DISSERTATION

THE SIMULTANEOUS INFLUENCE OF THERMODYNAMICS AND AEROSOLS ON

DEEP CONVECTION AND LIGHTNING

Submitted by

Douglas C. Stolz

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2016

Doctoral Committee:

Advisor: Steven A. Rutledge

Jeffrey R. Pierce
Susan C. van den Heever
Steven C. Reising

ABSTRACT

THE SIMULTANEOUS INFLUENCE OF THERMODYNAMICS AND AEROSOLS ON

DEEP CONVECTION AND LIGHTNING

The dissertation consists of a multi-scale investigation of the relative contributions of

thermodynamics and aerosols to the observed variability of deep convective clouds in the

Tropics. First, estimates of thermodynamic quantities and cloud-condensation nuclei (CCN) in

the environment are attributed to convective features (CFs) observed by the Tropical Rainfall

Measuring Mission (TRMM) satellite for eight years (2004-2011) between 36°S-36°N across all

longitudes. The collection of simultaneous observations was analyzed in order to assess the

relevance of thermodynamic and aerosol hypotheses for explaining the spatial and temporal

variability of the characteristics of deep convective clouds. Specifically, the impacts of

normalized convective available potential energy (NCAPE) and warm cloud depth (WCD) as

well as CCN concentrations (D $\geq$ 40 nm) on total lightning density (TLD), average height of 30

dBZ echoes (AVGHT30), and vertical profiles of radar reflectivity (VPRR) within individual

CFs are the subject of initial curiosity.

The results show that TLD increased by up to 600% and AVGHT30 increased by up to 2-

3 km with increasing NCAPE and CCN for fixed WCD on the global scale. The partial

sensitivity of TLD/AVGHT30 to NCAPE and CCN individually are found to be comparable in

magnitude, but each independent variable accounts for a fraction of the total range of variability

observed in the response (i.e., when the influences of NCAPE and CCN are considered

simultaneously). Both TLD and AVGHT30 vary inversely with WCD such that maxima of TLD

and AVGHT30 are found for the combination of high NCAPE, high CCN, and shallower WCD. The relationship between lightning and radar reflectivity is shown to vary as a function of CCN for a fixed thermodynamic environment. Analysis of VPRRs shows that reflectivity in the mixed phase region (altitudes where temperatures are between 0$^{\mathbf{o}}$C and -40$^{\mathbf{o}}$C) is up to 5.0-5.6 dB greater for CFs in polluted environments compared to CFs in pristine environments (holding thermodynamics fixed).

A statistical decomposition of the relative contributions of NCAPE, CCN, and WCD to the variability of convective intensity proxies is undertaken. Simple linear models of TLD/AVGHT30 based on the predictor set composed of NCAPE, CCN, and WCD account for appreciable portions of the variability in convective intensity ($R^2 \approx 0.3$-0.8) over the global domain, continents, oceans, and select regions. Furthermore, the results from the statistical analysis suggest that the simultaneous contributions from NCAPE, CCN, and WCD to the variability of convective intensity are often comparable in magnitude. There was evidence for similar relationships over even finer-scale regions [$O(10^6 \text{ km}^2)$], but differences in the relative prognostic ability and stability of individual regression parameters between regions/seasons were apparent. These results highlight the need to investigate the connection between statistical behavior and local meteorological variability within individual regions.

Following the global and regional analyses, data from Dynamics of the Madden-Julian Oscillation (DYNAMO) field campaign (2011-2012; central equatorial Indian Ocean (CIO)) and other sources was used to assess the relative impact of aerosols on deep convective clouds within a fine-scale environment with spatially homogeneous thermodynamics and variable aerosols in a pristine background over the CIO (CCN ~50-100 cm^{-3}, on average; NCAPE and WCD are hypothesized to be approximately constant, spatially). The experiment was designed to compare

differences in the convective cloud population developing in more-polluted and pristine regions, north and south of the equator, respectively. Analysis of the covariability of rainfall, cold cloud frequency, CCN, NCAPE, and lightning/radar reflectivity in deep convective clouds over multiple (> 20) episodes of the Madden-Julian Oscillation (MJO) leads to a hypothesis for a potential bi-directional interaction between aerosols and convective clouds that develop in association with the MJO. Close scrutiny of the results from climatology leads to the conclusion that thermodynamics and aerosols both influence deep convective cloud behavior over the CIO in a manner similar to that observed on the global scale, but the possibility that other factors are required to reproduce the full range of variability of deep convective clouds on fine-scales is acknowledged.

The research presented in this dissertation constitutes one of the first efforts to link the documented variability of radar reflectivity and lightning within convective features observed by the TRMM satellite to their environment using novel representations of thermodynamic and aerosol quantities from reanalysis and a chemical transport model, respectively. The independent variables studied here (i.e., NCAPE, CCN, and WCD) were chosen specifically to address preeminent hypotheses in the literature and the results from this investigation suggest that NCAPE, CCN, and WCD each contribute significantly to the variability of deep convective clouds throughout the Tropics and Subtropics (and perhaps seasonally). Implications of the findings from the current investigations and the relevance of these results to future studies are discussed.

ACKNOWLEDGEMENT

DEDICATION

I dedicate this work, the shining pinnacle of my scientific career to date, to my best friend and strongest advocate, Aliea; without your patience, loyalty, and love, this would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

From the jungles of Borneo, the Congo, and the Amazon, to the Sahel, to the lee side of the Andes and Rocky Mountains, to the remote reaches of the western Pacific and central Indian Ocean, the evolution from clear sky, to convective initiation, to the mature phase of deep convection and intense cloud electrification occurs with varying frequency on the global scale [*Boccippio et al.*, 2000; *Christian et al.*, 2003; *Zipser et al.*, 2006; *Liu et al.*, 2012; *Cecil et al.*, 2014]. To the common observer, strobe-like luminous discharges from a distant convective tower over the jungle or transient pulses illuminating the anvils of distant cumulonimbi rooted somewhere beyond the horizon warrant mere afterthoughts. For the atmospheric scientist, as he or she observes deep convective clouds (and the minute subset of them that produce lightning), a rapid flurry of curiosity surrounding cloud dimension, internal structure, morphology, and lightning flash rate ensues. *Why are regional populations of deep convective clouds incongruous in frequency, height, internal structure, and number? Why do some species of cumulonimbi produce little rainfall and copious lightning while other cumulonimbi produce significant rainfall and infrequent lightning? Furthermore, is the variable behavior of deep convection and lightning on large-scales attributable to specific characteristics of the tropospheric environment (e.g., conditional instability or the concentration of cloud nuclei)?*

This dissertation revisits long-standing questions surrounding an explanation for the observed variability of deep convection and lightning in the Tropics and Subtropics between continents and oceans as well as in individual regions. Two hypotheses stand out in the literature to explain land-ocean and regional contrasts in convective intensity, i.e., the thermodynamic

hypothesis and the aerosol hypothesis. Although a vast collection of literature relevant to thermodynamic and aerosol influences on deep convection exists [see *Williams and Stanfill*, 2002; *Williams et al.*, 2005; *Rosenfeld et al.*, 2008; and *Tao et al.*, 2012], recent advances (within the last two decades) in the atmospheric science community's ability to observe and model important aspects of the atmospheric system have revitalized the prospect of new understanding on the subject. The stated objective of this research is *to determine the relative roles of thermodynamics and aerosols as they contribute to the observed variability in deep convection and lightning on global and regional scales*.

A prerequisite to this research is a basic physical understanding of deep convective clouds, cloud electrification, and lightning. The launch of the TRMM satellite in 1997 initiated a collection of passive microwave and k_u-band radar observations from 36^oS to 36^oN latitude to allow scientists to develop side-by-side semi-global climatology of tropical lightning and precipitation. Subsequently, seminal analyses of individual precipitation features elucidated the intricate joint-frequency behavior of radar quantities (and the inferred presence of liquid and ice hydrometeors) and lightning [e.g., *Nesbitt et al.*, 2000; *Cecil et al.*, 2005; *Zipser et al.*, 2006; and *Liu et al.*, 2008], thereby leading to an important link between theory for thunderstorm electrification [e.g., *Takahashi*, 1978 and many others] and global-scale observations. Gradually, researchers studying the TRMM climatology began to note systematic discrepancies in radar-lightning correspondence [e.g., *Petersen and Rutledge*, 2001; *Liu et al.*, 2012] and general feature characteristics between continents and oceans, confirming earlier speculation about inherent phenomenological differences between convective clouds developing over continents and oceans, respectively.

The establishment of archetypes for "continental" and "maritime" convective clouds dates back to the middle part of the twentieth century when *Squires* [1958] noted distinct differences in cloud droplet spectra within warm convective clouds (i.e., growing cumulus clouds lacking ice with cloud tops below the local height of the 0°C isotherm) between maritime (oceanic) and continental locales. A strong inverse relationship between droplet number concentration and median diameter was found for "continental", "transitional", and "maritime" populations of cumuli, with "continental" ("maritime") cumuli exhibiting relatively small (large) median diameter and high (low) droplet number concentrations [*Squires*, 1958]. From the middle part of the twentieth century until the present, scientists have used data from various field campaigns throughout tropical and middle latitudes, e.g., the Thunderstorm project [*Byers and Braham*, 1948], Global Atmospheric Research Program's Atlantic Tropical Experiment [GATE; *Houze and Betts*, 1981], Tropical Ocean-Global Atmosphere-Coupled Ocean-Atmosphere Response Experiment [TOGA-COARE; *Webster and Lukas*, 1992], and Dynamics of the Madden-Julian Oscillation field campaign [DYNAMO; *Yoneyama et al.*, 2013], to develop a more rigorous separation for continental and maritime convective clouds (with and without lightning) globally. For example, investigations by *Byers and Braham* [1948], *LeMone and Zipser* [1980], *Zipser and LeMone* [1980], *Jorgensen and LeMone* [1989], *Szoke et al.* [1986], and *Zipser and Lutz* [1994] were among the earliest to document distinct differences in characteristic core widths, vertical motions, and reflectivity structures between convective clouds in continental and maritime environments.

Following previous hypotheses, land-ocean or regional contrasts in convective vigor and lightning are presumably related to the local availability of condensation nuclei, though it is possible that the thermodynamic characteristics of the environment differ between regimes

simultaneously [*Tao et al.*, 2012; *Altaratz et al.*, 2014]; such postulation is a microcosm for the main factors inhibiting understanding in studies that attempt to determine relationships between convective intensity and thermodynamics or aerosols individually.

The simplest approach to address the outstanding uncertainty surrounding thermodynamic/aerosol influences on deep convection is to evaluate "singular" hypotheses that argue for convective cloud variability being governed by thermodynamics *or* aerosols (of course there is the relevant null hypothesis that other environmental factors account for the variability of convective clouds). A considerable amount of previous research has demonstrated correlation between global attributes of deep convection, such as cloud top height/pressure, cloud fraction, precipitation rate/amount, and aerosol (cloud condensation nuclei, CCN) concentrations in the environment [e.g., *Rosenfeld et al.*, 1999; *Sherwood et al.*, 2006; *Koren et al.*, 2012; and many more]; however, many of the aforementioned studies cite uncertainty surrounding the covariability of thermodynamic metrics, such that the true magnitude of the aerosol influence remains unknown.

Research investigating the relative influence of aerosols on deep convection while controlling for environmental thermodynamics either by limiting the investigation region [e.g., *May et al.*, 2009,2011; *Altaratz et al.*, 2010; *Yuan et al.*, 2011; *Storer et al.*, 2014], instituting simultaneous observational strategies [e.g. *Williams et al.*, 2002; *Koren et al.*, 2010,2012; *Wall et al.*, 2014], or explicitly accounting for multiple independent variables in controlled cloud model experiments [e.g., *Storer et al.*, 2010; *van den Heever et al.*, 2011; *Storer and van den Heever*, 2013; *Li et al.*, 2013] are less common; results from these studies are in some cases apparently contradictory. For example, *Li et al.*, [2013] and *Sheffield et al.* [2015] contend that aerosols are responsible for the transition from shallow convective to congestus (or deep convective) modes,

4

especially in pristine-background environments, such as over remote reaches of the West Pacific Warm Pool and areas of the Atlantic Ocean. *Storer et al*., [2010] found approximately comparable increases in ice water path (~30-50%) in simulations of deep convective clouds for a 575% increase in CAPE (holding aerosols constant) and 400% increase in CCN (holding thermodynamics constant). Based on the findings of *Storer et al.* [2010] and theory for thunderstorm charging [e.g., *Saunders*, 1993], it is reasonable to expect comparable responses in lightning frequency with respect to CAPE and CCN. In contrast, in a study over the Amazon, *Williams et al.* [2002] observed negligible changes in lightning/radar reflectivity in the pre-monsoon period (decidedly continental conditions with high CAPE and relatively high cloud base height) as CCN trended from 3000 cm^{-3} to ~500 cm^{-3} between October and November 1999. *Williams et al.* [2002] concluded that their results, "…[cast] doubt on a primary role for the aerosol in enhancing the electrification," and go on to cite appreciable correlations between CAPE and CCN for their time period of interest over the Amazon, thus highlighting further ambiguity in their results surrounding the relative influence of thermodynamics and aerosols on deep convection.

Attempts to separate the simultaneous influence of aerosols on rainfall and convective cloud characteristics from background thermodynamics using observations on the global scale and in individual regions for continental and maritime populations of deep convective clouds are less numerous [e.g., *Koren et al.*, 2012]. *Koren et al.* [2012] matched rainfall observations from the TRMM satellite to aerosol optical depth measurements and reanalysis using a simple grid-based attribution scheme; strong increases in rain rate with increasing mid-level vertical velocity (a synoptic-scale proxy for thermodynamic instability) were noted and for a given thermodynamic environment, rainfall increased monotonically with increasing aerosol optical

depth, indicating simultaneous dependence in their sample. In a regional study of deep convective clouds over the tropical Atlantic Ocean, *Koren et al.* [2010] argued that small variations in the slopes of the relationships between cloud properties (e.g., cloud top pressure and cloud fraction) and aerosol quantities in different thermodynamic environments indicated near-independent forcing by aerosols and thermodynamics, respectively.

Hence, results from global and regional studies suggest that both thermodynamics and aerosols are important contributors to the observed variability of deep convective clouds on the large-scale, but the relative importance of each forcing mechanism, as they may contribute to lightning and radar reflectivity variability, has yet to be systematically quantified using global observations that are currently available. These investigations will use global satellite observations, reanalysis, and output from a global chemical transport model, to investigate the variability of lightning and deep convection in order to test a simultaneous interpretation of the thermodynamic and aerosol hypotheses (a more-detailed treatment of the "simultaneous" hypothesis follows in Sec. 2.1). As it will be shown, new understanding surrounding the variability of deep convective clouds and lightning may be gleaned by studying the variability of convective intensity proxies on multiple scales, while ensuring that the diversity of the observations spans as many combinations of environmental factors as possible; results from the analyses are presented in Chs. 2-4. The results are subsequently synthesized and contextualized within the scope of the relevant literature prior to a summary of research and conclusions in Ch. 5. Chapter 2 of this dissertation was published in the Journal of Geophysical Research – Atmospheres on 3 June 2015 (Sec. 2.5 constitutes a statistical investigation of the regional variations in the relationships between convective intensity proxies and environmental factors; the results from the statistical assessment have been amended to Ch. 2 and a separate manuscript

is in preparation). Chs. 3-4 have been combined into a separate manuscript for submission in the

Journal of the Atmospheric Sciences.

CHAPTER 2

Global and regional perspectives – The response of deep convection and lightning to thermodynamics and aerosols

2.1 Background

Over the past several decades, satellite observations have proven to be indispensable for characterizing convective clouds in the Tropics and Subtropics. Climatologies derived from the burgeoning data record have identified several important findings surrounding the phenomenon of deep convection. One prominent result is that lightning-producing convective features (LPCFs) occur predominantly over continental regions in the Tropics, whereas the majority of precipitating features occur over tropical ocean [*Orville and Henderson*, 1986; *Mackerras et al.*, 1998; *Boccippio et al.*, 2000, 2005; *Christian et al.*, 2003; *Cecil et al.*, 2005]. *Boccippio et al.* [2000] refined the understanding when they conducted an individual convective feature-based analysis across the Tropics and concluded that total lightning rates in continental thunderstorms are a factor of 2-3 larger compared to oceanic thunderstorms.

Regional analyses employing satellite instruments and ground-based radars have documented distinct differences in vertical precipitation structure throughout the Tropics associated primarily with intraseasonal variability [*Szoke et al.*, 1986; *Williams et al.*, 1992; *Rutledge et al.*, 1992; *Zipser and Lutz*, 1994; *Petersen et al.*, 1996; *DeMott and Rutledge*, 1998; *Rosenfeld and Lensky*, 1998; *Petersen and Rutledge*, 2001; *Cifelli et al.*, 2002, *Williams et al.*, 2002; *Petersen et al.*, 2006]. *Petersen and Rutledge* [2001] found a continuum of vertical reflectivity distributions corresponding to remote oceanic, coastal, and continental convection. A key finding was the higher frequency-of-occurrence of 30 dBZ radar echoes at temperatures colder than -10^{0}C (greater than ~6 km MSL) in tropical continental convection (compared to

tropical oceanic convection). The relative frequency of 30 dBZ echoes at these heights was considerably less over coastal and (especially) over remote oceanic regions.

