., 60*, 1060–1074, 2003a.

Neggers, R., A., P. G. Duynkerke, and S. M. A. Rodts, Shallow cumulus convection: A validation of Large-Eddy simulations against aricraft and Landsat observations, *Q. J. R. Meteorol.*

Soc., 129, 2671–2696, 2003b.

Pilewskie, P., J. Pommier, R. Bergstrom, W. Gore, S. Howard, M. Rabbette, B. Schmid, P.V. Hobbs, and S.C. Tsay, Solar spectral radiative forcing during the Southern African Regional Science Initiative. *J. Geophys. Res., 108,* 2003.

Plank, V. G., The size distribution of cumulus cloud in representative Florida populations, *J. Appl. Meteor., 8*, 46–67, 1969.

Siebesma, A. P, and J. W. M. Cuijpers, Evaluation of parametric assumptions for shallow cumulus convection. *J. Atmos. Sci., 52*, 650–666, 1995.

Siebesma, A. P, and co-authors, A large eddy simulation intercomparison study of shallow cumulus convection. *J. Atmos. Sci., 60*, 1201-1219, 2003.

Twomey, S., Pollution and the planetary albedo, *Atmos. Environ., 8*, 1251–1256, 1974.

Twohy, C. H, and co-authors, Evaluation of aerosol indirect effect in marine stracocumulus clouds: Droplet number, size, liquid water path, and radiative impace. *J. Geophys. Res., 110,*D08293, doi:10.1029/2004JD005116, 2005.

Tzivion, S., G. Feingold, and Z. Levin, An efficient numerical solution to the stochastic collection equation. *J. Atmos. Sci., 44*, 3139–3149, 1987.

Walko, R. L., L. E. Band, J. Baron, T. G. F. Kittel, R. Lammers, T. J. Lee, D. Ojima, R. A. Pielke Sr., C. Taylor, C. Tague, C. J. Tremback, and P. L. Vidale, Coupled Atmosphere-biophysics-hydrology models for environmental modeling, *J. appl. meteorol., 39,* 931–944, 2000.

Warner, J., A reduction in rainfall associated with smoke from sugar-cane fires – an inadvertent weather modification? *J. Appl. Meteorol., 7,* 247–251, 1968.

Warren, S. G., C.J.Hahn, J. London, R.M. Chervin,and R. L. Jenne, Global distribution of total cloud cover and cloud type amounts over land, *NCAR/TN-273+STR*, 29 pp. + 200 maps,

National Center for Atmos. Res., Boulder, Colo, 1986.

Warren, S. G., C.J.Hahn, J. London, R.M. Chervin,and R. L. Jenne, Global distribution of total

cloud cover and cloud type amounts over the ocean, *NCAR/TN-317+STR*, 42 pp. + 170 maps,

National Center for Atmos. Res., Boulder, Colo, 1988.

Wielicki, B. A. and R. M. Welch, Cumulus cloud properties derived using Landsat satellite data.

J. climate. appl. meteorol., 25, 261–276, 1986.

Xue, H, and G. Feingold, Large eddy simulations of trade wind cumuli: investigation of aerosol

indirect effects. *J. Atmos. Sci., 63,* 1605–1622, 2006.

Zhao, G. and L. Di Girolamo, Cloud fraction errors for trade wind cumuli from EOS-Terra

instruments. *Geophys. Res. lett. 33, L20802,* doi:10.1029/2006GL027088, 2006.

Zhao, G. and L. Di Girolamo, Statistics on the macrophysical properties of trade

wind cumuli over the tropical western Atlantic. *J. Geophys. Res. 112, D10204,*

doi:10.1029/2006JD007371, 2007.

Table 1. Flight number, Day in September 2006, Time (UTC), sample points, and number of clouds. Number of clouds denotes the total number of clouds that meets the definition of "cloud" (LWC > 0.05 g m^{-3}; $L \geq 300$ m). Number of cloud updraft cores represents the total number of clouds meeting the more strict "core" criteria (LWC > 0.05 g m^{-3}, $w > 1$ m s^{-1}, and $L \geq 500$ m.

			OBS		MODEL	
Flight number	Day in Sept	Time* (UTC)	No. of clouds	No. of cloud updraft cores	No. of clouds	No. of updraft cores
Flt 15	6	1900-2030	16	3	446	151
Flt 16	7	1900-2030	23	3	326	22
Flt 17	8	1900-2030	16	2	370	279
Flt 19	11	1530-1720	18	3	862	289
Flt 22	15	1630-1800	19	0	824	177
total			92	10	2828	918

* For observational data, time includes the level flight legs only.

Table 2. Three hour (1400-1700 local time) time-averaged model output fields from September 6,7,8,11 and 15, 2006. Numbers in parentheses are the standard deviations.

Date	LWP	CF	σ_w	zb_{min}	zt_{max}	SHF	LHF
	g m^{-2}	%	m s^{-1}	m	m	W m^{-2}	W m^{-2}
060906	141.7 (42.0)	10.2 (3.2)	1.175 (0.166)	1388 (93)	3370 (479)	116.5 (29)	336.0 (58)
060907	79.8 (21.9)	5.3 (2.4)	1.198 (0.095)	1750 (94)	2915 (327)	95.4 (24)	328.4 (41)
060908	175.8 (54.7)	8.5 (1.4)	1.129 (0.119)	1314 (109)	3645 (249)	109.1 (26)	352.6 (51)
060911	166.2 (57.4)	17.4 (2.9)	0.923 (0.119)	1076 (108)	3479 (498)	107.5 (26)	308.6 (47)
060915	164.0 (29.6)	10.2 (1.5)	0.961 (0.072)	1153 (134)	3229 (264)	100.7 (22)	336.5 (60)
mean	145.5 (41.1)	10.3 (2.3)	1.077 (0.114)	1336 (107)	3327 (363)	105.8 (26)	332.4 (51)

Table 3. Sensitivity of model output to sample frequency. Data shown are averaged over 3 h (1400-1700 local time) on September 6 2006

Sample frequency	LWP	CF	σ_w	zb_{min}	zt_{max}	SHF	LHF	number of clouds
min	g m^{-2}	%	m s^{-1}	m	m	W m^{-2}	W m^{-2}	
1	141.7	10.2	1.175	1388	3370	116.5	336.0	446
2	141.4	10.2	1.174	1387	3364	116.5	336.0	220
5	140.6	10.1	1.172	1385	3352	116.4	336.1	60

Table 4. Sensitivity of model output to initial volumetric soil moisture (VSM). Data shown are averaged over 3 h (1400-1700 local time) on September 6 2006