In turn, these disparities in vertical precipitation structure have been related to appreciable differences in lightning. On average, deep convection that produces significant amounts of lightning is frequently the most vertically developed from the radar perspective [*Dye et al.*, 1989; *Rutledge et al.*, 1992; *Williams et al.*, 1992; *Zipser*, 1994; *Petersen et al.*, 1996; *Carey and Rutledge*, 2000; *Nesbitt et al.*, 2000; *Cecil et al.*, 2005; *Zipser et al.*, 2006; *Liu et al.*, 2012; *Stolz et al.*, 2014], whereas reduced lightning activity in deep convection is often associated with reduced reflectivities above the freezing level. Such close correspondence between lightning and radar characteristics is substantiated by both theoretical and laboratory studies which support the non-inductive mechanism (charge separation that occurs during collisions between ice particles in the presence of supercooled liquid within a cloud's mixed phase region, i.e., in the temperature range from 0°C to -40°C). It is generally accepted that this mechanism is the most plausible explanation for lightning initiation [*Reynolds et al.*, 1957; *Takahashi*, 1978; *Baker et al.*, 1987; *Saunders*, 1993; *Baker and Dash*, 1994]. Indeed, global distributions of ice-water path observed using satellites exhibit strong correlations with lightning and radar echo top climatology [*Petersen and Rutledge*, 2001; *Petersen et al.*, 2005].

Thus, a multitude of observations clearly illustrate fundamental differences in the make-up and electrical behavior of deep convection between tropical continental and oceanic regions. We are then left to question the underlying driver(s) of the observed variability in lightning and convective intensity. There are two hypotheses in the literature – the thermal hypothesis and the aerosol hypothesis – that attempt to explain the aforementioned variability.

The thermal hypothesis holds that the variability in lightning and convective intensity over continental and oceanic regions in the Tropics can be explained by differences in thermodynamic instability [*Rutledge et al.*, 1992; *Williams et al.*, 1992, 2002; *Williams and Stanfill*, 2002; *Williams and Sátori*, 2004]. One major caveat to the thermal hypothesis has been demonstrated in a number of analyses: the average convective available potential energy (CAPE), which represents the column-integrated energy available to accelerate parcels vertically, *is approximately the same over continents and oceans* [*Williams and Renno*, 1993; *Lucas et al.*, 1994a,b; *Halverson et al.*, 2002; *Williams and Stanfill*, 2002]. *Williams and Stanfill* [2002] argue that tropical land surfaces, with their relatively low heat capacity and high Bowen ratio, respond strongly to solar radiation and excite larger, more buoyant (energetic) parcels which ascend through a deep boundary layer and therefore are less susceptible to dilution via entrainment. Parcels in oceanic regimes have reduced thermal buoyancy in the lower and middle troposphere and may undergo more significant entrainment as a direct consequence of smaller parcel widths, resulting from ascent through a shallower boundary layer. The differences would suggest mixed-phase microphysics and non-inductive charging are both more robust in continental convection compared to oceanic convection, as wider updrafts contribute to a more efficient conversion of CAPE to updraft kinetic energy [e.g., *Williams et al.*, 2005]. *Williams et al.* [2005] found that flash rates increased with increasing cloud base height for a sample of tropical convective clouds (i.e., assuming that updraft width is proportional to cloud base height).

Citing results of both observation- and model-based studies, advocates of the aerosol hypothesis argue that the number of cloud condensation nuclei (CCN) in the environment of convective clouds significantly influences their microphysical and vertical development [*Rosenfeld and Lensky*, 1998; *Rosenfeld*, 1999; *Ramanathan*, 2001; *Andreae et al.*, 2004; *Graf*,

2004; *Khain et al.*, 2005; *Lohmann and Feichter*, 2005; *Sherwood et al.*, 2006; *van den Heever et al.*, 2006; *Hudson and Mishra*, 2007; *van den Heever and Cotton*, 2007; *Andreae and Rosenfeld*, 2008; *Bell et al.*, 2008; *Lee et al.*, 2008; *Rosenfeld et al.*, 2008 hereafter R08; *Koren et al.*,2010, 2012; *Li et al.*, 2011; *May et al.*, 2011; *Yuan et al.*, 2011, 2012; *Heiblum et al.*, 2012; *Niu and Li*, 2012; *Fan et al.*, 2013; *Li et al.*, 2013; *Lebo and Morrison*, 2014; *Storer et al.*, 2014; *Wall et al.*, 2014]. In their theoretical work, R08 provided a conceptual description of so-called aerosol-induced convective invigoration which has since been cited in many studies on the subject. Following their model, in convective clouds that develop in an environment with high aerosol concentrations (> 500 CCN cm^{-3}) precipitation formation by the collision-coalescence mechanism is hindered relative to clouds drawing on lower CCN concentrations. Substantial cloud water is then transported above the freezing level which upon freezing releases latent heat contributing to increased thermal buoyancy, stronger vertical motions, and greater charge separation. Note that R08 also emphasize the importance of offloading condensate in order for the maximum invigoration effect to be realized in and above the mixed-phase region (see Fig. 3 of that study). In more pristine environments (< 100 CCN cm^{-3}), collision/coalescence becomes very efficient which leads to rapid generation of precipitation, thereby reducing supercooled water contents in the mixed-phase region. This may explain the relatively low occurrence of lightning over remote oceanic regions.

We hypothesize that warm cloud depth (WCD), defined as the vertical distance between the lifted-condensation level (LCL) and the freezing level, could simultaneously influence the growth of cloud droplets by determining the duration of ascent through a cloud's warm phase and the subsequent development of precipitation by condensation or collision/coalescence [e.g., *Carey and Buffalo*, 2007; *Albrecht et al.*, 2011] in combination with aerosol indirect effects.

Takahashi [1978] communicated the importance of liquid water content for the charge separation process. For deep WCD, a relatively long trajectory through the cloud's warm phase implies that the probability of precipitation formation increases as autoconversion proceeds; therefore, cloud liquid may be lost prior to its arrival in the mixed-phase region where charge separation may decrease subsequently. Conversely, for shallower WCD, the likelihood that more cloud liquid reaches the mixed-phase increases for a fixed updraft and aerosol concentration. Lastly, for very shallow WCD (most applicable outside the Tropics), the duration of ascent through a cloud's warm phase is shortest for a fixed updraft velocity. In this case, the time period during which aerosols may be able to impact collision/coalescence is short and as a result there may be little if any sensitivity to aerosol concentration [e.g., *Li et al.*, 2011].

These ideas are supported by a number of model-based analyses of the impacts of aerosols on convective clouds in environments where WCD ranges from shallow to deep. *Fan et al.*, [2007] found that the influence of aerosols on the microphysical development of deep convective clouds was negligible when surface dew-point temperature depressions were large (i.e., dry boundary layers with resulting shallower WCD). In 3-D simulations of convective clouds in the deep Tropics (where WCD is deep), *Wang* [2005] found increasing precipitation as the initial CCN concentration was varied from 50 cm^{-3} to more than 5000 cm^{-3}, but decreases in precipitation efficiency were not observed even under heavily polluted conditions. In contrast, 2-D simulations of isolated convection in environments with intermediate WCD (WCD ~ 2 km) depicted increases in max updraft velocity of 5–6 m s^{-1} and enhanced lightning as aerosol concentrations were increased from 50 cm^{-3} to > 500 cm^{-3} [*Mansell and Ziegler*, 2013].

Results from regional and cloud-resolving model studies that investigate the robustness of aerosol effects in varied thermodynamic environments throughout the Tropics and Subtropics

depict non-trivial modifications to a cloud's hydrometeor population and subsequent dynamics [*Khain et al.*, 2005, 2012; *van den Heever et al.*, 2006; *van den Heever and Cotton*, 2007; *Lee et al.*, 2008; *Fan et al.*, 2013; *Storer and van den Heever*, 2013; *Venevsky*, 2014]. These results generally show enhanced pristine ice, as well as increases in graupel and hail at the expense of rain and drizzle in the cloud's warm phase when aerosol concentrations exceed ~100-400 cm^{-3} [e.g., *Storer and van den Heever*, 2013]. Collectively, these studies strengthen the case for aerosol effects on deep convection and lightning, but considerable uncertainty remains [e.g., *Stevens and Feingold*, 2009; *van den Heever et al.*, 2011; *Lee*, 2012; *Wall et al.*, 2014]. For example, differences in precipitation/updraft intensity within an individual cloud with respect to changes in aerosol concentrations may be buffered [e.g., *Stevens and Feingold*, 2009] or overwhelmed by the compensating circulations in the ensemble of deep convective clouds across a wider domain [e.g., *Lee*, 2012].

In their recent global analysis of the impacts of aerosol indirect effects on convective clouds, *Wall et al.* [2014] concluded, "…the true magnitude of the aerosol indirect effect [on deep convection] remains elusive on the global scale". They demonstrated regional sensitivity of deep convective clouds to aerosol load in the atmospheric column while accounting for meteorological factors, but their uncertainty may be the result of their choice to use column-integrated aerosol quantities estimated from satellites. *Wall et al.* were diligent in addressing the limitations of their chosen satellite aerosol data [e.g., *Várnai et al.*, 2013], but other investigators have found aerosol number concentrations to be more illustrative in studies of aerosol-convection interactions [e.g., *Hudson and Mishra*, 2007; *Koren et al.*, 2010].

We recognize that there may be a mutual-dependence between aerosols and thermodynamics that leads to stronger convection. Following R08, aerosols may influence how

much of the thermodynamic potential energy (i.e., CAPE) is realized by a parcel within a mature deep convective cloud. They argue that a shift from invigorated to suppressed convection should occur as a function of aerosol concentrations (as was exemplified by *Altaratz et al.,* 2010 over the Amazon). According to R08, the optimum aerosol load is that which balances the effects of modifying hydrometeor size distributions and subsequent differences in latent heating against water loading within rising parcels. From another point of view, the environmental thermodynamics dictate whether or not aerosol indirect effects on deep convection will be significant [e.g., *Morrison and Grabowski*, 2013]. In the absence of large-scale forcing (i.e., within baroclinic zones or due to topography), conditional instability in the thermodynamic environment is necessary to accelerate parcels with variable aerosol concentrations to their respective LCLs and potentially to sub-freezing temperatures in order for any microphysical response to aerosols to be observed.

This study addresses the question of how aerosols impact characteristics of deep convective clouds and attendant lightning production in the Tropics (over continents and oceans). Based on the above reasoning, we will consider the impacts of aerosols within the context of thermodynamic instability and WCD. Hence, we investigate the hypothesis that aerosols modulate the amount of available potential energy realized throughout the lifecycle of a convective cloud and this interaction is sensitive to WCD. We utilize reanalysis and the GEOS-Chem transport model to estimate relevant thermodynamic and boundary layer aerosol number concentrations globally in lieu of satellite methods for retrieving environmental thermodynamic and aerosol characteristics. By this method, we assure a large, representative sample of deep convective clouds in the Tropics. Additionally, we contribute new insight about the simultaneous importance of WCD.

2.2 Data and Methodology

This study employs data from the Tropical Rainfall Measuring Mission (TRMM), which in itself represents the most spatially and temporally comprehensive, global-scale data set consisting of simultaneous high-resolution lightning and radar observations (TRMM PR and LIS observations span all longitudes between 36^0N-36^0S and 38^0N-38^0S, respectively). The TRMM Convective Feature (CF) database (available, http://trmm.chpc.utah.edu) is comprised of individual groups of contiguous convective pixels observed by the TRMM precipitation radar (PR) that are determined by version 7 of the 2A23 'raintype' PR algorithm. A number of relevant radar and lightning characteristics are defined for each CF (e.g., maximum height of reflectivity echo tops, total lightning flash count, estimated time within the satellite's field-of-view, number of PR pixels within the CF). The TRMM Lightning Imaging Sensor (LIS) detects total lightning, i.e., contributions from intracloud and cloud-to-ground lightning sources, but is unable to distinguish between the two types of lightning. An eight-year temporal subset (2004-2011) of the full CF database (1998-2014) was chosen to maximize overlap with the available global GFED3 biomass burning inventory [*van der Werf et al.*, 2010] for the development of the aerosol data component, which will be described next.

A novel aspect of this research is its reliance on a global chemical transport model, GEOS-Chem (www.geos-chem.org) with the online aerosol microphysics module TOMAS [*Adams and Seinfeld*, 2002; *Pierce and Adams*, 2009; *D'Andrea et al.*, 2013, *Pierce et al.*, 2013], to provide estimates of lower tropospheric aerosol number concentrations on the global scale. GEOS-Chem-TOMAS simulates the particle size distribution from 3 nm - 10 µm in 15 size bins, and it tracks sulfate, sea-salt, organics, black carbon and dust aerosol species within these size sections. In this analysis, the GEOS-Chem-TOMAS aerosol fields are used to provide the

simulated number concentration of aerosols with diameters larger than 40 nm (D ≥ 0.04 μm).

This concentration is referred to as N40. Thus, this study does not account for differences in

aerosol composition because integrated number above cutoff diameters (as well as variability in

maximum supersaturations) accounts for most of the variability in cloud droplet number

concentrations [*Dusek et al.*, 2006]. It is expected that variability with respect to different

aerosol species (e.g. giant CCN from sea-spray, black carbon, and dust) could affect the results

and these sensitivities should be explored in future work.

In an analysis of particulate matter present within a variety of continental and marine

airmasses, *Dusek et al.* [2006] found that the CCN efficiency (i.e., the ratio of CCN to cloud

nuclei) of aerosols with diameters larger than 40 nm was strongly sensitive to the level of

supersaturation within cloudy parcels. For low supersaturations (< 1%), N40 was shown to

overestimate CCN concentrations, while for supersaturations a few tenths of a percent higher

than 1%, the CCN efficiency of N40 rapidly increased up to values exceeding 0.8. Both

observations and two-dimensional cloud scale simulations of growing congestus and

cumulonimbus clouds in the Tropics show that supersaturations at various heights above cloud

base range from a few tenths of a percent up to 3% or more [*Khain et al.*, 2012; 2013]. This

study focuses on deep convective clouds with strong updrafts that are capable of producing high

supersaturations immediately above cloud base. Therefore, N40 should be an appropriate proxy

for CCN in the context of this investigation; note that several aspects of this analysis were also

conducted using N80 data from the GEOS-Chem model runs and the results were approximately

unchanged (the correlation between N40 and N80 in our data sample was high, $r > 0.93$).

The simulations were run globally at a horizontal resolution of 2.5 degrees longitude by 2

degrees latitude (roughly 270 km by 220 km at the equator) for 47 vertical levels between

approximately 1000-0.01 hPa. In our analysis we use output between the 1000-850 hPa levels

(10 layers). Output was provided every 6 hours during the time period of interest over the

domain spanning all longitudes and between the latitudes 38°S - 38°N (corresponding to the

latitudinal extent of coverage for the TRMM LIS instrument). The spatial and temporal

autocorrelation of aerosol quantities has been shown to be near 0.8 for time and space scales of

200 km and 10 hours respectively in global observations [*Anderson et al.*, 2003]. Therefore, the

model resolution should adequately capture aerosol variability on the scales of interest for the

current study.

The choice to use a chemical transport model arose from the documented uncertainties

surrounding global satellite aerosol retrievals. The method requires observations as close as

possible to deep convective clouds, but the probability of contamination for passive retrievals of

aerosol characteristics by a satellite increases with decreasing distance to clouds [e.g., *Koren et

al.*, 2010]. Additionally, passive imagers flying onboard satellites suffer from the inability to

discern where in the atmospheric column the radiation reflected by aerosols originates and have

difficulty differentiating aerosols from clouds, especially over land surfaces and near optically

thick clouds where 3-D radiative effects are significant [*Várnai and Marshak,* 2009; *Várnai et

al.*, 2013].

Errors in GEOS-Chem simulated aerosol fields may impact the aerosol-cloud

relationships determined in this analysis. If errors in aerosol variability are independent of

cloud/lightning variability (e.g., random noise in the predicted aerosol fields as opposed to

systematic bias), then the errors will contribute to reduced sensitivity in the trends of convective

cloud characteristics versus aerosol concentrations. Therefore, it is argued that the results should

represent a lower bound for quantifying the influence of aerosols on the microphysical

development of deep convective clouds and lightning production. In an independent assessment

of the performance of the GEOS-Chem model [e.g., *D'Andrea et al.*, 2013], a comparison of

predicted and observed aerosol concentrations at 21 ground sites throughout North America and

Europe illustrated that the log-mean bias of predicted N40 in GEOS-Chem was less than ~|0.067|

(a factor of ~1.17) for all sensitivity experiments conducted.

The CF database incorporates thermodynamic information from the Interim Reanalysis

(ERAi) developed by the European Center for Medium Range Weather Forecasts (ECMWF)

[*Dee et al.*, 2011]. For each CF, ten levels of meteorological variables (Independent sensitivity

analysis illustrated that the global probability distribution of CAPE computed using all thirty-

seven levels of available reanalysis data is more accurately reproduced using just sixteen levels

of thermodynamic data (1000, 975, 950, 925, 900, 850, 800, 750, 700, 600, 500, 400, 300, 250,

200, 100 hPa) as opposed to the original ten levels. In the interest of computational efficiency,

the aforementioned sixteen levels of data were chosen at 2.5^0 horizontal resolution for this

analysis and are provided from the nearest neighboring reanalysis grid point. These variables are

linearly interpolated to the time of the TRMM overpass and are used to compute thermodynamic

variables of interest. Global aerosol climatology and satellite observations of aerosols have been

integrated into the ECMWF's data assimilation plan [*Morcrette et al.*, 2009; *Benedetti et al.*,

2009] and are accounted for in the radiative transfer scheme within the model [*Dee et al.*, 2011].

The methods for computing aerosol and thermodynamic quantities and attributing them to

individual CFs will now be described in more detail. For each CF between the years 2004 and

2011 (inclusive), we computed a vector-average of the horizontal wind components from ERAi

reanalysis between 1000 hPa and 850 hPa. We then defined an arbitrary "inflow" swath by a 90°

sector centered on the computed direction (0-360^0, with a direction of 360^0 denoting north).

Next the direction to all aerosol grid points relative to the geographic coordinates of each CF was calculated and the grid points within the upstream swath sector at a distance of less than 350 km were identified. After linearly interpolating the aerosol data to the time of the TRMM overpass, boundary-layer average (~1000-850 hPa) CCN concentrations (N40) were computed at the identified grid points and averaged. The average of boundary layer N40 in the upstream swath was then assigned to individual CFs.

ERAi profiles of temperature and moisture at the nearest reanalysis grid point within the upstream swath were used to calculate mixed-layer (lowest 50 hPa) pseudoadiabatic CAPE. The difference between the geopotential heights of the approximate levels of free convection and neutral buoyancy within each reanalysis sounding were found; this quantity represents the depth over which the idealized parcel's perturbation temperature was positive. Normalized CAPE (NCAPE) was determined for each CF by dividing the mixed-layer pseudoadiabatic CAPE by the depth of the positive area in the sounding [*Blanchard*, 1998]. Following earlier studies [e.g., *Lucas et al.*, 1994b], the NCAPE represents the amount of thermodynamic instability in the environment but takes the "shape-of-the-CAPE" into account via the normalization factor. For example, NCAPE = 0.1 J kg^{-1} m^{-1} could be representative of CAPE = 1000 J kg^{-1} distributed over a depth of 10 km (as is common in the Tropics). Comparison of CAPE and NCAPE across the global domain for the eight-year period illustrated a generally high correlation, but greater spread in the distribution was noted for generally high values of both metrics.

The derivations of CAPE and NCAPE incorporate a number assumptions that have been the subject of debate in the community, e.g., a critical assumption of no mixing between the parcel and the environment. In fact, parcels do entrain a significant amount of ambient air [e.g., *Romps and Kuang*, 2010; *McGee and van den Heever*, 2014] throughout the course of ascent

(especially true over tropical oceans where thermal/plume widths are relatively small) and subsequent mixing processes serve to homogenize thermal and moisture perturbations within parcels. Furthermore, there are known moisture biases within the reanalysis product used in this analysis [e.g., *Vesperini*, 2002] that would lead to increasing uncertainty in our calculations. We acknowledge these shortcomings of the chosen data and emphasize that these metrics for thermal instability (CAPE and NCAPE) are only estimates of the potential intensity of deep convection based on the available data. That being said, these datasets are considered the best available to work with in large scale analyses.

This study emphasizes the role of NCAPE as an estimator for the potential intensity of deep convection. It is possible that NCAPE can be ambiguous in this context because similar values may be found for different thermodynamic environments. For example, the NCAPE computed for a sounding with a shallow layer of positive area with a relatively small value of CAPE may be comparable to the NCAPE in a sounding with a deep layer of positive area and large CAPE. In each case, convection that develops is likely to be very different. When we looked at the distribution of the depth of positive area in each sounding in our data subset, we found a sharp peak in frequency near a depth of 10 km (depths < 4 km accounted for roughly just 5% of our sample) in line with unstable conditions observed in the tropical atmosphere. The fact that the majority of soundings with NCAPE > 0 J kg^{-1} m^{-1} in our sample exhibited deep layers of positive parcel buoyancy increased our confidence in using NCAPE as metric for thermodynamic instability in the investigation.

The LCL was approximated by taking the difference between a parcel's surface temperature and dew-point temperature, and then multiplying by a constant, $c=0.12$ km K^{-1} [e.g., *Iribarne and Godson*, 1981]. Next, the local height of the 0°C isotherm was estimated by linear

interpolation from the nearest neighbor in the ERAi vertical temperature profile. The approximations for the LCL and the local freezing height were found to agree generally well with observations from the global upper-air network. We then calculated the WCD for each CF by finding the vertical distance between the approximate LCL and the freezing level.

The mean vertical profile of radar reflectivity (VPRR) was computed for each CF by cross-referencing the version 7 2A25 attenuation-corrected reflectivity profile [*Iguchi et al.*, 2000]. The geographic centroid of each CF was noted and the nearest *convective* pixel (according to the version 7 2A23 raintype algorithm) was identified as the reference position (ray and scan indices) within the orbital 2A25 data array. Next, "SEARCH2D" software from the Interactive Data Language (www.exelisvis.com) was used to index array positions that had both convective precipitation and a continuous path of connectivity to the reference position. The angle between the local zenith and the slant path of the radar beam was then used to calculate the height of radar returns in each VPRR using basic trigonometry. The mean VPRR for individual CFs was then computed by taking the mean of the linear reflectivity at indexed array positions (in the horizontal dimensions) at all 80 heights assuming a 250 m interval in the vertical. To mitigate the effects of near-surface ground clutter, we restricted the VPRR to all PR returns above 1.5 km altitude (approximately less than PR range bin 73).

Importantly, the average height of 30 dBZ echoes (AVGHT30) is defined to be the peak altitude in the *mean* VPRR where the reflectivity was between 30.0-39.9 dBZ (inclusive). Note, however, that the definition of AVGHT30 differs from the maximum height of 30 dBZ echoes (MAXHT30) in the original CF database since the latter refers to the *single maximum altitude* within a CF where the reflectivity is greater than or equal to 30 dBZ. AVGHT30 takes the area-average behavior of the 30 dBZ echo top height surface into account, whereas MAXHT30

indicates the behavior of the peak(s) of the 30 dBZ echo top height surface (i.e., the single strongest convective pixel/core resolvable with TRMM PR). Lastly, the elevation of the surface (i.e., the 'elev' parameter in the CF dataset) has been subtracted, such that AVGHT30 is representative of height relative to the ground surface. By definition, the grouped TRMM PR pixels that constitute a CF contain shallow and deep convective pixels and we adhere to this standard definition in the current study (as opposed to excluding shallow convective pixels within each CF/LPCF). Therefore, in the computation of mean VPRR and AVGHT30, the presence of shallow convective pixels within the larger CF would tend to offset the contributions from deep convective pixels and possibly weight the 30 dBZ echo height statistics toward lower altitudes.

The CF database algorithm attributes lightning flashes to individual CFs if the flash location falls within the boundary of grouped convective pixels. Total lightning flash rate was estimated by taking the quotient of the flash count and the viewtime, or the estimated time during which the CF was in LIS's field-of-view (units of flashes min.$^{-1}$). Total lightning density (TLD) for each CF was then computed by taking the estimated total lightning flash rate and normalizing by the approximate feature area (km^2). While TRMM LIS observations extend from 38^0S-38^0N, CFs and LPCFs are defined only within the latitudinal limit of the TRMM PR observations, i.e., 36^0S-36^0N; thus our results apply to CFs and LPCFs within the domain of TRMM PR observations.

There are multiple potential "noise" sources that could result from the above feature attribution scheme coupled with the TRMM satellite observation strategy. Most importantly, given the myriad of convective scale processes that are not resolved by the relatively coarse thermodynamic and aerosol grids and the impossibility of knowing *exactly* which

thermodynamic environments and aerosol concentrations each CF interacts with, scatter about some central value is expected. For example, our preliminary analyses showed appreciable total lightning flash rates occurred in LPCFs where our scheme attributed the lowest values of NCAPE. It was determined that these LPCFs occurred in the vicinity of large gradients in thermodynamic instability; these LPCFs accounted for a small fraction of the dataset (~6% of LPCFs) and were excluded from further analysis due to the uncertainties surrounding their respective thermodynamic environments.

The TRMM satellite has a forward propagation speed of ~7 km s^{-1} such that any CF may be in the instrument field of view for a maximum of ~90 s. Then it is expected that "snapshot" observations by TRMM PR/LIS will depict CFs at various stages in their respective life cycles in any given orbit − constituting another potential source of variability. Additionally, close inspection of orbital level data from PR/LIS for individual CFs illustrated that larger CFs sometimes have significant lightning observed beyond the PR swath boundary so the radar/lightning correspondence may be compromised. We tried filtering CFs within varying distances from the PR swath edge (cross-track PR pixels 0 and 49) and found that our results were insensitive to this limitation in the data. We chose to include all CFs, regardless of their proximity to swath boundaries, in an effort to maximize the strength of our statistical findings.

Separate data populations of CFs for continental and oceanic regions, both with and without lightning, were simultaneously stratified by the three independent parameters (NCAPE, CCN, and WCD) to test various aspects of our hypothesis. The data were first separated by WCD, then the data were binned by CCN and NCAPE using set intervals in order to facilitate comparison between specific environments of interest. A considerable effort went to into understanding how WCD varies throughout the tropical/subtropical domain. The probability

density functions of WCD were computed at every reanalysis gridpoint and time step for the 8-year period (regardless of whether there was an overpass by TRMM); the median WCD in the domain of interest was found to be very close to 4200 m. Therefore, discussion of differences with respect to "shallower" and "deeper" WCD in this paper are relative to WCD = 4200 m.

In our examination of VPRR for the global sample of CFs, we investigated the potential impacts of aerosols in fixed background thermodynamic environments. It is difficult to rigorously define pristine and polluted aerosol environments between continents and oceans since the underlying aerosol distributions are so different [e.g., *Williams et al.*, 2002 and references therein]. For this reason, we introduce stratifications for the range from pristine to heavily polluted environments (CCN < 100 cm^{-3}, 100-200 cm^{-3}, 200-500 cm^{-3}, > 500 cm^{-3}) based on the cumulative probability distribution function for the global population. We use average 30 dBZ echo top height (AVGHT30), TLD, and VPRR as proxies for the variability of convective intensity throughout the Tropics in the following investigation [e.g., *Zipser et al.*, 2006; and many others].

2.3 Results

2.3.1 Global Climatology

Roughly 12.2 million CFs were observed by the TRMM satellite between 2004 and 2011 but we decided to include only CFs and LPCFs with 1) AVGHT30 > 5 km and 2) with collocated aerosol and thermodynamic variables within the upstream swath in our final data subset. The choice to truncate the data to CFs/LPCFs with AVGHT30 > 5 km reduced our data subset to just under 1.5 million CFs (~260,000 LPCFs; see Table 2.1 for a sample sizes of each population). However, by this method, we attempt to isolate quasi-upright, deep convective

features since we expect shallow and deep CFs to respond differently to perturbations to environmental thermodynamics and aerosol concentrations [e.g., *Tao et al.*, 2012; *Rosenfeld et al.*, 2014]. We then binned these populations of CFs and LPCFs in the 8-year temporal subset geographically on a 1^0 grid and produced global distributions of CF/LPCF frequency, mean AVGHT30, mean TLD, as well as, mean distributions of NCAPE, CCN, and WCD for both CFs (Figs. 2.1a-e) and LPCFs (Figs. 2.1 f-j).
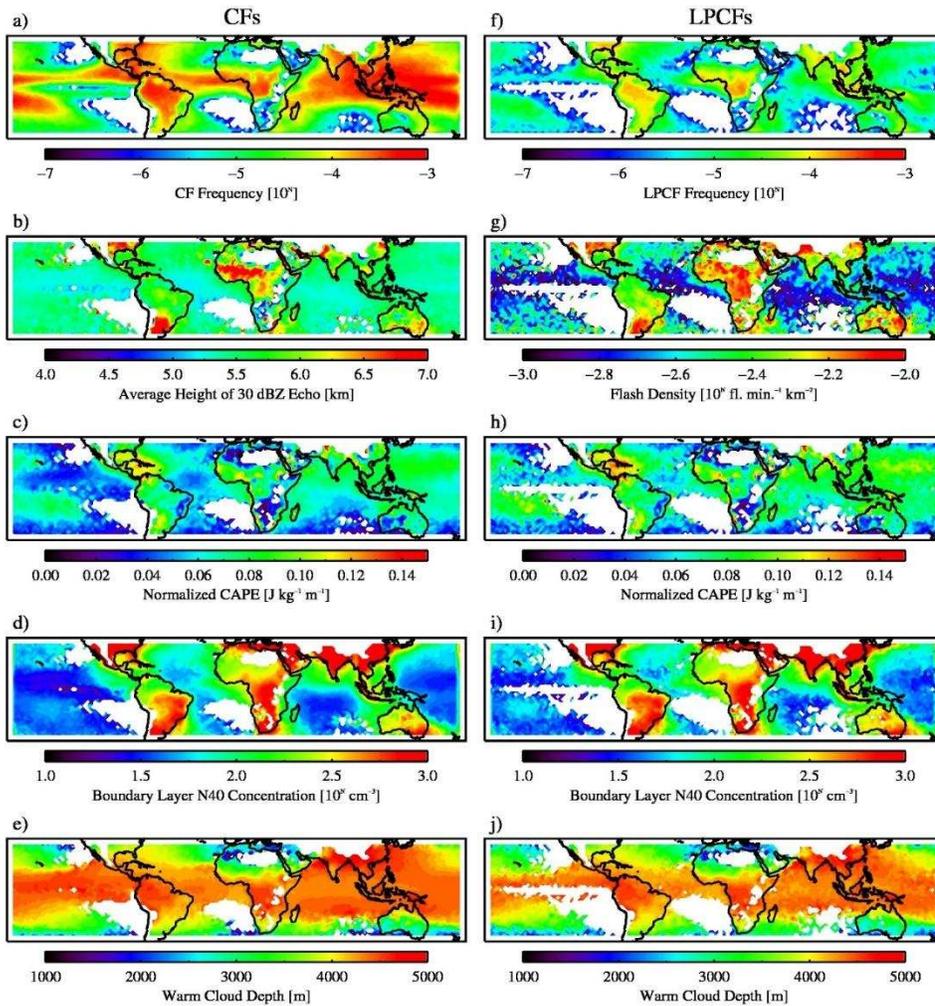
Fig. 2.1. Annual mean distributions of a) CF frequency, b) AVGHT30, c) NCAPE for CFs, d) CCN for CFs, e) WCD for CFs, f) LPCF frequency, g) TLD, h) NCAPE for LPCFs, i) CCN for LPCFs, and j) WCD for LPCFs averaged over 1^0 grid boxes between 38^0S and 38^0N for the years 2004-2011 (see text for definitions). CFs (LPCFs) with AVGHT30 > 5 km (with flash rates above the minimum detection threshold of the TRMM Lightning Imaging Sensor) and were analyzed (both constraints in the case of LPCFs).

Globally, the frequency of deep CFs maximizes in the Intertropical Convergence Zone

and over the western Pacific Ocean (Fig. 2.1a). Deep CFs are nearly absent from subsidence

regions on the eastern periphery of major ocean basins. The global distribution of AVGHT30 for

the subset of deep CFs (Fig. 2.1b) shows heights of < 6 km over oceans and 6 km to more than 7

km over continents in the Tropics and Subtropics on average. These differences generally agree

with the spatial variability of lightning shown in Fig. 2.1g, but the strength of correlation

between these two fields appears to vary between continents and oceans. The maxima in the

AVGHT30 distribution occur over the Sahel, Southern Great Plains of the United States, in lee

of the central Andes Mountains, and over parts of northwest India. Notably, the distribution of

MAXHT30 for LPCFs (and to some extent, for CFs) in our data subset show roughly similar

values, ~9 km, between continents and oceans over much of the TRMM domain (not shown).

The homogeneity of the MAXHT30 climatology for LPCFs coupled with the striking land-ocean

contrast in TLD constitutes an important set of findings and will be discussed in a subsequent

section.

Table 2.1. Census of CFs and LPCFs[a]

Domain	CFs	LPCFs	CFs (AVGHT30 > 5 km)	LPCFs (AVGHT30 > 5 km)
Global	12,232,564	503,133	1,457,919	263,378
Continents	2,148,492	360,190	388,859	162,593
Oceans	10,084,072	142,943	1,069,060	100,785

[a]Populations of convective features (CFs) and lightning-producing convective features (LPCFs; i.e., CFs with lightning flash rates above TRMM LIS's minimum detection threshold) over the TRMM domain for the years 2004-2011. The first two columns represent populations in the original CF/LPCF database, while the latter two columns represent the sample sizes of the data subset for the current analysis.

In agreement with previous studies, the frequency of LPCFs over continents is

consistently an order of magnitude larger than the frequency of LPCFs over ocean (Fig. 2.1f).

However, the vast majority of CFs (including CFs without lightning) occur over ocean (cf.,

Table 2.1). CFs with lightning flash rates above LIS's minimum detection threshold (~0.7

flashes min.$^{-1}$) represent only a small fraction of the total number of CFs observed (approximately 3.4% of CFs observed by the TRMM satellite between 2004-2011). Although LPCFs in our data subset account for a small fraction of the full CF database, prominent aspects of the general circulation like the Intertropical and South Pacific Convergence Zones, the African Easterly Jet, and midlatitude storm tracks are readily noticeable in Fig. 1f as these areas are favorable for the development of deep convective clouds.

The annual mean distribution of TLD (Fig. 2.1g) depicts the strong climatological land-ocean contrast in lightning that has been previously documented [*Christian et al.*, 2003]. TLD over continents is greater than TLD over ocean by a factor of 2-5 [e.g., *Boccippio et al.*, 2000]. While intense convection capable of producing copious lightning does occur over remote tropical oceans [e.g., *Kelley et al.*, 2010], Fig. 2.1f shows that such convection occurs very infrequently compared to continental areas. Abrupt decreases in TLD near coastlines are also shown in Fig. 2.1g in agreement with previous global lightning climatologies.

In general, the difference in NCAPE between continental and oceanic regions across the TRMM domain for both CFs and LPCFs is small in the annual mean sense (Figs 2.1c,h). At first, this result was surprising given the emphasis on the importance of thermal instability in moist convective processes in the literature to date. However, several researchers have argued that the difference in conditional instability between continents and oceans in the Tropics is small on average [e.g., *Williams and Renno*, 1993; *Lucas et al.*, 1994a,b], consistent with the results found here. The annual mean distributions of the depth of the positive area (i.e., the normalization factor in the NCAPE computation) in a given sounding depict general longitudinal homogeneity (depth ~ 10 km) and general decreases in the poleward direction (not shown). The

differences in the distributions of NCAPE for CFs (Fig. 2.1c) and NCAPE for LPCFs (Fig. 2.1h) are not readily discernible, but will be discussed in more detail below.