Initial VSM	LWP	CF	σ_w	zb_{min}	zt_{max}	SHF	LHF	Bowen ratio
%	g m^{-2}	%	m s^{-1}	m	m	W m^{-2}	W m^{-2}	
45	179.2	12.1	1.193	1397	3457	117.9	340.8	0.341
40	207.9	16.2	2.266	1859	3821	266.9	99.7	2.835
35	202.2	16.6	2.411	1914	3780	287.6	66.2	4.676
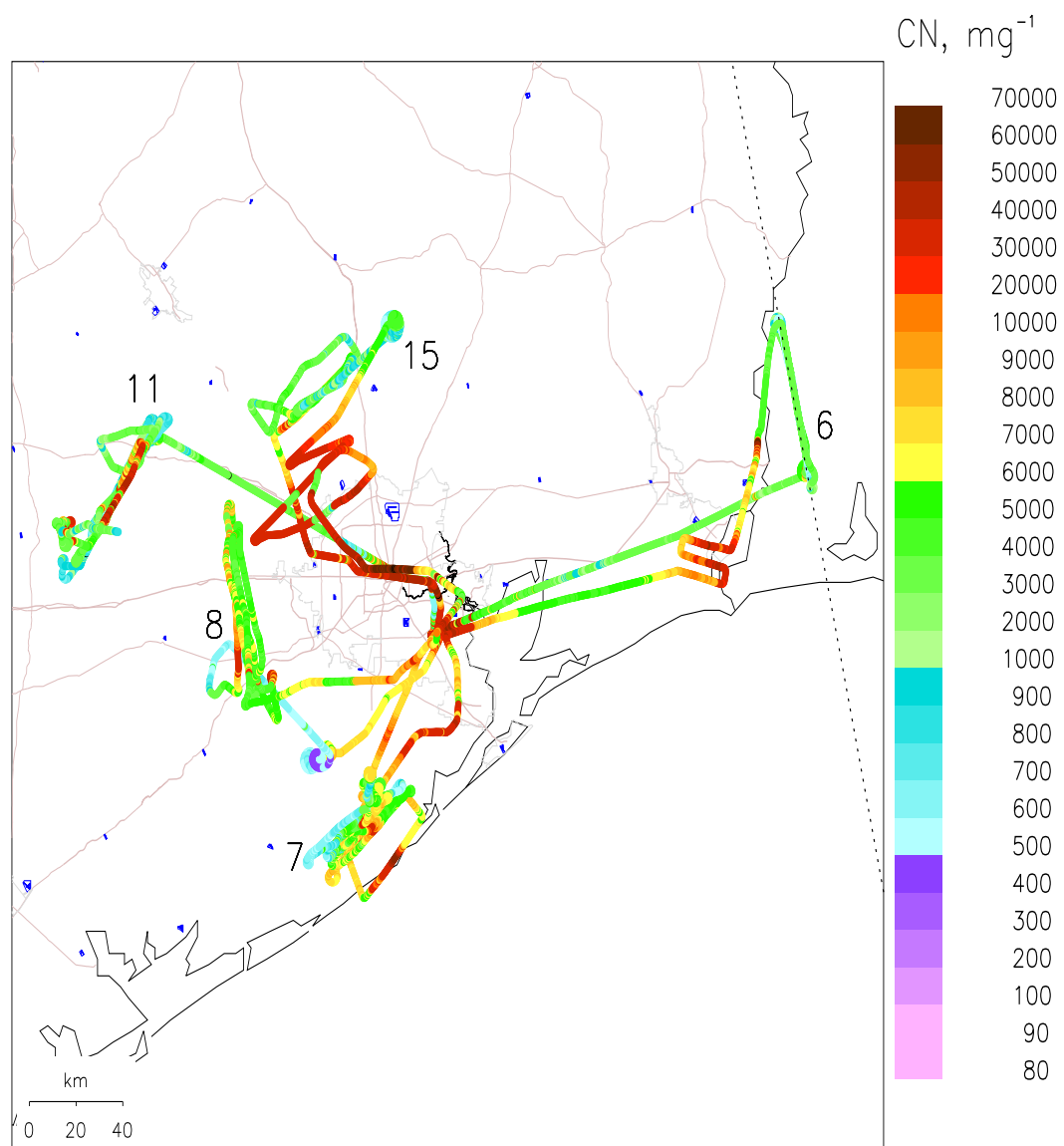
Figure 1. Flight tracks for the 5 selected days (September 6, 7, 8, 11 and 15, 2006) color-coded

by CN number mixing ratio. Tracks are labeled by day in September. The dashed line at right

is the A-Train satellite track on September 6.
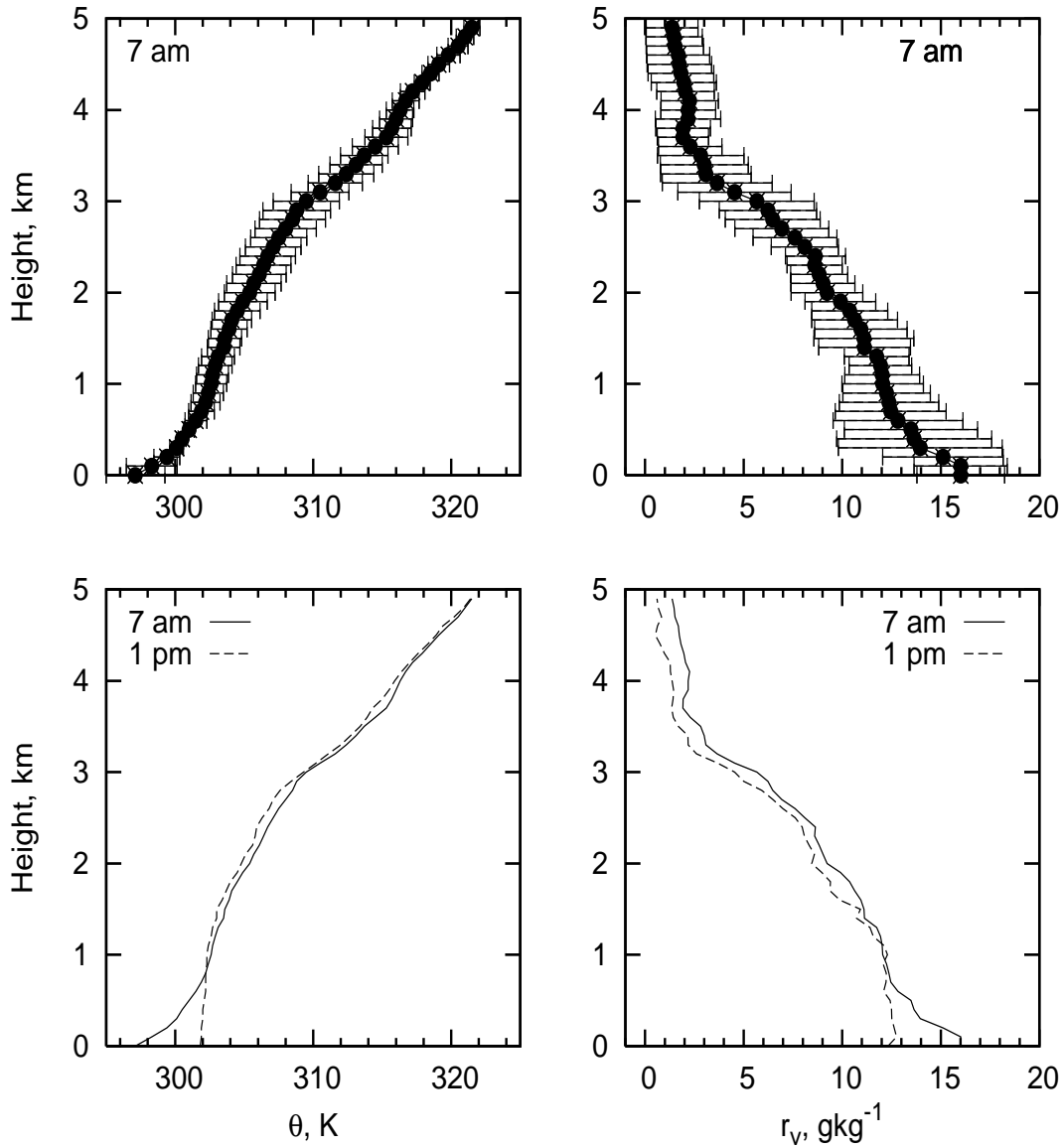
Figure 2. Initial sounding profiles averaged over the 5 days simulated by the LES. The horizontal bars denote standard deviation from day-to-day. Top panels show the initial soundings at 7 am local time, and the bottom panels show both 7 am and 1 pm local time soundings without the standard deviation.