The annual distributions of boundary-layer CCN for CFs and LPCFs in Figs. 2.1d,i are consistent with observations of an order-of-magnitude difference in the mean aerosol concentrations between continents and oceanic regions. The most pristine environments are the equatorial regions of the Pacific and Indian Oceans with mean CCN values near a few 10's cm^{-3}. Globally, CCN concentrations are maximized over continental regions in developing counties and in areas where seasonal biomass burning takes place (e.g., central equatorial Africa and the southern/southeastern portions of the Amazon). Several oceanic regions are subject to offshore aerosol transport as evidenced by CCN reaching as high as a few hundred per cubic centimeter (e.g., downstream of the eastern United States, South America, South Africa, and eastern Australia).

The global distributions of WCD (Figs. 2.1e,j) show considerable homogeneity in the deep Tropics. The greatest distance between the freezing level and LCL, on average is found over parts of the Amazon, equatorial Africa, Southeast Asia, and the Indian and Pacific Oceans for both of the populations of CFs and LPCFs. The fact that WCD generally decreases with increasing latitude is likely due to decreases in boundary layer moisture and the lowering of the local freezing level. Zonal heterogeneity of WCD is most pronounced in the northern hemisphere over continents (e.g., between the United States, northern Africa, and Southeast Asia).
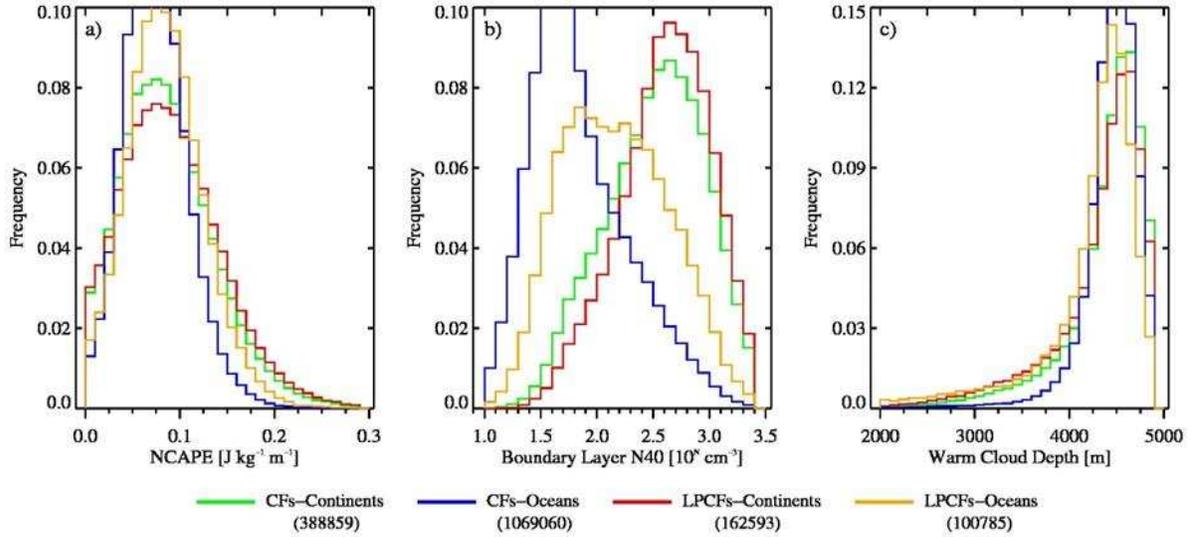
Fig. 2.2. Probability density functions of a) NCAPE, b) CCN, and c) WCD for CFs over continents (green), CFs over oceans (blue), LPCFs over continents (red), and LPCFs over oceans (orange). The number of CFs or LPCFs is provided in parentheses for each population. CFs (LPCFs) with AVGHT30 > 5 km (with flash rates above the minimum detection threshold of the TRMM Lightning Imaging Sensor) and were analyzed (both constraints in the case of LPCFs).

Probability density functions for NCAPE, CCN, and WCD for the populations of CFs and LPCFs over both continents and oceans are presented in Fig. 2.2. The median NCAPE for CFs (for LPCFs) is 0.08 J kg^{-1} m^{-1} (0.09 J kg^{-1} m^{-1}) over continents and 0.07 J kg^{-1} m^{-1} (0.08 J kg^{-1} m^{-1}) over oceans. While the distributions of NCAPE (Fig. 2.2a) peak near 0.08 J kg^{-1} m^{-1} for each population of CFs and LPCFs, there is a tendency for CFs over continents and LPCFs to develop in environments with higher NCAPE more frequently.

There are distinct differences between the distributions of CCN for CFs and LPCFs over continents and oceans. Fig. 2.2b shows that CFs over oceans develop in environments that are most frequently characterized by CCN < 100 cm^{-3}. In contrast, the distribution of CCN for CFs that develop over continents peaks around 300-500 cm^{-3}. LPCFs over oceans develop in environments with CCN values that are intermediate between those values typical for oceanic and continental CFs. Lastly, the distribution of CCN for LPCFs over continents peaks at approximately 300-500 cm^{-3}; however, the relative frequency of occurrence of LPCFs over

29

continents in heavily-polluted environments (CCN > 1000 cm^{-3}) is higher than that for any other subset of the total CF population considered.

The distributions of WCD for CFs and LPCFs (Fig. 2.2c) each peak near 4500 m (slightly deeper than the global median value computed from all reanalysis time steps) and there is a subtle tendency for LPCFs over continents and oceans to occur more frequently at shallower WCD when compared to the relative frequency of occurrence for the populations of CFs over continents and oceans. It is also apparent that there are relatively few observations of WCD shallower than 2000 m in our data subsets and therefore we limit the analysis to WCD between 2000-5000 m.

While the global climatological and probability distributions in Figs. 2.1-2.2 lend physical credence to the behavior of each variable, they only allow for qualitative characterization of potential relationships between proxies for convective intensity and NCAPE, CCN, and WCD. Furthermore, these climatological distributions support the idea that the variability in TLD and AVGHT30 could result from some combination of NCAPE, CCN, and WCD. We now turn to a discussion of this topic.

2.3.2 Three-parameter stratification: total lightning density and average height of 30 dBZ

We begin our discussion by examining the relationships between NCAPE, CCN, WCD, and TLD. For the global population of LPCFs, we find that TLD increases from < 0.0014 fl. min.$^{-1}$ km^{-2} up to 0.01 fl. min.$^{-1}$ km^{-2} (a 600% increase) with *both increasing CCN and NCAPE for all simultaneous stratifications of WCD* (Figs. 2.3a-d). Additionally, the highest TLD occurs for CCN > 1000 cm^{-3} and NCAPE > 0.15 J kg^{-1} m^{-1} while the lowest TLD is generally found below CCN < 500 cm^{-3} and NCAPE < 0.15 J kg^{-1} m^{-1} for all WCD stratifications. Finally, as

WCD trends from shallower (Fig. 2.3a) to deeper (Fig. 2.3d), the highest median values of TLD become restricted to higher values of both NCAPE and CCN, suggesting a modulation of the NCAPE-CCN-TLD relationship by WCD. Fig. 2.3 also illustrates important conditional or partial sensitivities of TLD to NCAPE and CCN (e.g., changes in TLD along a constant vertical/horizontal parameter space trajectory holding other variables constant). The partial sensitivity of TLD with respect to changes in NCAPE and CCN is systematically detailed in Sec. 2.3.3 below.
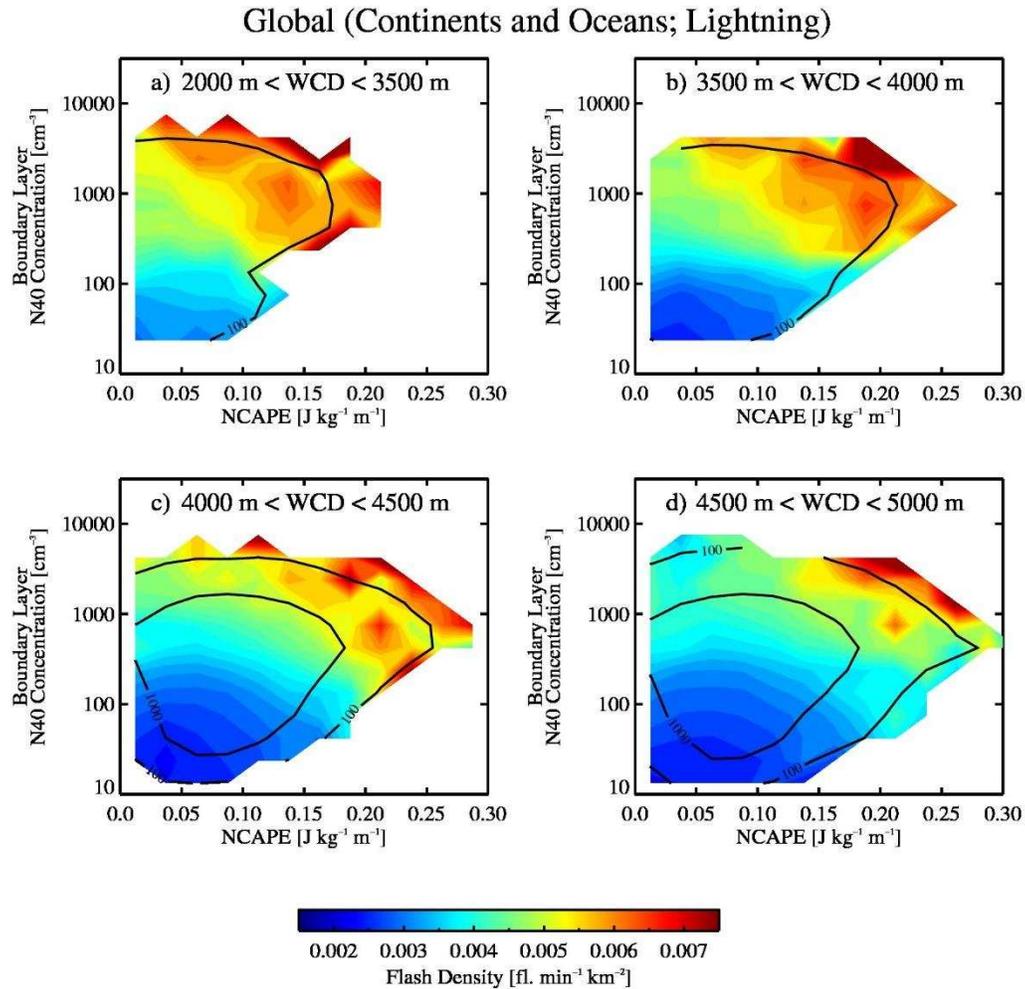
Fig. 2.3. Total lightning density (TLD; shaded) plotted as a two dimensional function of NCAPE and CCN for CFs across the global TRMM domain (continents and oceans) for a) 2000 m < WCD ≤ 3500 m, b) 3500 m < WCD ≤ 4000 m, c) 4000 m < WCD < 4500 m, and d) 4500 m < WCD < 5000 m. Black solid contours indicate the number of observations in each bin within the parameter space. LPCFs with flash rates above the minimum detection

threshold of the TRMM Lightning Imaging Sensor and AVGHT30 > 5 km were analyzed. A threshold of twenty LPCFs was set for a given bin before the output was plotted.

Next, we examined the response of AVGHT30 in the population of CFs with AVGHT30 > 5 km (shallow CFs, i.e., features with AVGHT30 < 5 km, were excluded). The distributions of AVGHT30 for these CFs in response to NCAPE and CCN for simultaneous stratifications of WCD (Figs. 2.4a-d) are broadly consistent with the results shown in Fig. 2.3. AVGHT30 increases from 5.25 km up to more than 7.5 km as both NCAPE and CCN increase, with the peak AVGHT30 found for CCN > 1000 cm^{-3} and NCAPE > 0.15 J kg^{-1} m^{-1} while minima are found for CCN < 500 cm^{-3} and NCAPE < 0.15 J kg^{-1} m^{-1} for all WCD stratifications. Analogous to Fig. 2.3, there are again aspects of the two-dimensional dependence of AVGHT30 on NCAPE and CCN, i.e., partial sensitivities, and these sensitivities will be discussed in Sec. 2.3.3 below.
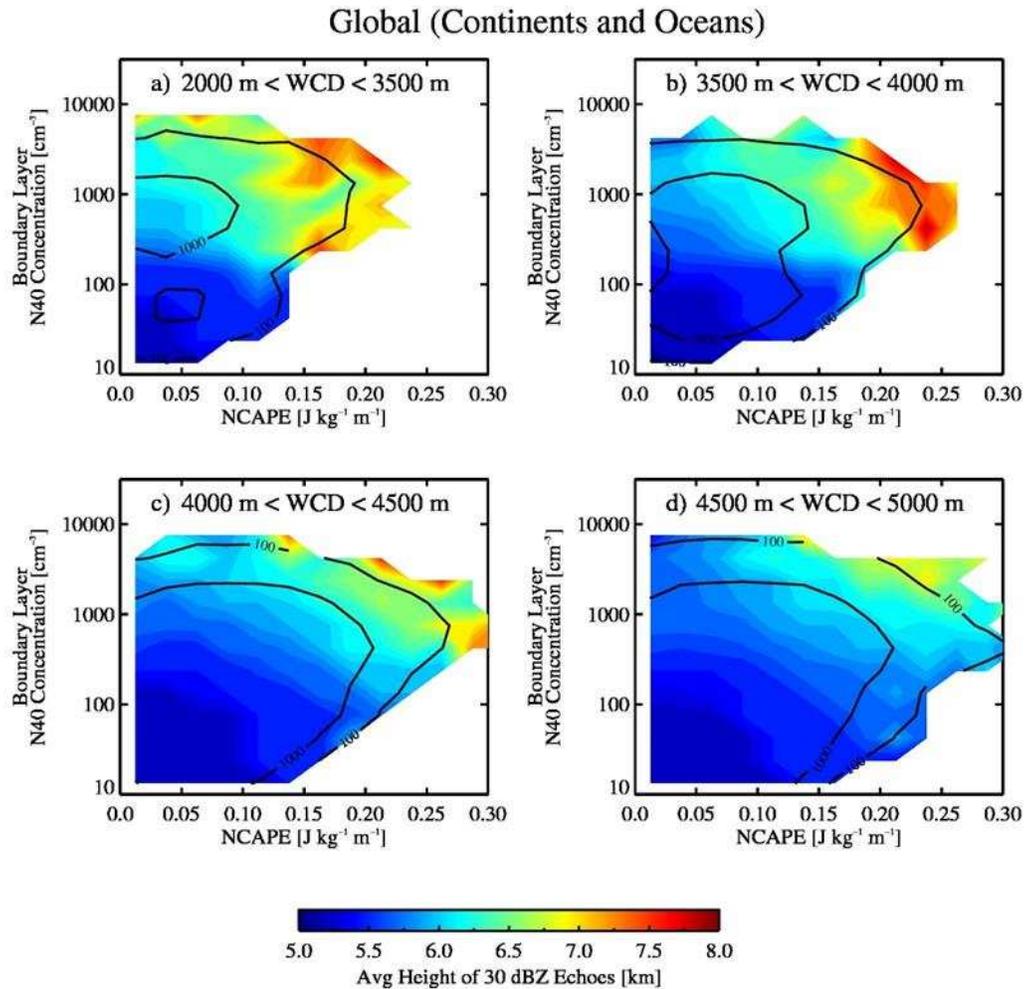
Fig. 2.4. As in Fig. 2.3, but for the average height of 30 dBZ echoes (AVGHT30; shaded). CFs with AVGHT30 > 5 km were analyzed. A threshold of twenty CFs was set for a given bin before the output was plotted.

Our next objective was to determine whether the response of these proxy measures for convective intensity varies for CFs/LPCFs between continents and oceans. From climatology in Fig. 2.1, LPCFs occur predominantly over continents. On average, TLD is greater and AVGHT30 is higher over continents compared to oceanic regions. LPCFs over continents and oceans are shown to populate different sections of the NCAPE-CCN parameter space (Fig. 2.5); LPCFs over continents occur in more polluted environments across a wide range of NCAPE values while LPCFs over oceans occur in environments with lower NCAPE and generally low aerosol number concentrations. Here, we present only the extremes of the WCD stratifications in

33

our sample to illustrate potential variability with respect to shallower and deeper WCD. Do the

differences in the intrinsic properties of the environments for continental and oceanic CFs impact

the resulting sensitivities of lightning and radar quantities to NCAPE and CCN?
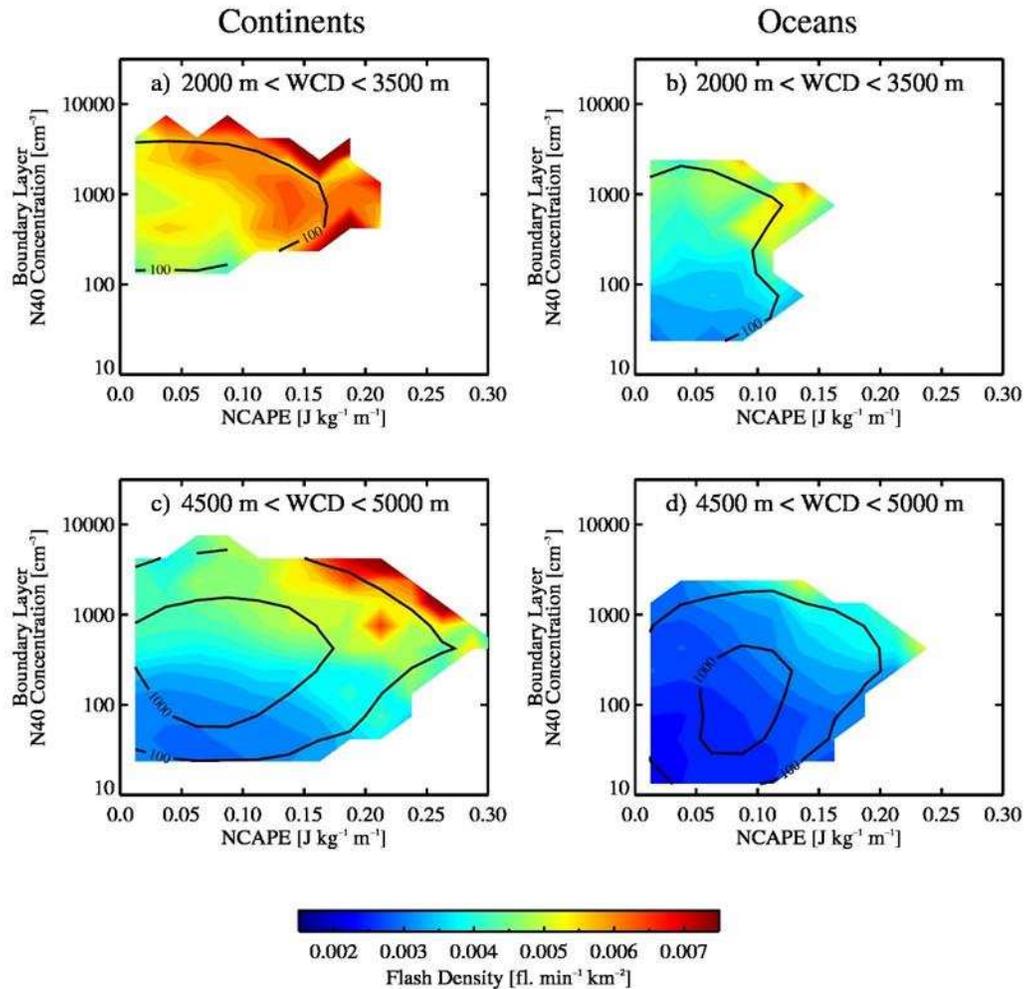
Fig. 2.5. Total lightning density (TLD; shaded) plotted as a two dimensional function of NCAPE and CCN for LPCFs over continents (left panel) and oceans (right panel) for a,b) 2000 m < WCD ≤ 3500 m and c,d) 4500 m < WCD < 5000 m. Black solid contours indicate the number of observations in each bin within the parameter space. LPCFs with flash rates above the minimum detection threshold of the TRMM Lightning Imaging Sensor and AVGHT30 > 5 km were analyzed. A threshold of twenty LPCFs was set for a given bin before the output was plotted.

Over continents for shallower WCD (Fig. 2.5a), TLD is shown to increase from about

0.0033 fl. min.$^{-1}$ km^{-2} up to more than 0.012 fl. min.$^{-1}$ km^{-2} (a 268% increase) as NCAPE and

CCN increase together. Over oceans for shallower WCD (Fig. 2.5b), TLD increases from

approximately 0.0018 fl. min.$^{-1}$ km^{-2} up to 0.006 fl. min.$^{-1}$ km^{-2} (roughly a 300% increase) as

NCAPE and CCN simultaneously increase for fixed WCD. For deeper WCD (Figs. 2.5c,d), TLD increases by roughly 300-400% again as both NCAPE and CCN increase simultaneously. In all cases (both shallower and deeper WCD), the minimum TLD is found for NCAPE < 0.10 J $kg^{-1} m^{-1}$ and CCN < 500 cm^{-3} and the maximum TLD is generally found for NCAPE > 0.15 J $kg^{-1} m^{-1}$ and CCN > 1000 cm^{-3}.

We conducted a similar analysis for the sensitivity of AVGHT30 over continents and oceans and the results for shallower and deeper WCD are shown in Fig. 2.6. Over continents (Figs. 2.6a,c), AVGHT30 increases from approximately 5.5 to 8.0 km for shallower WCD and approximately 5.25 to 7.25 km for deeper WCD as NCAPE and CCN increase simultaneously. Over oceanic regions for shallower and deeper WCD (Figs. 2.6b,d), AVGHT30 increases from roughly 5.25–6.75 km and 5.25-6.25 km, respectively, as both NCAPE and CCN increase, with the maximum values of AVGHT30 found for NCAPE > 0.10 J $kg^{-1} m^{-1}$ and CCN ~ 1000 cm^{-3}. The minimum values of AVGHT30 are again found for low NCAPE and low CCN.

Together the results for the sensitivity of TLD and AVGHT30 with respect to NCAPE and CCN simultaneously over continents and oceans are consistent with the behavior observed for the global populations of CFs and LPCFs (Figs. 2.3-2.4). Furthermore, the variability of AVGHT30 over continents and oceans mirrors the response of TLD to combinations of NCAPE/CCN for fixed WCD, emphasizing potential correlation between 30 dBZ radar reflectivity and lightning characteristics in convective features (analysis of the relationship between lightning and 30 dBZ radar reflectivity will be provided in a subsequent section).
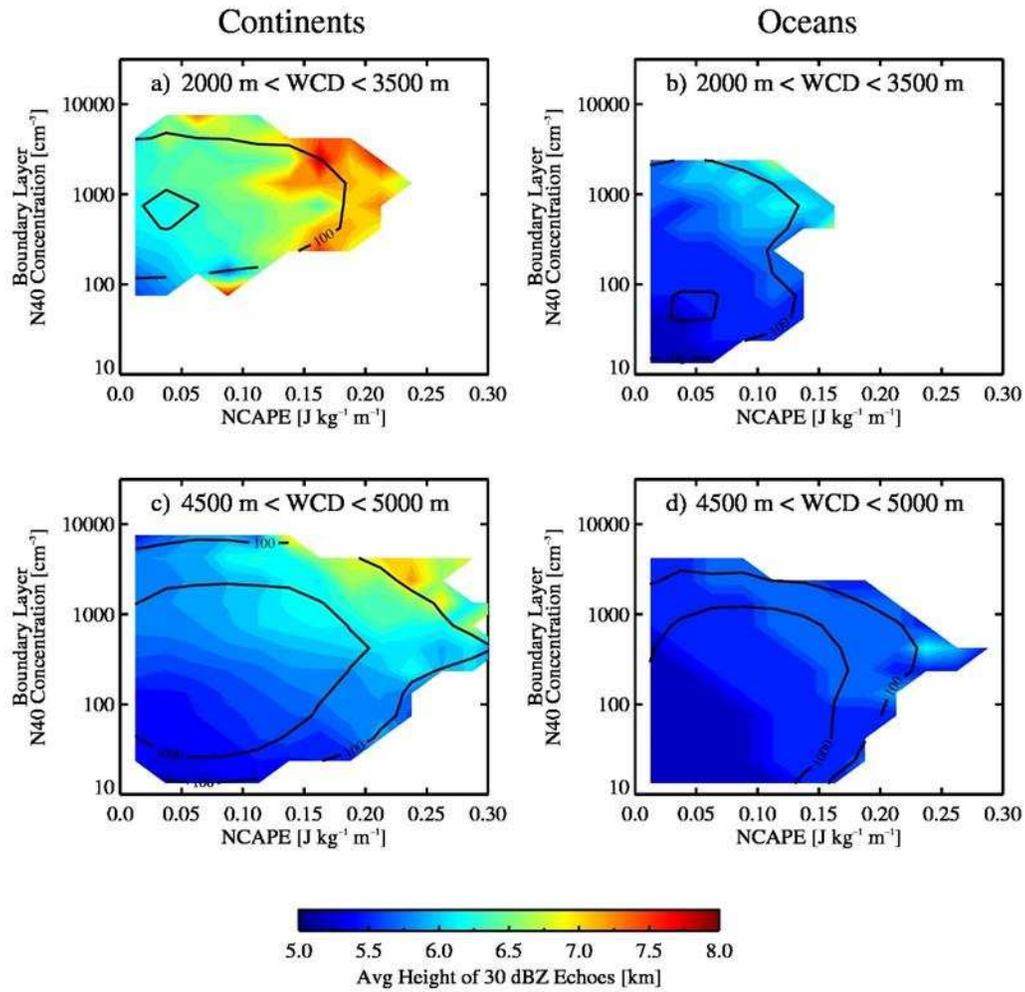
Fig. 2.6. As in Fig. 2.5, but for the average height of 30 dBZ echoes (AVGHT30; shaded). CFs with AVGHT30 > 5 km were analyzed. A threshold of twenty CFs was set for a given bin before the output was plotted.

2.3.3 Partial sensitivity of TLD/AVGHT30 to NCAPE and CCN

To this point, we have focused on the sensitivity of lightning and radar reflectivity to simultaneous changes in NCAPE and CCN. However, it is important to explore how the response of deep convection varies with respect to individual predictor variables with other factors held fixed. Hence we focus on partial sensitivities of TLD/AVGHT30 (e.g., ΔAVGHT30/ΔNCAPE) and how the magnitudes of these terms compare to the observed total response of TLD/AVGHT30 to simultaneous changes in NCAPE and CCN presented in Sec.

2.3.2. The partial sensitivities of proxy variables (i.e., TLD and AVGHT30) with respect to NCAPE have been plotted for fixed CCN above and below the median value of CCN over the global domain, over continents, and over oceans (for shallower and deeper WCD). Likewise, the relationships between TLD and AVGHT30 with respect to CCN are plotted for fixed NCAPE above and below the median value of NCAPE for shallower and deeper WCD for the global domain, over continents, and over oceans (Figs. 2.7-2.8). The partial sensitivity of TLD and AVGHT30 with respect to WCD was also examined; these results will be detailed in Sec. 2.3.4 below.

Table 2.2. Partial sensitivities of AVGHT30 and TLD[b]

Partial Sensitivity	Global	Continents	Oceans
$\left[\dfrac{\Delta AVGHT30}{\Delta CCN}\right]_{NCAPE\,Low}$	(+1.26 km/+0.69 km)	(+1.01 km/+0.65 km)	(+0.75 km/+0.25 km)
$\left[\dfrac{\Delta AVGHT30}{\Delta CCN}\right]_{NCAPE\,High}$	(+1.49 km/+1.13 km)	(+1.11 km/+1.00 km)	(+1.00 km/+1.25 km)
$\left[\dfrac{\Delta AVGHT30}{\Delta NCAPE}\right]_{CCN\,Low}$	(+0.75 km/+0.68 km)	(+0.62 km/+0.47 km)	(+0.50 km/+0.25 km)
$\left[\dfrac{\Delta AVGHT30}{\Delta NCAPE}\right]_{CCN\,High}$	(+1.38 km/+0.76 km)	(+1.28 km/+0.71 km)	(+1.50 km/+0.75 km)
$\left[\dfrac{\Delta TLD}{\Delta CCN}\right]_{NCAPE\,Low}$	(+408%/+199%)	(+211%/+112%)	(+276%/+106%)
$\left[\dfrac{\Delta TLD}{\Delta CCN}\right]_{NCAPE\,High}$	(+421%/+263%)	(+178%/+171%)	(+138%/+150%)
$\left[\dfrac{\Delta TLD}{\Delta NCAPE}\right]_{CCN\,Low}$	(+153%/+131%)	(+61%/+74%)	(+193%/+145%)
$\left[\dfrac{\Delta TLD}{\Delta NCAPE}\right]_{CCN\,High}$	(+53%/+80%)	(+45%/+66%)	(+209%/+141%)

[b]Partial sensitivities of AVGHT30 [km] and TLD [%] with respect to NCAPE and CCN for (shallower/deeper) WCD for CFs/LPCFs over the global domain, over continents, and over oceans. Each value represents the max range of the one-dimensional comparisons in Figs. 2.7-2.8; the first value is the change when WCD is shallower and the second value (see text for stratification methodology).

First, we present the partial sensitivity of TLD as functions of NCAPE and CCN holding other independent variables fixed (Fig. 2.7). Note that there was a minimum threshold of 20 LPCFs required before the bin median was plotted in each case; error bars represent the interquartile range in each bin. The average range of the relationship between TLD and each predictor variable (e.g., NCAPE and CCN) for the global domain, for continents, and for oceans are 104%, 62%, and 172% with respect to NCAPE (Figs. 2.7a-c) and 322%, 168%, and 168% with respect to CCN (Figs. 2.7d-f). The range of the individual relationships described above are given in Table 2.2 for reference. Next, in the legend of each panel, we provide the linear correlation between NCAPE and CCN for each stratification in an effort to address the possibility of the trends being the result of simultaneous correlation. For all stratifications over the global domain, over continents, and over oceans, the linear correlation is generally smaller than 0.20, suggesting that NCAPE and CCN are approximately independent in each case and that variations in the predictand are more likely attributable to the changes in the abscissa.
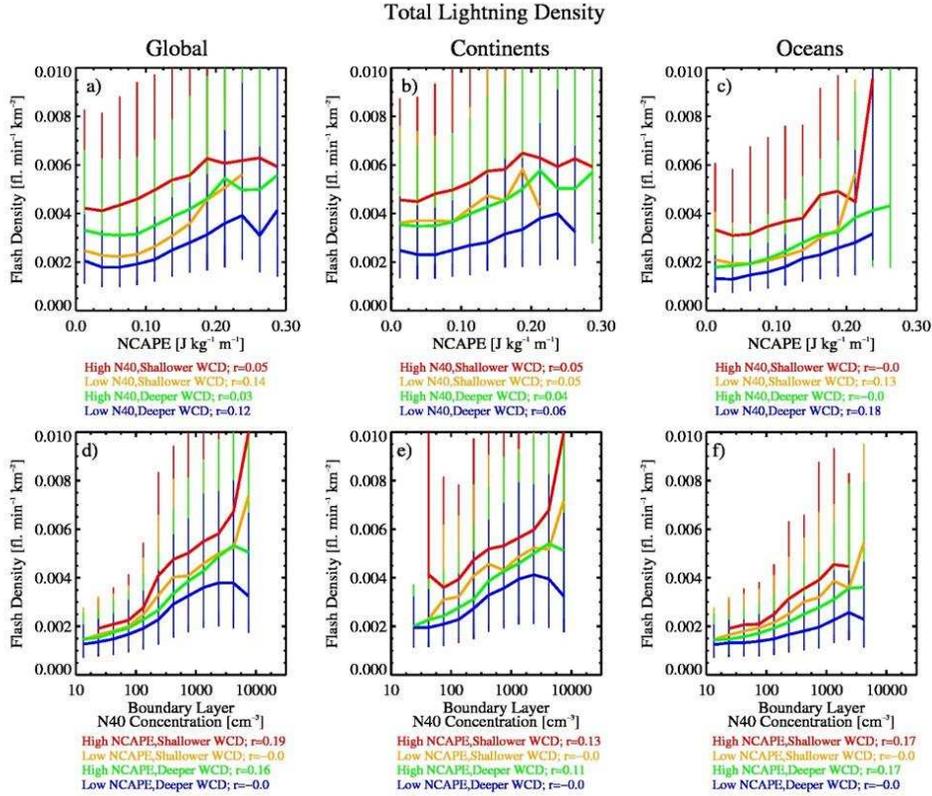
Fig. 2.7. The partial sensitivity of total lightning density (TLD) with respect to a-c) NCAPE and with respect to d-f) CCN for LPCFs over the global TRMM domain (left), over continents (center), and over oceans (right) (*see text for definition of individual stratifications*). Medians of TLD are plotted and the interquartile range in each bin is represented by the error bars. The linear correlation, r, between NCAPE and CCN within each data subset is provided in each panel's legend. LPCFs with flash rates above the minimum detection threshold of the TRMM Lightning Imaging Sensor and AVGHT30 > 5 km were analyzed. A threshold of twenty LPCFs was set for a given bin before the output was plotted.

From Fig. 2.7, it is clear that the average ranges of variability in TLD with respect to CCN over the global domain and over continents are larger when compared to the sensitivity with respect to NCAPE (the behavior of TLD over oceans being the exception). Furthermore, TLD is greater at a given value of NCAPE when CCN is above the median value (true for both shallower and deeper WCD). Additionally, the slope of the relationship between TLD and NCAPE remains roughly constant for the different stratifications of CCN and WCD (Figs. 2.7a-c) while there is a considerable amount of variability in the slope of the relationship between TLD and CCN for different stratifications of NCAPE and WCD (Figs. 2.7d-f). The steepest slope in the latter cases is found for NCAPE above the median value and shallower WCD and

39

the smallest slope is found for low NCAPE and deeper WCD. The slope for the TLD vs. CCN relationship, specifically for low NCAPE and deeper WCD, diminishes and becomes negative beyond CCN > 1000 cm[-3]; this result agrees with the changes in lightning behavior with respect to aerosols that were observed by *Altaratz et al.* [2010] during four dry seasons over the Amazon.

Decreasing TLD with increasing CCN (CCN > 1000 cm[-3]) is consistent with earlier hypotheses put forth in the literature – that the combined effects of a reduction in conditional instability (as a consequence of aerosol absorption/diminished shortwave fluxes at the surface) and increased water loading (owed to reduced coalescence) in ascending parcels reduces updraft intensity for aerosol concentrations beyond some optimum value. The observed behavior of TLD for high CCN (CCN > 1000 cm[-3]) is consistent with a decrease in riming efficiency as cloud droplets become small enough under high aerosol concentrations to yield an overall decrease in the collision efficiency with graupel particles [e.g., R08; *Storer and van den Heever*, 2013].

We also examined the partial sensitivity of AVGHT30 with respect to changes in NCAPE (Figs. 2.8a-c) and CCN (Figs. 2.8d-f) in a similar manner and the results were found to be largely consistent with what was shown for TLD in Fig. 2.7. The average ranges of the AVGHT30 vs. NCAPE relationship (holding CCN and WCD fixed) were 0.89 km, 0.77 km, and 0.75 km for the global domain, continents, and oceans respectively (statistics for individual relationships are listed in Table 2.2). By comparison, the average ranges of the relationship between AVGHT30 and CCN (holding NCAPE and WCD constant) were 1.14 km, 0.94 km, and 0.81 km. Thus, the average ranges in AVGHT30 with respect to CCN are slightly larger than the ranges in AVGHT30 with respect to NCAPE over the global domain and continents, and over

oceans. Again, we find that AVHT30 is higher at a given value of NCAPE when CCN is above

the global median (for fixed WCD; Figs. 2.8a-c) and AVGHT30 levels off or begins decreasing

slightly for values of CCN > 1000 cm^{-3} (e.g., Figs. 2.8d-f). Although the linear correlations

between NCAPE and CCN in each stratification of AVGHT30 (in the legend of each panel in

Fig. 2.8) are slightly larger compared the values for the stratifications of TLD, the r values

remain below about 0.25, again lending to the idea that NCAPE varies approximately

independently of CCN.