September 6, 7, 8, 11, and 15, 2006
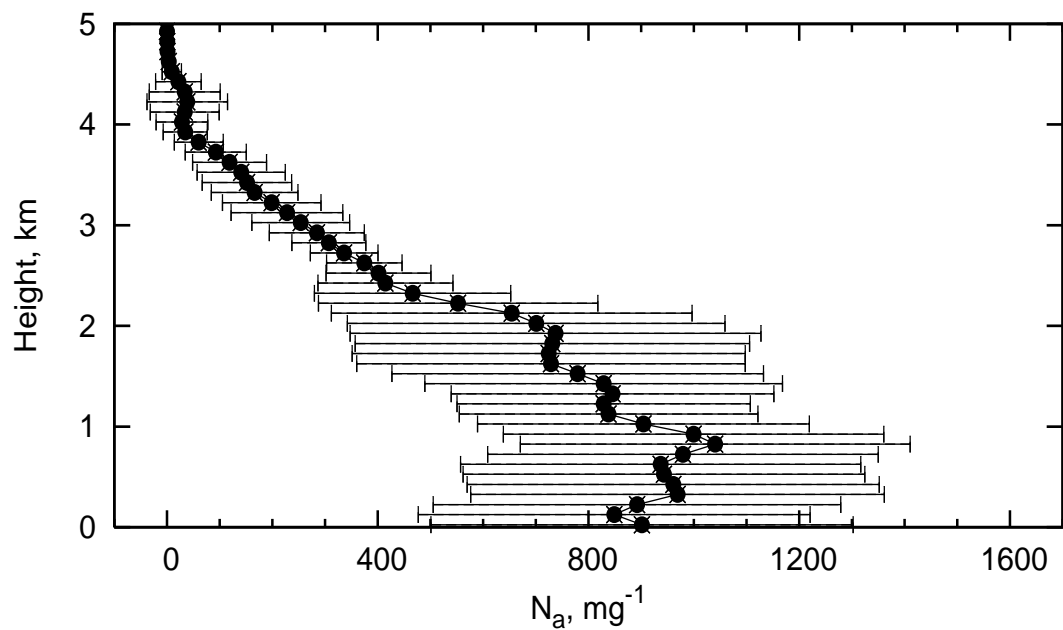
Figure 3. Initial aerosol profiles based on PCASP measurements averaged over the 5 days

selected. See text for details. The horizontal bars denote standard deviation.

06 September 2006
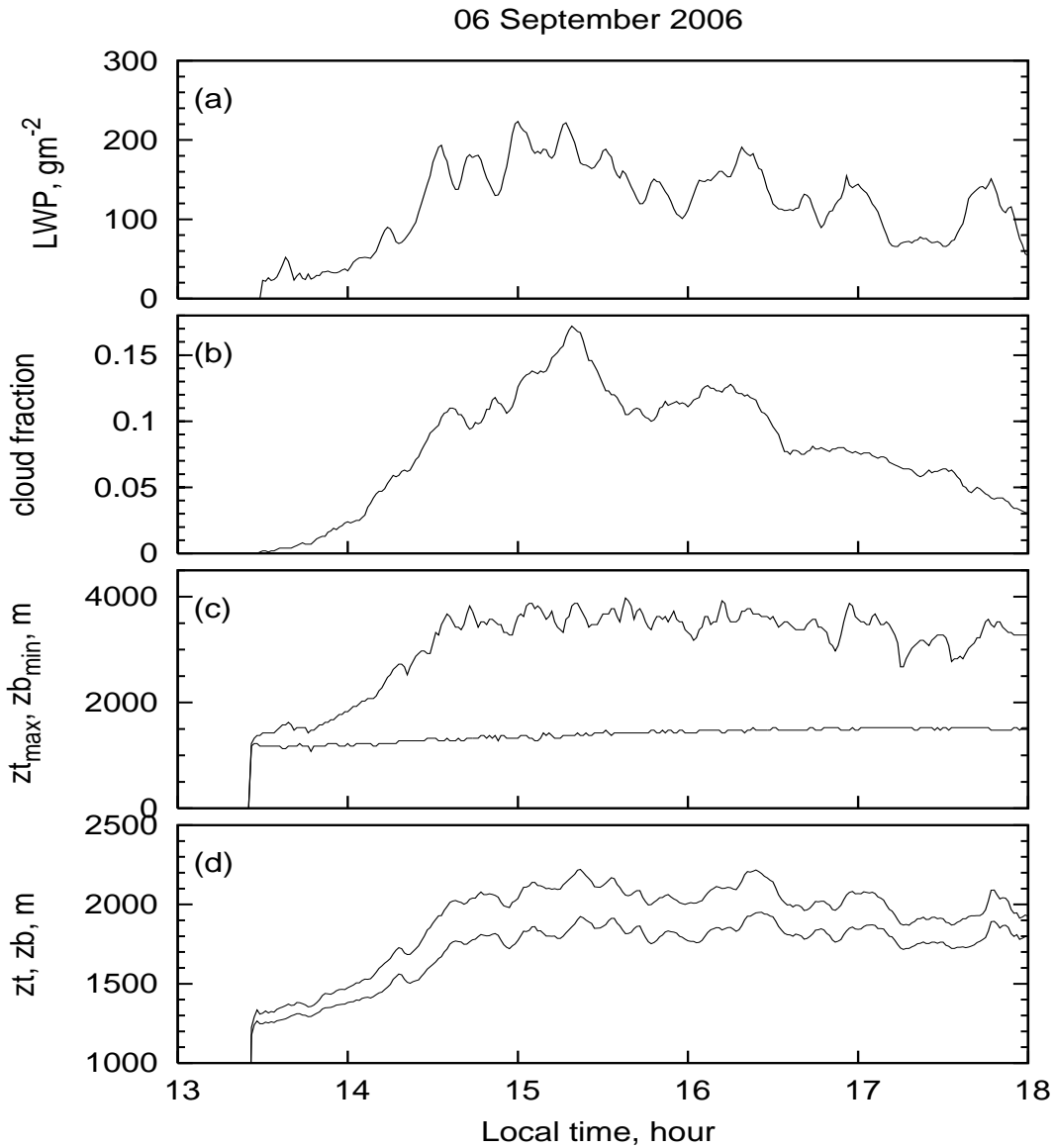
Figure 4. Time series of (a) LWP, (b) cloud fraction, (c) minimum cloud base, and maximum cloud top, (see text for definitions), and (d) average cloud base and top derived from model output for September 6 2006.
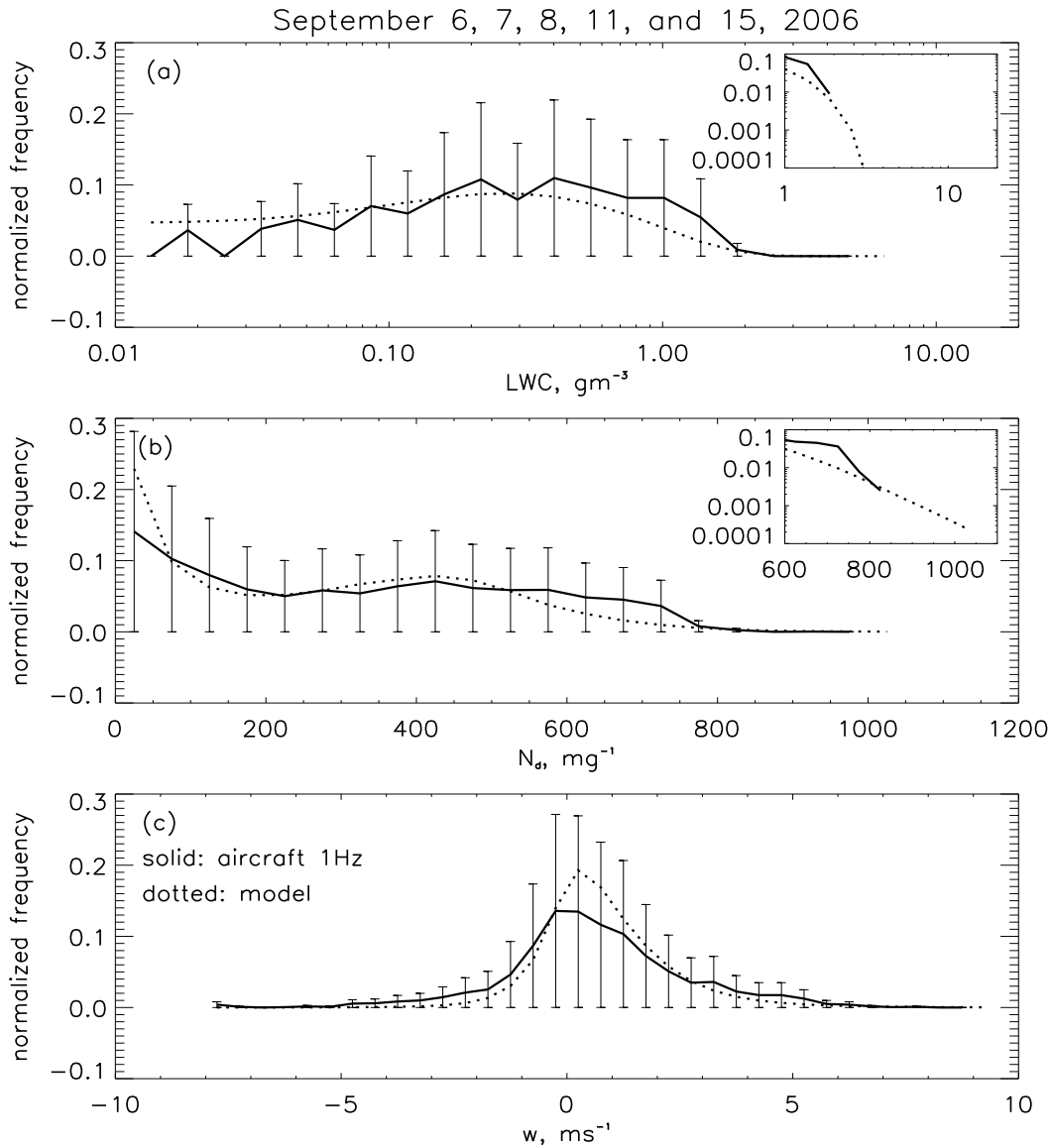
Figure 5. Normalized frequency of occurrence of (a) LWC, (b) N_d, and (c) w for a five-day composite for both modeled results and aircraft data. Solid line denotes the aircraft data, and the dashed line denotes the model output. The vertical bars denote the standard deviation from