For CFs with and without lightning across the global Tropics, the resulting partial

sensitivities of both TLD and AVGHT30 to NCAPE and CCN (for fixed WCD) support the

hypothesis that aerosols modulate the amount of conditional instability realized by ascending

parcels [e.g., R08]. TLD (AVGHT30) was found to be greater (higher) when NCAPE increased

in the presence of higher aerosol concentrations (e.g., Figs. 2.7a-c and 2.8a-c) and the slope of

both the TLD vs. CCN and AVGHT30 vs. CCN relationships decreased to zero and became

negative for CCN > 1000 cm^{-3} (e.g., Figs. 2.7d-f and 2.8d-f). We emphasize that the average

ranges of the TLD/AVGHT30 vs. NCAPE and CCN relationships separately amounted to a

fraction of the total changes that were observed above in Sec. 2.3.2. Therefore, our results

suggest that 1) both NCAPE and CCN are important factors that contribute to the development

of high TLD and AVGHT30 in our data subsets and 2) that there may be other factors that are

simultaneously influencing these attributes of deep convection.

Average Height of 30 dBZ
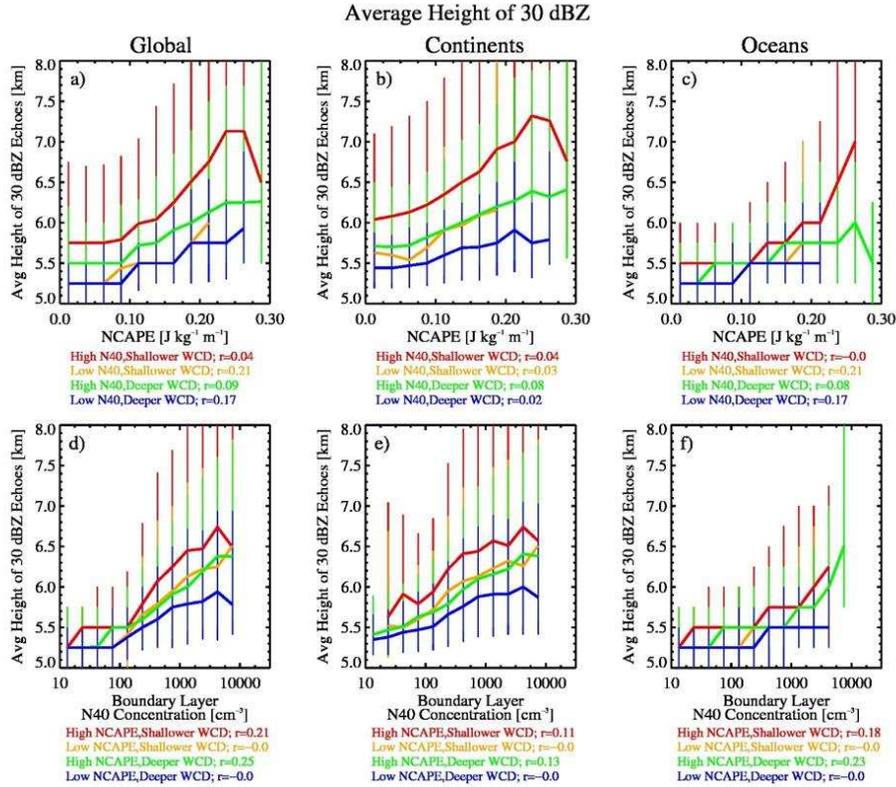
Fig. 2.8. As in Fig. 2.7, but for the partial sensitivity of the average height of 30 dBZ echoes (AVGHT30). CFs with AVGHT30 > 5 km were analyzed. A threshold of twenty CFs was set for a given bin before the output was plotted.

2.3.4 The importance of warm cloud depth

In the above discussion, the sensitivity of lightning and radar reflectivity to simultaneous changes in NCAPE and CCN as well as partial sensitivities of these quantities to NCAPE and CCN holding other independent variables constant were shown to be appreciable. However, our results illustrate considerable variability in both TLD and AVGHT30 with respect to WCD. In Fig. 2.9, we show the difference in TLD and AVGHT30 between shallower WCD (WCD < 4200 m) and deeper WCD (WCD > 4200 m) populations at each point in the NCAPE-CCN parameter space for CFs and LPCFs in the global TRMM domain, over continents, and over oceans, respectively. In each case presented, differences in TLD (percent differences; left panel) and AVGHT30 (height differences; right panel) with respect to WCD are mostly positive – TLD is

greater and AVGHT30 is higher when WCD is shallower as opposed to deeper. We note that there is an exception for AVGHT30 over continents for low CCN across a range of NCAPE; it is likely that this behavior results from seasonal and regional dependence, i.e., CFs occurring in select regions or preferred times of year (to be discussed in Sec. 2.5).

We investigated the significance of the differences between populations of TLD or AVGHT30 between shallower and deeper WCD at each NCAPE/CCN point using a Wilcoxon rank-sum test. In context, the Wilcoxon rank-sum test assumes the null hypothesis that the medians of the populations of TLD or AVGHT30 for shallower and deeper WCD are equal. The maximum increase in TLD (Fig. 2.9; left panel) with respect to WCD at a given NCAPE-CCN point was found to be +91% for the global domain, +86% over continents, and +42% over oceans and these differences were all significant at the P=0.05 level (i.e., these results were in favor of rejecting the null hypothesis at the 5% level). The maximum positive increases for TLD for the global domain and continents (oceans) occurred for NCAPE > 0.25 J kg^{-1} m^{-1} and CCN > 1000 cm^{-3} (NCAPE > 0.1 J kg^{-1} m^{-1} and CCN > 500 cm^{-3}). Negative differences in TLD with respect to changing WCD were very small (less than 1%) and were relatively rare. The maximum increase in AVGHT30 (Fig. 2.9; right panel) with respect to WCD was +1.46 km for the global TRMM domain, +1.19 km over continents, and +1.00 km over oceans and these differences were also significant at the P=0.05 level. The maximum positive increases of AVGHT30 over both the global domain and over continents (over oceans) occur for NCAPE > 0.2 J kg^{-1} m^{-1} and CCN > 1000 cm^{-3} (NCAPE > 0.08 J kg^{-1} m^{-1} and CCN > 1000 cm^{-3}).

These results suggest that deep convection in the Tropics is stronger when WCD is shallower since similar responses are noted for both proxies for convective intensity (i.e., TLD and AVGHT30). Additionally, the differences in TLD and AVGHT30 with respect to

progressively shallower WCD are largest when NCAPE and CCN are high. In other words, in

environments with deeper WCD, higher NCAPE and CCN appear to be necessary to

"invigorate" a convective updraft during ascent through the cloud's warm phase.
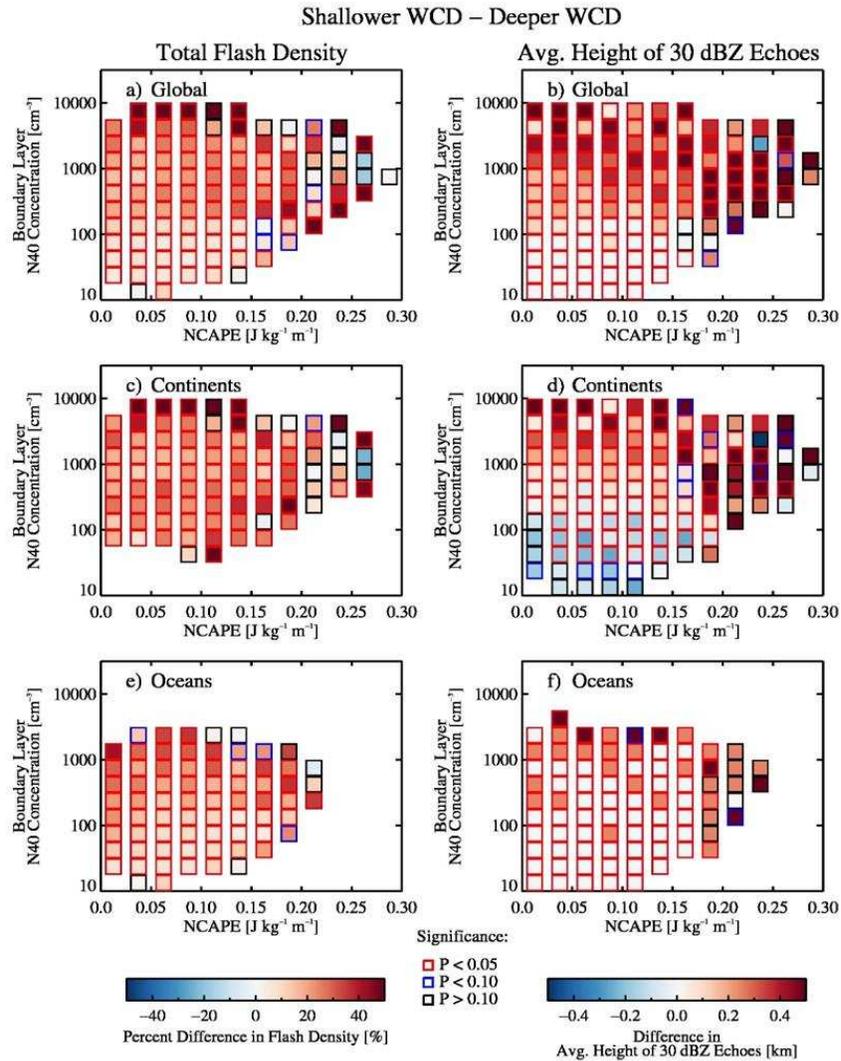
Fig. 2.9. Differences in total lightning density (TLD; left panel) and the average height of 30 dBZ echoes (AVGHT30; right panel) between shallower WCD (WCD < 4200 m) and deeper WCD (WCD > 4200 m) for a,b) the global population of LPCFs and CFs, c,d) LPCFs and CFs over continents, and e,f) LPCFs and CFs over oceans for a fixed NCAPE/CCN point. Shading indicates the percent difference in total lightning density and the difference in AVGHT30 in the left and right panels, respectively. Colored outlines for each point in the two-dimensional parameter space illustrate the significance of the difference determined by a Wilcoxon Rank-Sum test for the difference of medians. CFs (LPCFs) with AVGHT30 > 5 km (with flash rates above the minimum detection threshold of the TRMM Lightning Imaging Sensor) and were analyzed (both constraints in the case of LPCFs).

2.3.5 Radar reflectivity/lightning correspondence and sensitivity

Our results indicate that both TLD and AVGHT30 behave similarly in response to changes in both CCN and NCAPE for different stratifications of WCD. Meanwhile, previous studies have documented strong correspondence between radar reflectivity and lightning in many regions [e.g., *Dye et al.*, 1989; *Rutledge et al.*, 1992; *Petersen et al.*, 1996; *DeMott and Rutledge*, 1998; *Petersen and Rutledge*, 2001; *Liu et al.*, 2012; *Stolz et al.*, 2014]. The climatologies for TLD and AVGHT30 presented above (Fig. 2.1) corroborate the findings of these previous studies, however, our study allows for an attempt to quantify the correspondence between reflectivity and lightning proxies for convective intensity on the global scale. Specifically, we question whether the relationship between TLD and AVGHT30 varies as a function of NCAPE, CCN, and/or WCD.
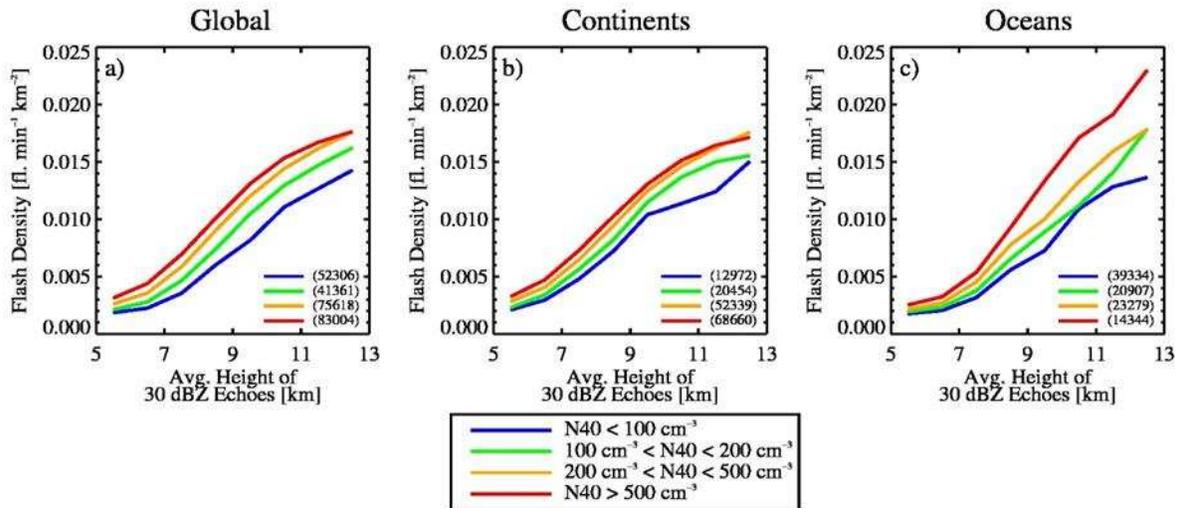
Fig. 2.10. Total lightning density (TLD) versus average height of 30 dBZ echoes for various aerosol stratifications (see legend) for a) the global population of LPCFs, b) LPCFs over continents, and c) LPCFs over oceans. LPCFs with flash rates above the minimum detection threshold of the TRMM Lightning Imaging Sensor and AVGHT30 > 5 km were analyzed. A threshold of twenty LPCFs was set for a given bin before the output was plotted. The number of LPCFs in each respective aerosol stratum is provided in parentheses in the lower right corner of each plot.

Note that this comparison is only possible for deep CFs that produce lightning and therefore we were limited to analyzing only 2.1% of the total CF population observed by TRMM during 2004-2011 (i.e., those features with lightning above the minimum detection threshold of TRMM LIS and AVGHT30 > 5 km). For this subset of CFs, when TLD is binned by AVGHT30 for a continuum of stratifications of CCN (Fig. 2.10) and the resulting bin medians are plotted, a strong positive relationship is readily apparent (r = 0.95-0.98). However, the sensitivity of the TLD versus AVGHT30 relationship decreases with decreasing CCN, suggesting that for convective clouds developing in pristine environments, the 30 dBZ reflectivity column has to be deeper in order to produce the same flash density. Evidence of this phenomenon in the continental subset of LPCFs (Fig. 2.10b) is subtle but is more apparent for the oceanic subset of LPCFs (Fig. 2.10c). Importantly, this finding points to the possibility of different microphysical properties of the mixed phase region in a convective cloud under different background aerosol concentrations. Clouds developing in more polluted environments may be more efficient in separating charge and producing lightning (as evidenced by steeper slopes in the relationship between TLD and AVGHT30).

In accordance with previous observational studies, AVGHT30 was investigated here since it may provide a rough measure of updraft intensity. The assumption is that stronger updrafts are more capable of lofting cloud liquid and frozen hydrometeors. However, AVGHT30 is an ambiguous metric in this context since the same radar reflectivity can be realized for different hydrometeor populations [e.g., *Carey and Rutledge*, 2000]. We next examined VPRR for different stratifications of CCN in various thermodynamic environments to verify if the observed increases in lightning and AVGHT30 with respect to aerosols are associated with a more developed reflectivity column (Fig. 2.11).
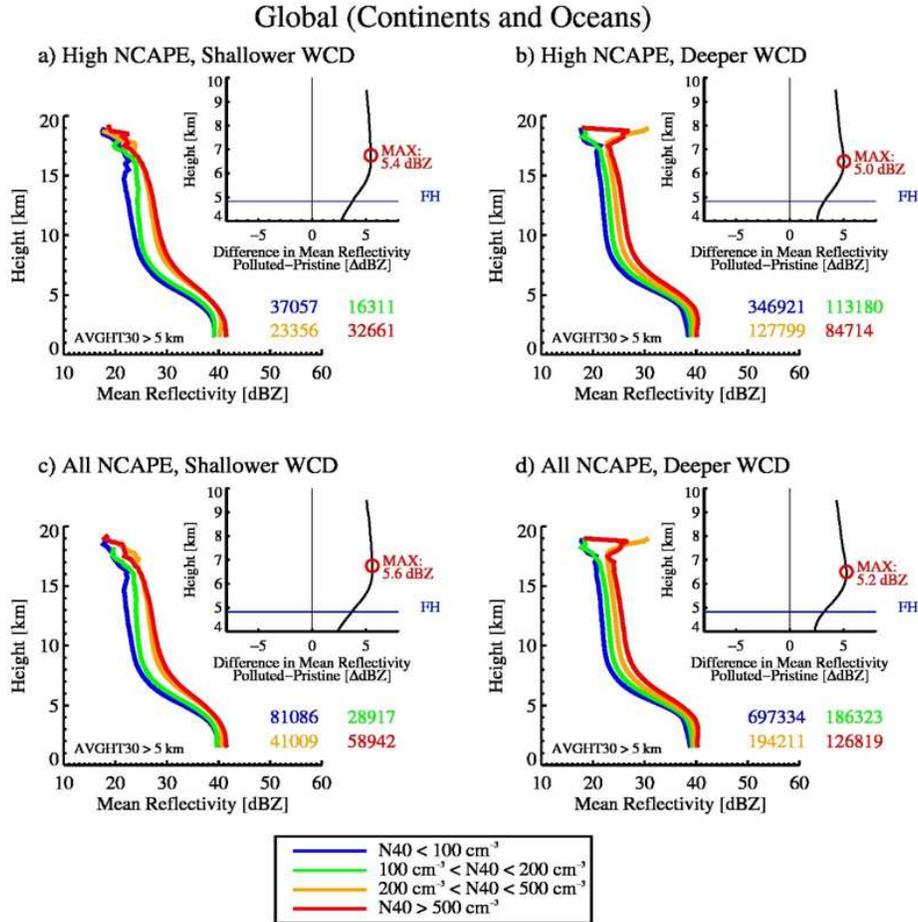
Fig. 2.11. Mean vertical profiles of radar reflectivity (VPRR) for CFs across the global TRMM domain (continents and oceans) for various levels of CCN for a) shallower WCD and b) deeper WCD for NCAPE above the global median value (0.07 J kg^{-1} m^{-1}) and c) shallower WCD and d) deeper WCD for all NCAPEs. Colored values in the lower right of the main plots indicate the number of observations within each aerosol stratum. The difference between the most polluted and pristine VPRR are plotted for altitudes between 4 and 10 km in each inset. The altitude and magnitude of the maximum difference are provided by red open circles and adjacent text. The thin solid blue line in each inset indicates the global average height of the 0^{0}C isotherm (4823 m). CFs with AVGHT30 > 5 km were analyzed.