the aircraft data. The tails of the LWC and N_d distributions are plotted on a logarithmic scale as insets in (a) and (b), respectively.
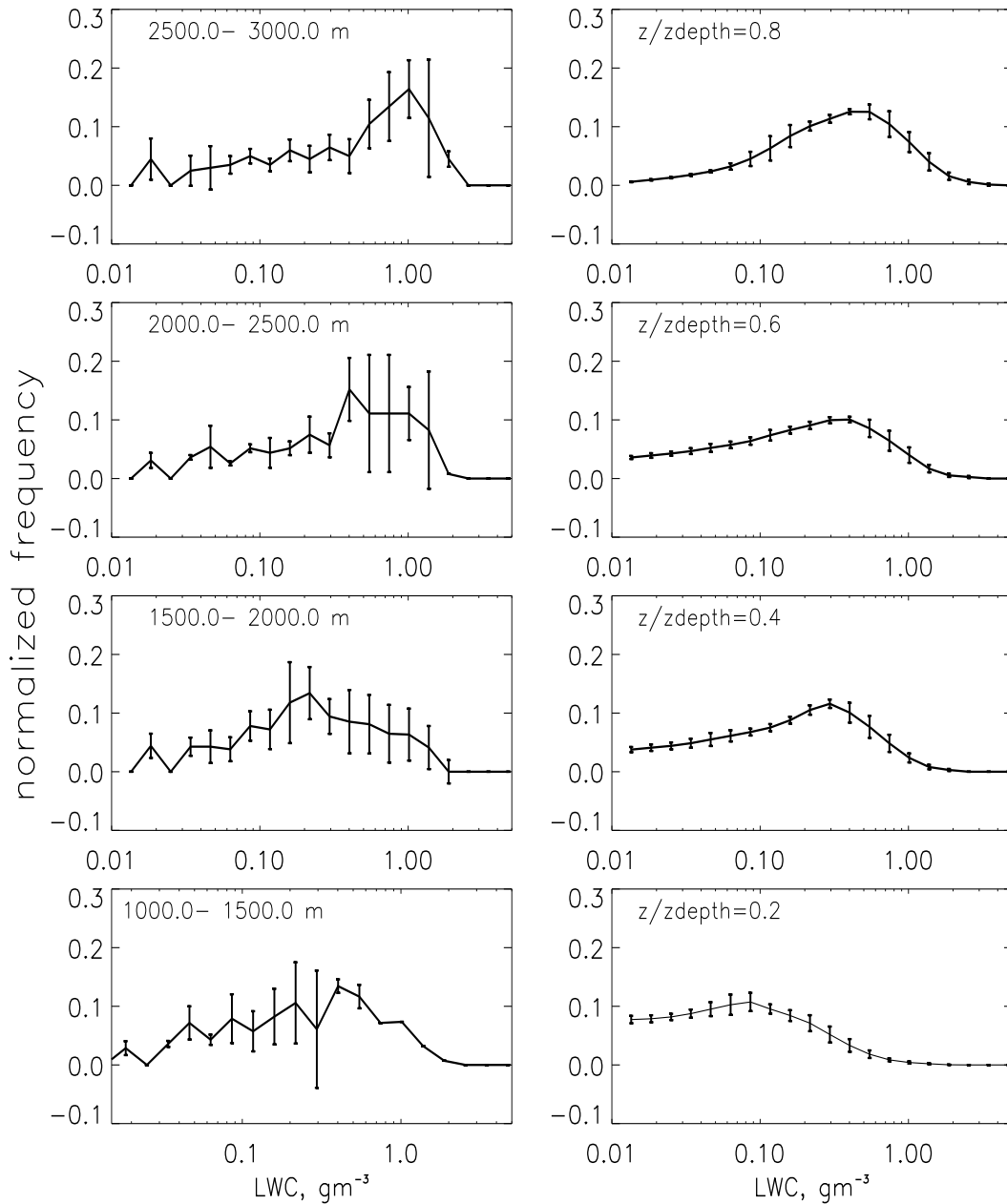
Figure 6. Normalized frequency of occurrence of LWC as a function of altitude. Left panels are the aircraft data and right panels denote model output. These are five-day composites. The vertical bars denote the standard deviation. The model output is plotted on a normalized height scale, 0 representing cloud base and 1 cloud top.
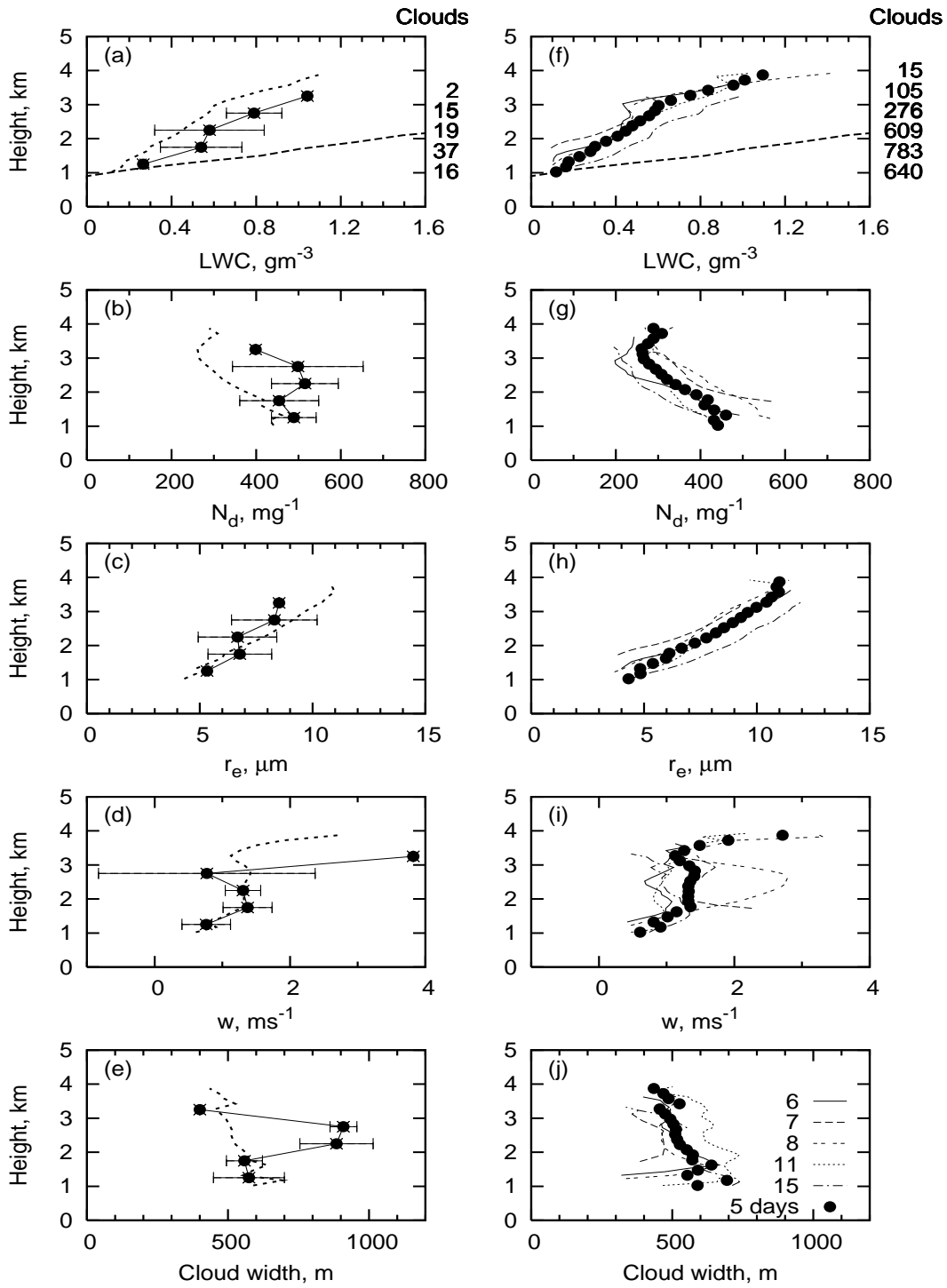
Figure 7. See next page for Figure 7 caption

Figure 7.   Profiles of LWC, droplet number mixing ratio (N_d), effective radius (r_e), updraft velocity (w), and cloud width averaged over the cloud region. The left panels denote aircraft data, and the right panels are the model output. The model output is plotted for all 5 days individually. The thick filled circle is the five-day mean. The horizontal bars on the left panels are standard deviations. The thicker dashed lines superimposed on (a) and (f) are adiabatic LWC profiles. The number of clouds sampled are listed on the right margin of (a) and (f) (see text for details). For ease of comparison, the five-day mean of the model output from (f) to (j) is superimposed on the aircraft data as a dashed line. Other line types are as labeled.
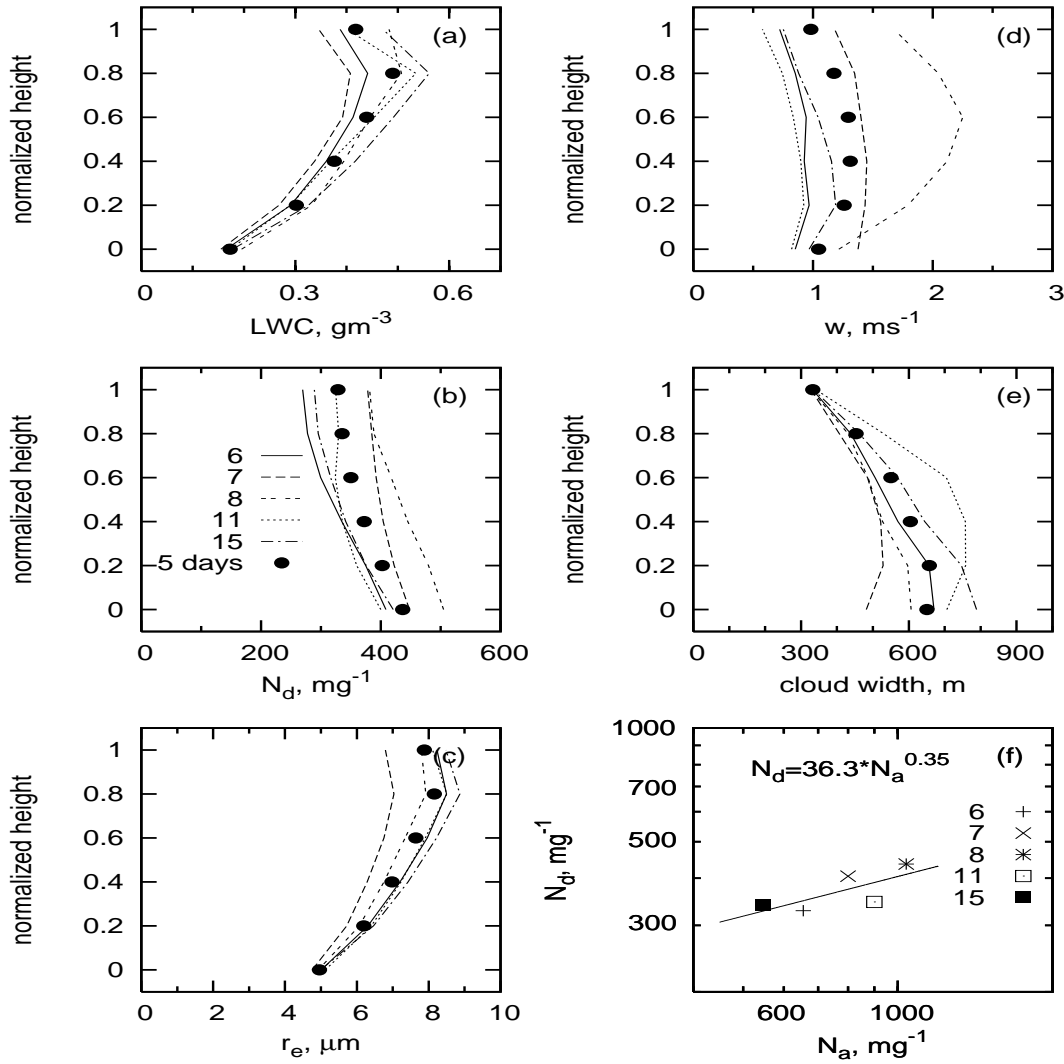
Figure 8. Vertically-normalized profiles of (a) LWC, (b) N_d, (c) r_e, (d) w, and (e) cloud width averaged over the cloudy region derived from model output. Line types are as labeled. (f) Number mixing ratio of cloud droplets (N_d; the average of the individual profiles in b) versus number mixing ratio of subcloud aerosol (N_a) for the 5 days. Each day is represented by one point, with different symbols. The power-law fit is included.
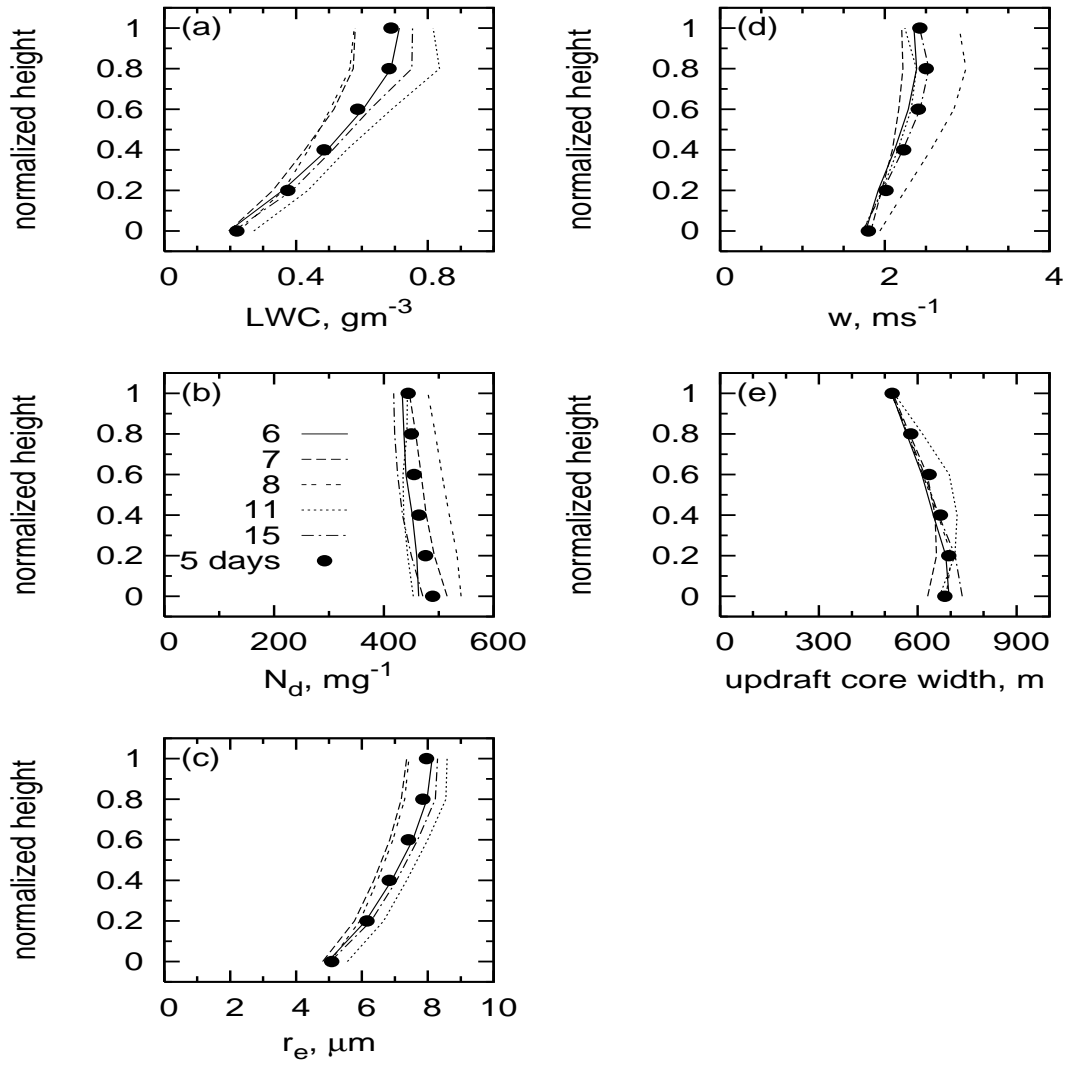
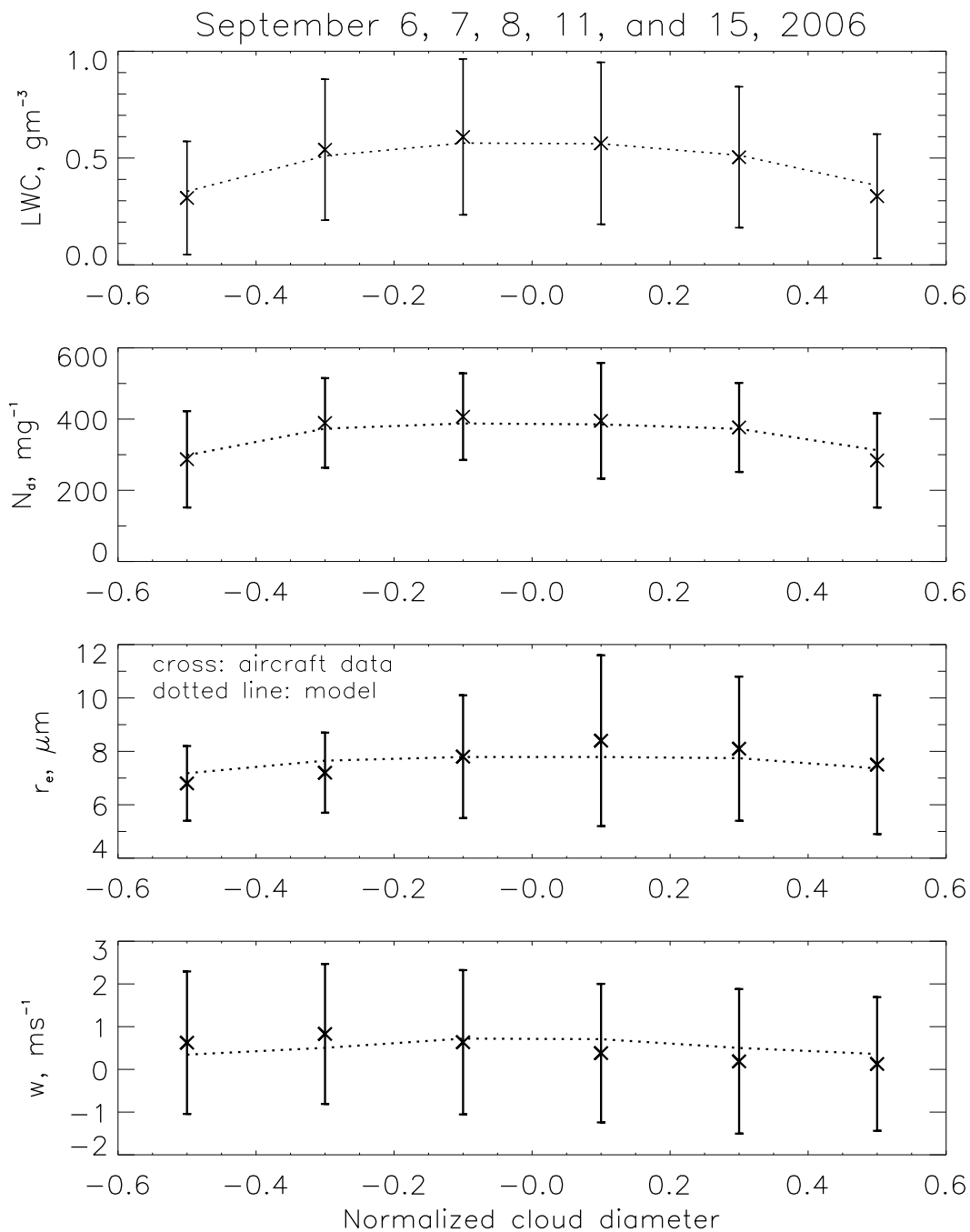**Figure 9.** As in Figure 8, but for cloud updraft core regions.