To investigate the possibility of simultaneous dependence on environmental thermodynamics, we stratified the data by NCAPE above the global median in addition to examining the full range of NCAPEs for both shallower and deeper WCD. Motivation for this part of the analysis came in part as a result of several aspects of the 30 dBZ echo top climatologies discussed in Sec. 2.3.3; MAXHT30 did not change appreciably between continents and oceans (especially true for the global population of LPCFs). By the above method, we

would be able to visualize potential differences in the response of vertical precipitation structure (i.e., the reflectivity column) in deep convective features at various levels of CCN in different thermodynamic environments.

For the global population of CFs (with and without lightning), a clear, systematic increase in reflectivity at a given altitude is shown as CCN increases from pristine levels (CCN < 100 cm^{-3}) to polluted levels (CCN > 500 cm^{-3}) for shallower and deeper WCD in all NCAPE environments. In agreement with the analysis in Secs. 2.3.2 and 2.3.4, the largest changes in VPRR at a given height with respect to CCN occur for shallower WCD (Figs. 2.11a,c) compared to deeper WCD (Figs. 2.11b,d). As shown in each inset, the difference in reflectivity between polluted and pristine profiles is maximized in the mixed phase region (greater for polluted features) and these differences are shown to vary slightly as a function of NCAPE. For all NCAPEs (Figs. 2.11c,d), the difference between polluted and pristine environments ranges from 5.2-5.6 dB in the mixed phase. The observed mean VPRR for high NCAPE (VPRR with NCAPE above the global median value; NCAPE=0.07 J kg^{-1} m^{-1}) illustrate that the magnitude of the difference in reflectivity within the mixed phase between polluted and pristine clouds is still appreciable (5.0-5.4 dB; Figs. 2.11a,b).

The mean VPRR for various aerosol concentrations in different thermodynamic environments therefore show that higher aerosol concentrations are associated with larger values of reflectivity at a given height in the mixed phase. The behavior of the mean VPRR, specifically in the mixed phase region, for different aerosol concentrations are consistent with previous hypotheses regarding aerosol indirect effects on deep convective clouds (e.g., R08), that greater latent heat release via transport of supercooled liquid water and subsequent riming/deposition processes lead to invigorated updrafts capable of lofting graupel.

Additionally, we looked at the median TLD for each of the aerosol stratifications in the VPRR analysis; the median TLD increased monotonically between the pristine and polluted CCN stratifications consistent with the idea that greater flash rates/densities are associated with more vertically developed reflectivity. The results of the VPRR analysis also suggest that the influence of CCN on vertical precipitation structure decreases as NCAPE increases. A discussion of the foregoing results follows in the next section.

2.4 Discussion

A summary of the behavior of deep convective features, associated lightning activity, and their mean VPRR signatures in differing background aerosol and thermodynamic environments (NCAPE, WCD) follows next. Our results suggest that *the highest TLD/AVGHT30 is found for deep convective features that develop in polluted environments where WCD is shallower and NCAPE is above the median value* (Figs. 2.3-2.6). Conditional sensitivities of convective intensity proxies (e.g., TLD and AVGHT30) illustrated that for a fixed WCD, increasing NCAPE resulted in greater TLD/higher AVGHT30 when CCN was above the median value (CCN > 100-300 cm^{-3}) (Figs. 2.7-2.8). Additionally, TLD and AVGHT30 were both nearly systematically larger for shallower WCD compared to deeper WCD (Fig. 2.9). Note that the median value of WCD in our data was found to be very close to 4200 m; therefore, "deeper" WCD probably represents the deepest limit of WCD globally. Thus, the results of this study favor a merged "simultaneous" hypothesis regarding the roles of thermodynamics and aerosols as they may influence the variability of deep convective clouds in the Tropics and Subtropics.

Though a considerable amount of effort went into isolating the impacts of aerosols from background thermodynamics in our investigation of the variability of lightning and radar

characteristics over continents and oceans, it is possible that other sources of potential variability exist within each respective environment (e.g., aerosol species, ice nuclei concentration, wind shear, and relative humidity in the middle troposphere). For example, the data from GEOS-Chem simulations used for this study made no distinction for ice nuclei but ice nuclei sources, concentrations, and activity as a function of temperature are likely to differ between continents and oceans [e.g., *Burrows et al.*, 2013]. In addition, environmental wind shear may impact the results of the current study through its connection to entrainment/mixing events in convective updrafts that could then influence CCN/ice nuclei concentrations and lead to changes in thermal buoyancy within cloudy parcels.

A combination of shallower WCD and more instability over continents implies that the warm-cloud residence time for an ascending parcel could be sufficiently short to allow for efficient transport of cloud water to the mixed phase region to promote electrification and frequent lightning. While observations of very shallow WCD were absent from our tropical dataset, we speculate that large TLD and high AVGHT30 may occur in these environments despite varying aerosol concentrations as the thermodynamics dictate the transport of cloud liquid water to the mixed phase region to allow for riming and charge separation. Additionally, for convective clouds with shallow WCD, the influence of aerosols on droplet growth processes may be limited as a result of the brief period of ascent through the cloud's warm phase. When WCD is > 4000 m (e.g., for many oceanic regions in the Tropics), our findings (Figs. 2.3-2.6) suggest that the activation of warm-rain microphysics and subsequent development of precipitation may still occur, even in the presence of high aerosol concentrations; this is substantiated by our results as the greatest TLD and highest AVGHT30 were found to occur in a relatively small region of the NCAPE-CCN parameter space (i.e., high NCAPE and high CCN).

Hence, it appears that the maximum sensitivity to aerosols in deep convective clouds should be found for environments with sufficient NCAPE for strong convective updrafts and shallower WCD (e.g., WCD < 4000 m) in agreement with previous studies [e.g., *Wang*, 2005; *Fan et al.*, 2007; *Mansell and Ziegler*, 2013].

TLD for a fixed AVGHT30 was found to vary considerably as a function of CCN especially for oceanic regions (Fig. 2.10). In more polluted environments the slope of the relationship between TLD and AVGHT30 is greater and a given lightning density was associated with higher AVGHT30 in more pristine environments. The results suggest that lightning is sensitive to both changes in the internal vertical structure of precipitation and 30 dBZ echo top height. Meanwhile, we have shown that modest enhancements to reflectivity in the mixed phase region occur in progressively polluted boundary layer environments. While the correlation between lightning and radar reflectivity proxies for convective intensity is generally strong, their response to changes in thermodynamics and aerosol concentration may result from different processes.

For example, laboratory experiments suggest that at a constant temperature (i.e., altitude), the amount of charge separated per collision between ice particles increases with increasing supercooled liquid water content [*Takahashi*, 1978]. Thus, it is possible that increases in lightning with respect to increasing aerosol concentrations in a given thermodynamic environment result from enhanced transport of liquid water to the mixed phase region as warm rain processes become less efficient for high aerosol concentrations, without invoking the convective invigoration hypothesis of R08. Meanwhile, increased reflectivities at a given height with increasing aerosol concentrations may again result from inefficient collision/coalescence and enhanced riming in the mixed phase, but larger particles have greater fall speeds. Therefore

a stronger updraft, possibly the result of enhanced latent heat release in the mixed phase when aerosol concentrations are high [e.g., R08], may be necessary to loft these particles to greater altitudes and contribute to larger reflectivity observed there.

In our examination of differences between the mean VPRR within deep CFs (i.e., CFs with AVGHT30 > 5 km) for various stratifications of CCN (Fig. 2.11), we noted distinct similarities in the mean vertical profiles, but systematically larger reflectivity at a given height for progressively higher CCN. *Szoke et al.* [1986] and *Zipser and Lutz* [1994] produced median VPRR for various samples of tropical oceanic, tropical continental, and midlatitude continental convective systems. The mean VPRR for CFs in pristine environments in our study depicted near constant reflectivity below the freezing level (near 40-45 dBZ) and rapid decreases in reflectivity above the freezing level – salient features of the VPRR shown by *Zipser and Lutz* [1994]. *Zipser and Lutz* [1994] showed differences of up to 10-15 dBZ above the freezing level between tropical oceanic and tropical continental convective systems.

We point out that the effect of aerosols is to "continentalize" the vertical precipitation structure by effectively enhancing the reflectivity above the freezing level. While differences in the mean VPRR with respect to aerosols are observed, they are still modest and vary quantitatively as a function of background thermodynamics. We interpret this finding as an indication that thermodynamics simultaneously influence the development of deep convection and are probably necessary to more accurately account for the full differences between tropical continental and tropical oceanic VPRRs and the observed lightning characteristics.

Lastly, it is possible that the response of TLD and AVGHT30 to thermodynamic characteristics of the environment and local concentrations of CCN varies seasonally and regionally. For example, Fig. 2.9d illustrated that AVGHT30 was higher when WCD was deeper

for low CCN and a considerable range of NCAPE for CFs over continents, in effect, countering

the hypotheses concerning the role of WCD put forth by previous researchers [e.g., *Williams and*

Stanfill, 2002; *Williams et al.*, 2005; and *Carey and Buffalo*, 2007]. Even though we exclude

CFs and LPCFs with AVGHT30 < 5 km in the current study to isolate the populations of deep

CFs and LPCFs, we have chosen to incorporate CFs and LPCFs from the global scale TRMM

domain and continents and oceans respectively. Therefore, it is possible that CFs and LPCFs

from different background regimes (e.g., subtropical/midlatitude vs. deep tropical) in different

stages of their respective lifecycles are analyzed together. A pressing question then becomes

whether the results found here are representative over continents and oceans everywhere or are

there notable exceptions over finer time and space scales. Despite the limitation of the smaller

sample sizes, a statistical decomposition of the relative influence of these independent variables

on deep convection and lightning in specific regions of the Tropics and Subtropics was

attempted.

2.5 Statistical evidence for the simultaneous roles of thermodynamics and aerosols

2.5.1 Introductory analysis
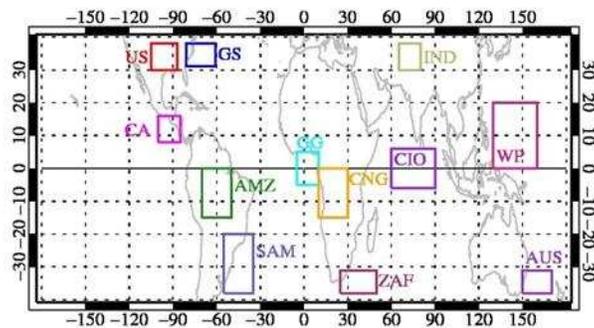
Fig. 2.12. Global map of the TRMM domain (spanning 38°S-38°N, 180°W-180°E). Domains of interest are outlined by colored lines and labeled for identification purposes (reference Table 2.3).

The foregoing discussion focused predominantly on land-ocean contrasts in convective behavior in response to changes in thermodynamics (NCAPE, WCD) and aerosols. A consistent simultaneous dependence of convective intensity proxies on NCAPE, CCN, and WCD was demonstrated for both continents and oceans. Noting the first-order dependence of convective intensity proxies on the chosen predictors that was illustrated in the global analysis, one may question whether the relative contributions of thermodynamics and aerosols to explaining the overall variability in convective intensity changes for environments with different thermodynamic and CCN conditions, i.e., between regions (Fig. 2.12; geographical bounds and abbreviations for each region are provided in Table 2.3). Note that these regions were chosen specifically for study based on documented seasonal changes in environmental characteristics [e.g. Williams et al, 2002] and in order to highlight maxima in global lightning/radar echo top height climatologies [e.g., *Petersen and Rutledge*, 2001; *Christian et al.*, 2003; *Zipser et al.*, 2006; *Liu et al.*, 2012] as well as land-ocean contrasts in convective intensity.

Table 2.3. Regions of Interest[c]

Region	Abbreviation	Geographical Limits
Globe	GLOBE	$[38^0S - 38^0N, 180^0W - 180^0E]$
Continents	CONT	$[38^0S - 38^0N, 180^0W - 180^0E]$
Oceans	OCEAN	$[38^0S - 38^0N, 180^0W - 180^0E]$
Gulf Stream	GS	$[31 - 38^0N, 61 - 81^0W]$
Gulf of Guinea	GG	$[5^0S - 5^0N, 5^0W - 10^0E]$
South America Offshore	SAM	$[20 - 38^0S, 35 - 55^0W]$
South Africa Offshore	ZAF	$[31 - 38^0N, 25 - 50^0E]$
Central America Offshore	CA	$[8 - 16^0N, 85 - 100^0W]$
Australia Offshore	AUS	$[31 - 38^0S, 150 - 170^0W]$
Warm Pool	WP	$[0 - 20^0N, 130 - 160^0E]$
Central Indian Ocean	CIO	$[6^0S - 6^0N, 60 - 90^0E]$
Southern Great Plains, USA	US	$[30 - 38^0N, 87 - 105^0W]$
Amazon	AMZ	$[0 - 15^0S, 50 - 70^0W]$
Congo	CNG	$[0 - 15^0S, 10 - 30^0E]$
India	IND	$[23 - 38^0N, 65 - 75^0W]$

[c]Identification, abbreviation, and geographical boundaries for the selected regions of interest.

The distribution of environmental characteristics for CFs/LPCFs spans a broad area in the three-dimensional parameter space defined by NCAPE (axial X direction), CCN (axial Y

direction), and WCD (axial Z direction), and it is readily apparent that continental and oceanic

LPCFs occupy different portions of the three-dimensional parameter space (Fig. 2.13).

Sensitivity of convective intensity proxies to NCAPE, CCN, and WCD can first be assessed by

searching in the geometric space defined by the three-independent parameters (NCAPE, CCN,

and WCD) for the path from the least to greatest values of convective intensity proxies; once

known, an estimate of the relative importance of each independent parameter may be ascertained

by comparing the magnitude of the projection of the trajectory onto each axis in the three-

dimensional NCAPE-CCN-WCD parameter space. In this section, such an heuristic analysis of

regional dependence of convective intensity proxies on the chosen predictor variables leads to a

more stringent statistical quantification of the relative influence of each variable using a multiple

regression approach following earlier hypotheses (the statistical method will be discussed in

more detail in a subsequent section). As it will be shown, the collection of a large number of

individual observations for many unique combinations of the independent predictors allows for

statistical isolation of the relative contributions of each independent variable studied in this

analysis to the variability of convective intensity.