Figure 10. Horizontally-normalized cloud cross-sections of LWC, N_d, r_e, and updraft velocity w including all data for the aircraft data, and at normalized cloud height $z' = 0.6$ for the model output (see text for details). The zero on the abscissa denotes the center of the normalized cloud width. The cross symbol denotes aircraft data with ± 1 standard deviation, and the dotted lines are the model output. These are five-day composites.
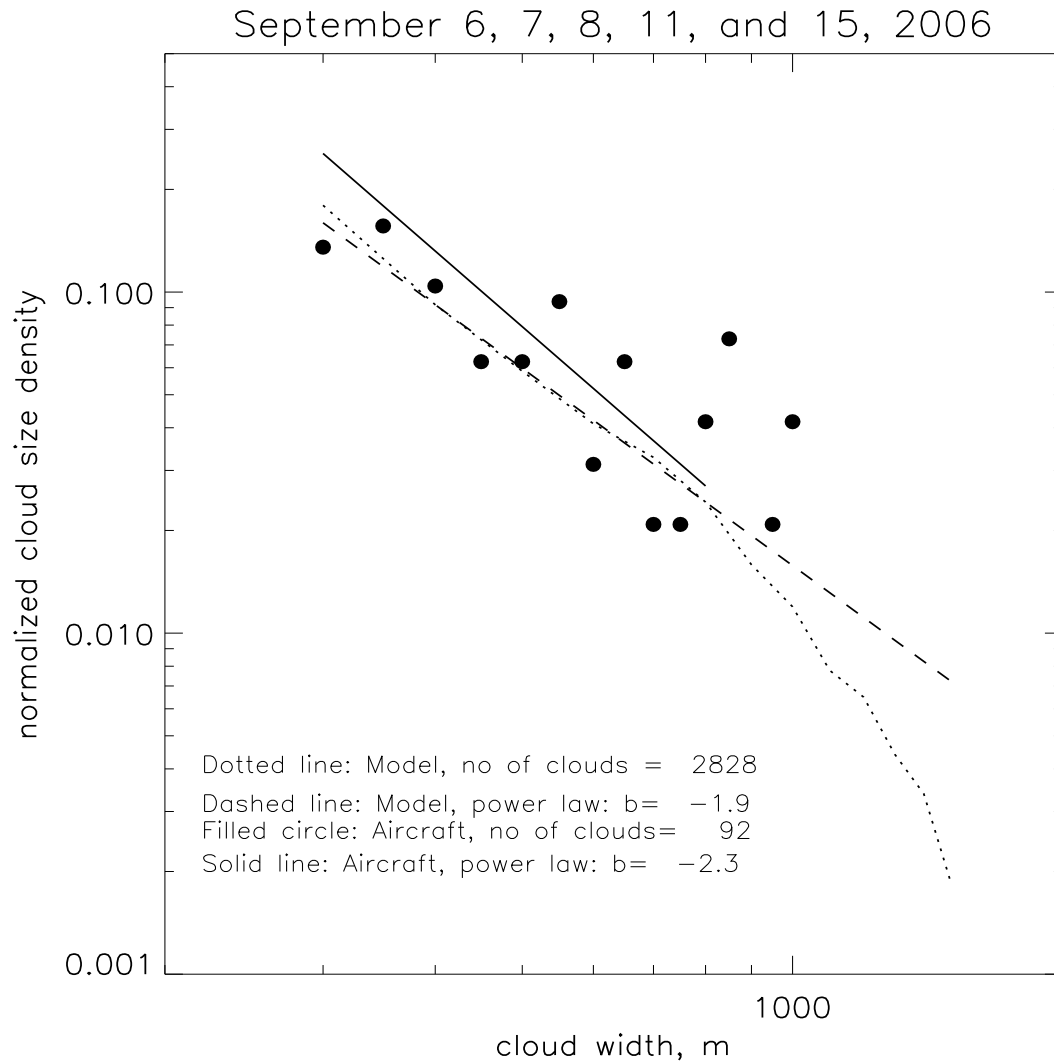
Figure 11. Normalized cloud size density. Filled circles denote aircraft data, and the dotted line denote the model output. The solid and dashed line are power-law fits to the aircraft data and model output, as indicated.
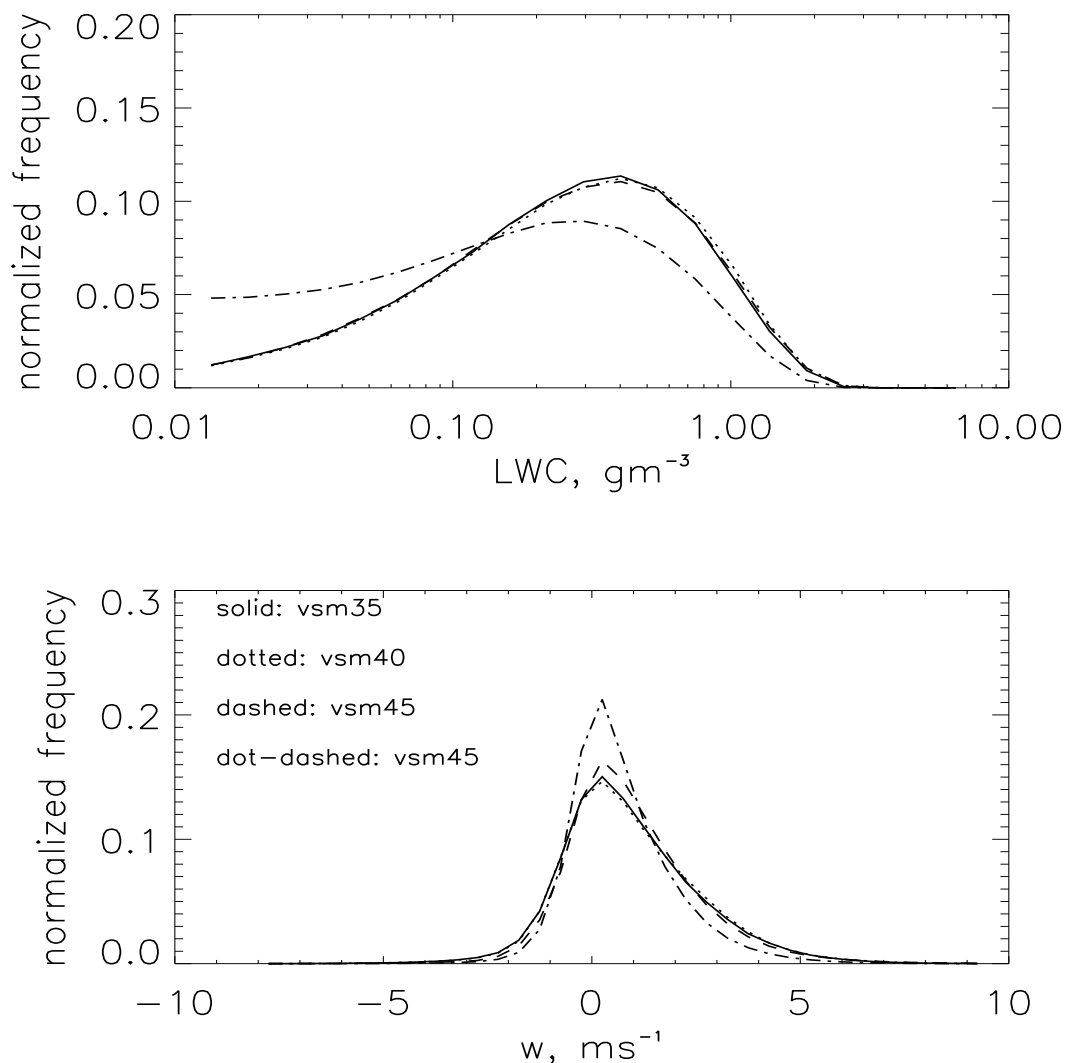
Figure 12. Normalized frequency distributions of (a) LWC, and (b) *w* for September 6 to test

the sensitivity of simulations to initial volumetric soil moisture (VSM). Solid line denotes an

initial wetness of 35 %, dotted line denotes 40 %, and dashed line denotes 45 %. These three

simulations used a simple saturation adjustment microphysical scheme (see text for details),

while the dot-dashed line denotes 45 % but using bin microphysics as in the rest of the paper.