As a first step to the analysis for individual regions, the convective intensity proxies were

sorted according to their magnitude from least to greatest (according to deciles for TLD and

altitude for AVGHT30). For each stratification, the associated values of the predictors were

indexed, averaged, and then catalogued. By making no assumption about the predictor variable

behavior *a priori*, strictly speaking, the procedure implicitly (incorrectly) treats the predictand as

an independent variable and the predictors as dependent variables; however, this elementary

method is instructive as one can begin to understand potential relationships between convective

intensity proxies and the specified predictors in each region. The trajectories of TLD/AVGHT30

(from least to greatest) were mapped in the three-dimensional parameter space of NCAPE, CCN, and WCD (X, Y, and Z axial directions, respectively; TLD shown in Figs. 2.14-2.16) for each region of interest globally, over continents, and over oceans. For a simplified view, projections of the three-dimensional predictand trajectories onto the NCAPE-WCD (X-Z), CCN-WCD (Y-Z), and NCAPE-CCN (X-Y) planes are also presented.
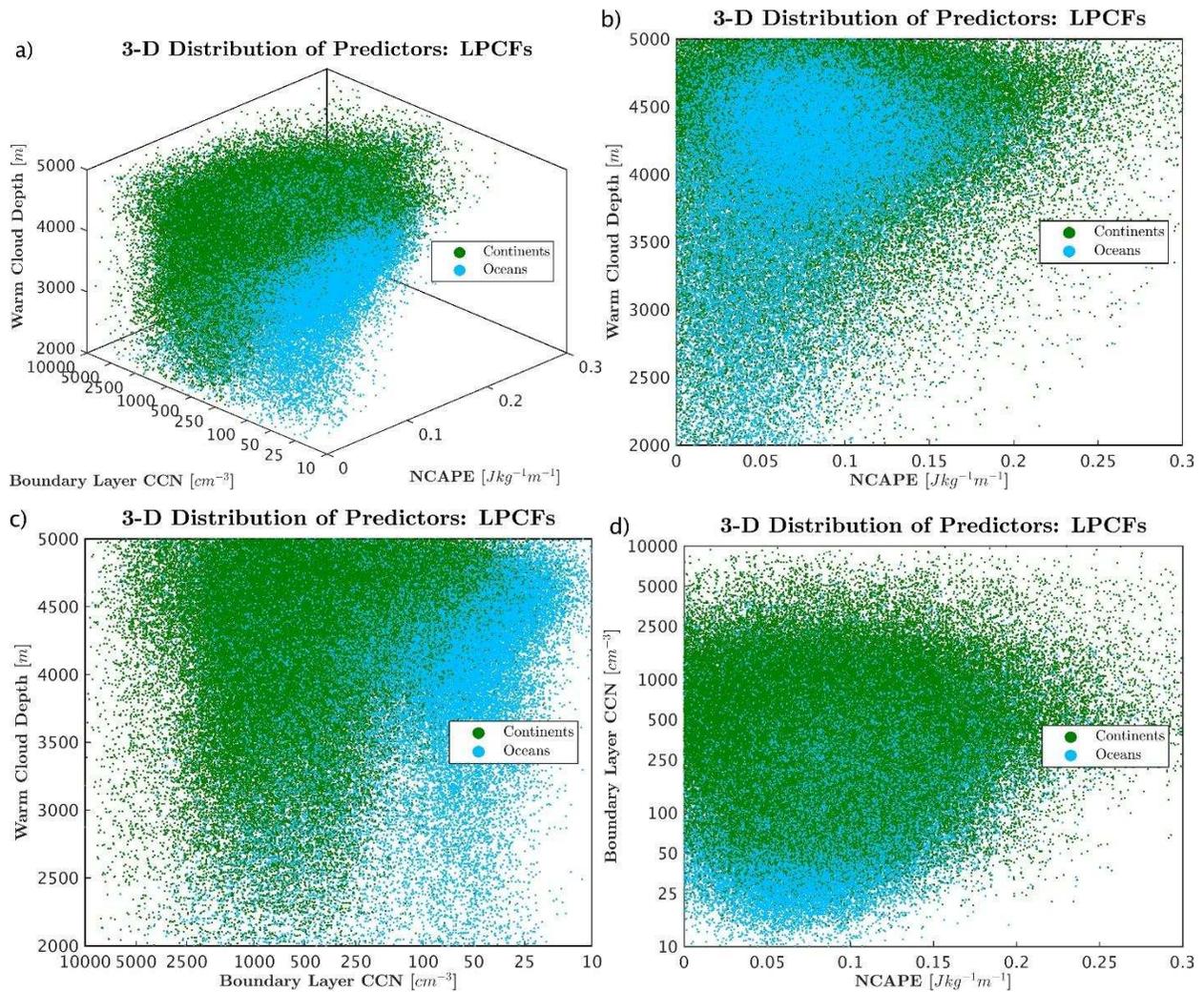
Fig. 2.13. Three-dimensional distribution of independent predictors attributed to LPCFs for the years 2004-2011. Data for LPCFs over continents (green dots) and over oceans (blue dots) have been a) plotted as a function of NCAPE [J kg^{-1} m^{-1}] (x-axis), Boundary Layer CCN concentration [cm^{-3}] (y-axis), and Warm Cloud Depth [m] (z-axis). Projections of the data onto each plane in the three-dimensional parameter space are shown in b), c), and d) (*see text for further description*).
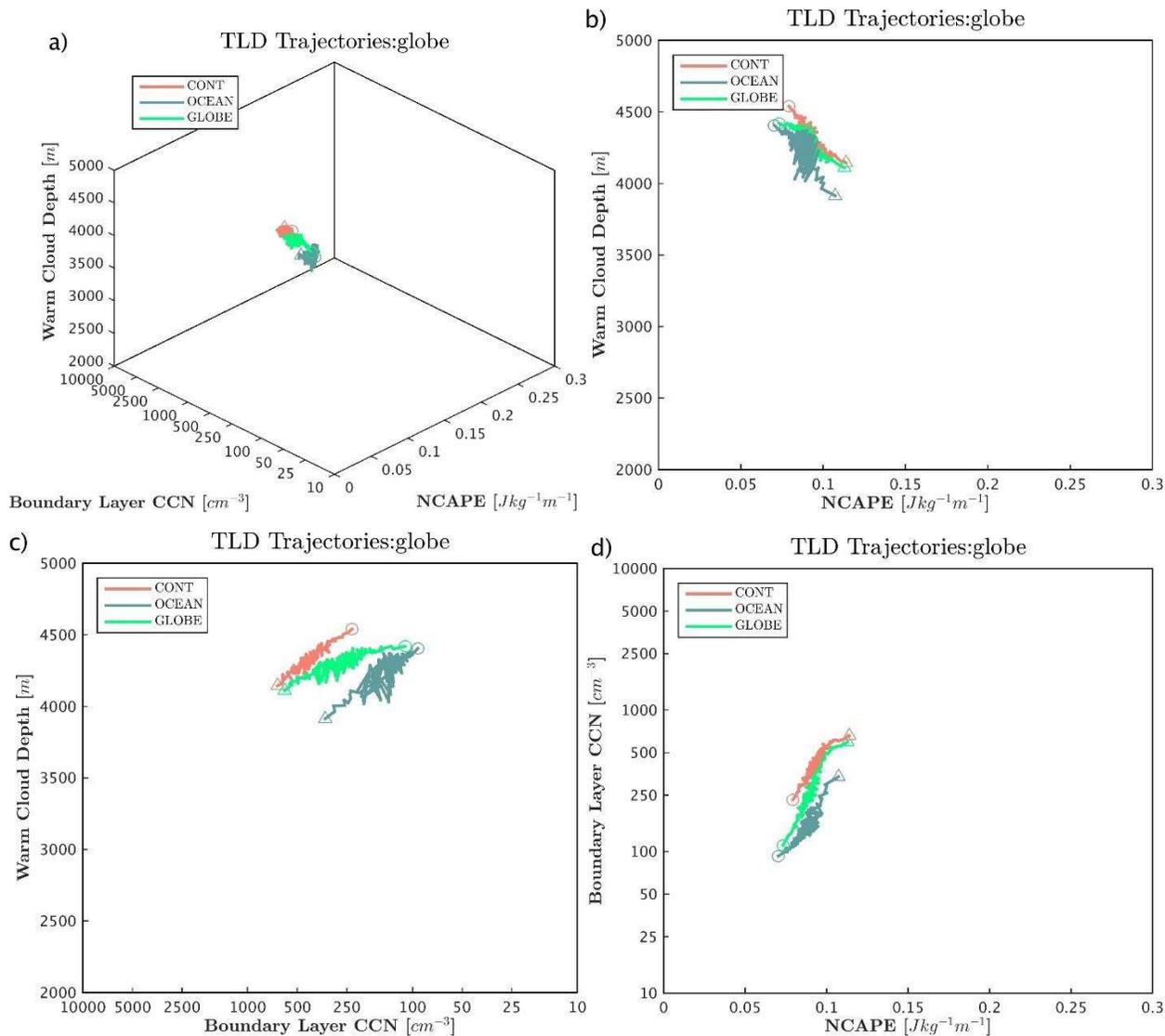
Fig. 2.14. Three-dimensional map of trajectories (solid) for TLD percentiles from least (open circles) to greatest (open triangles) for the globe, continents, and oceans during 2004-2011 (*see legend in each panel for region identification*).
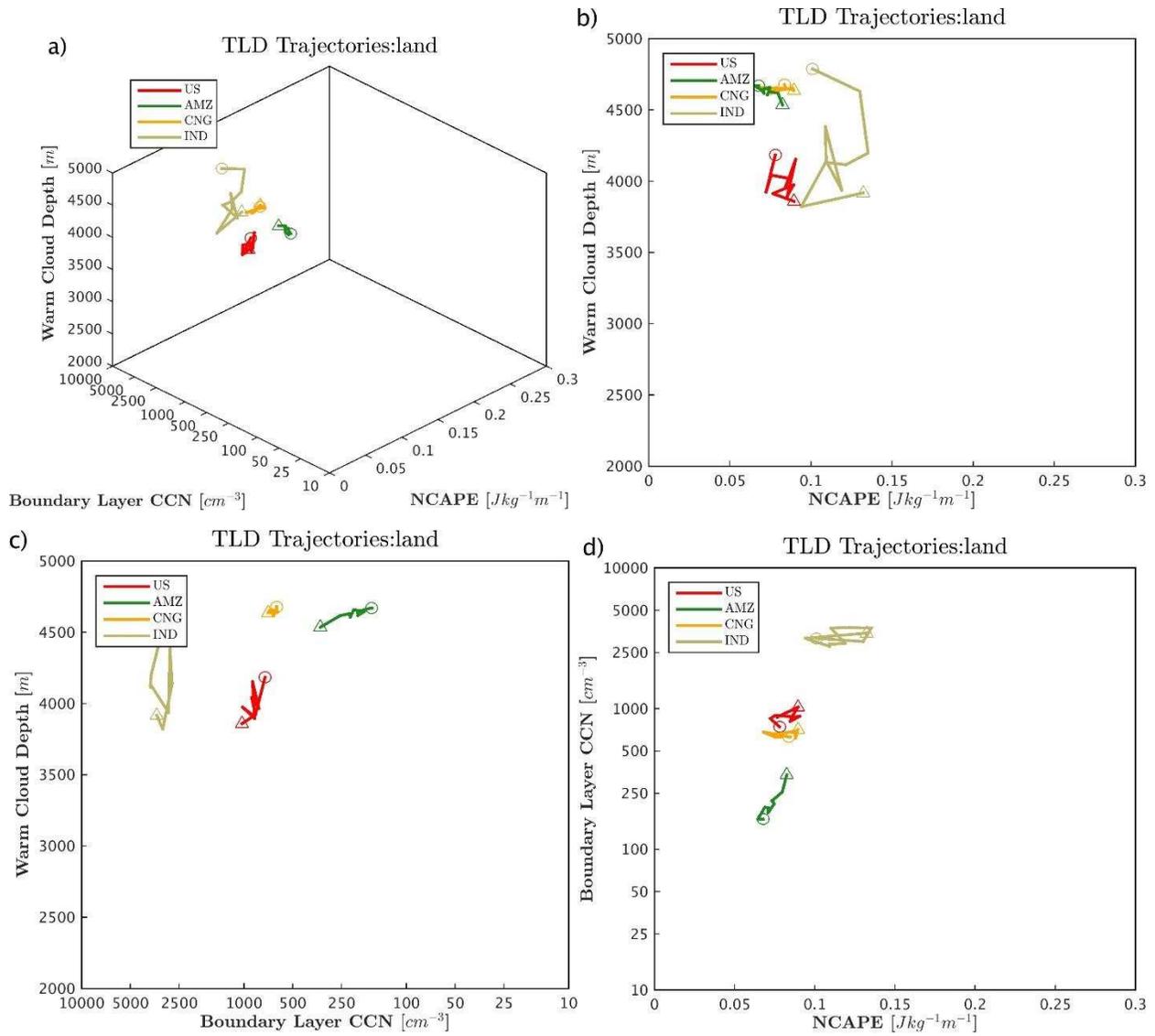
Fig. 2.15. Three-dimensional map of trajectories (solid) for TLD deciles from least (open circles) to greatest (open triangles) for regions of interest over continents during 2004-2011 (*see legend in each panel for region identification*).
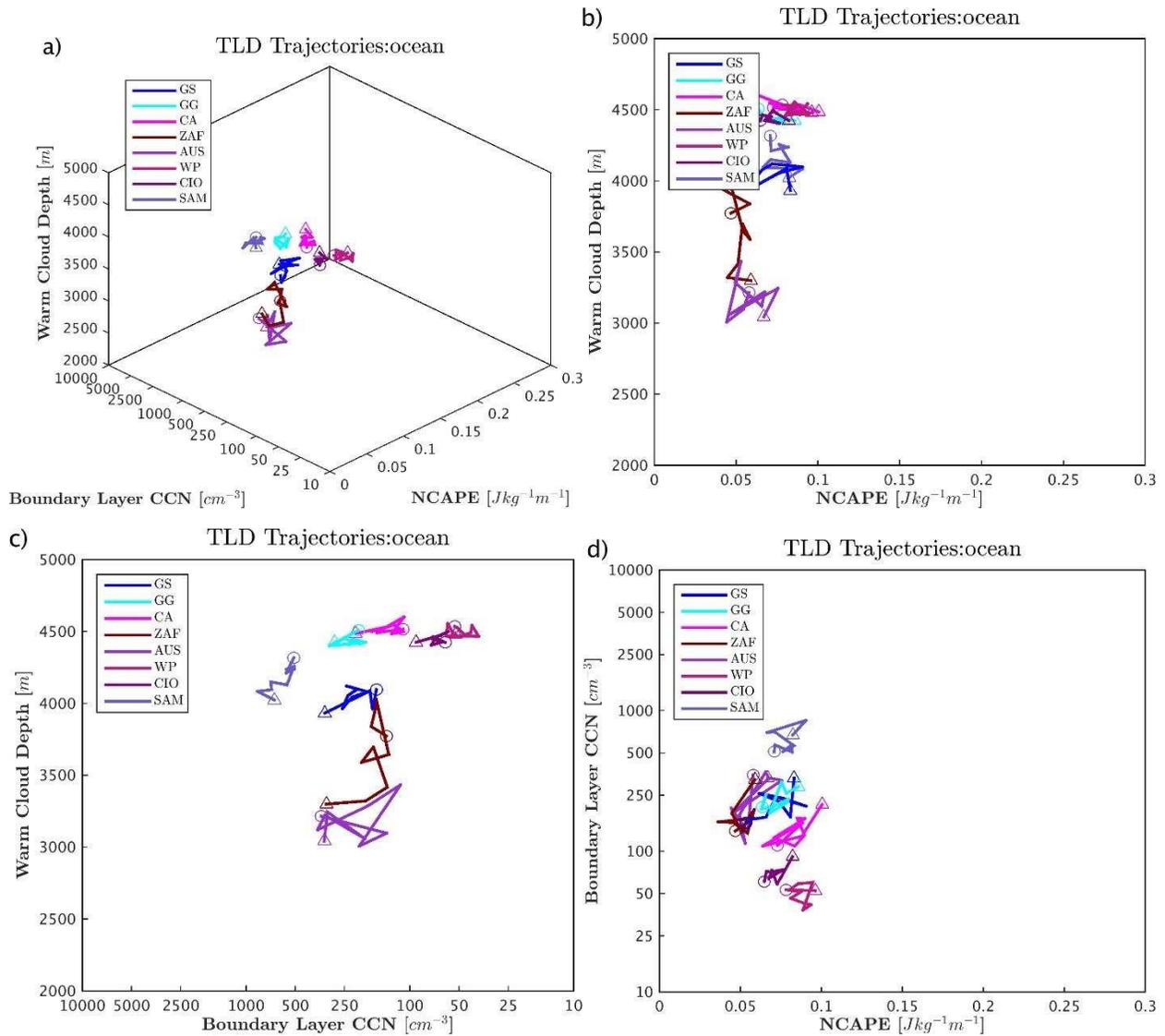
Fig. 2.16. Three-dimensional map of trajectories (solid) for TLD deciles from least (open circles) to greatest (open triangles) for regions of interest over oceans during 2004-2011 (*see legend in each panel for region identification*).

The trend from low to high TLD (AVGHT30 not shown) is marked by a path toward higher NCAPE, higher CCN, and shallower WCD over many of the regions investigated. However, for several regions studied it is clear that the points that make up the trajectory are restricted to certain subspaces within the three-dimensional parameter space, i.e., the Southern Great Plains region (US), the Amazon (AMZ), the regions offshore of South Africa/eastern Australia (ZAF/AUS), and over the Gulf Stream (GS). It is interesting to note that for regions

where CCN concentrations range from 100-500 cm^{-3}, the projections of the trajectories onto the NCAPE and CCN axes appear to be comparable. Meanwhile, for regions with CCN >500 cm^{-3} or <100 cm^{-3}, the trajectories project most strongly on the NCAPE axis and relatively less so on the CCN axis. Thus, the preliminary results suggest the possibility that the largest sensitivity of convective intensity proxies to CCN is found for environments in which the background concentrations are moderate (~100-500 cm^{-3}).

To investigate the linkage to background meteorological conditions, composite behavior of the predictors was examined for low and high values (i.e., below the 10th percentile and above the 90th percentile) of the predictand in a subset of the regions analyzed. For brevity, the results for TLD over regions that stand out as maxima in lightning climatology offshore/downstream of major continents are shown, as a definitive explanation for enhanced lightning in these regions is lacking. During the eight-year period, days with LPCFs that produced lightning densities above the 90th percentile were sequestered from days with LPCFs that produced lightning densities below the 10th percentile (also without producing lightning density above the 90th percentile; percentiles were computed within individual regions). The resulting daily averages of thermodynamic/meteorological variables of interest were composited for these two "exclusive" data populations and compared.

Composites of NCAPE, CCN, WCD, 850 hPa geopotential height (Z), and 1000-850 hPa vector-average flow over three regions downstream of the United States (GS), South Africa (ZAF), and Australia (AUS) during their respective warm convective seasons (JJA for GS and DJF for ZAF and AUS) are presented in Figs 2.17-2.21. For these coastal regions, deeper low-level troughs (evident in the composites of Z850 hPa) promote stronger cross-shore flow and transport higher CCN concentrations farther offshore. Additionally, both convective inhibition

(not shown) and NCAPE were generally higher, contributing to a more volatile thermodynamic profile once parcels reach the level of free convection. WCD composites depict ambiguous results in the comparison between low and high lightning density, possibly indicative of a secondary importance of WCD in the cases presented. Collectively, the composite distributions illustrate that higher lightning density in the warm season over regions downstream of major continents is associated with more favorable thermodynamics and CCN simultaneously when compared to the composites for the lowest decile of lightning density in the same season.
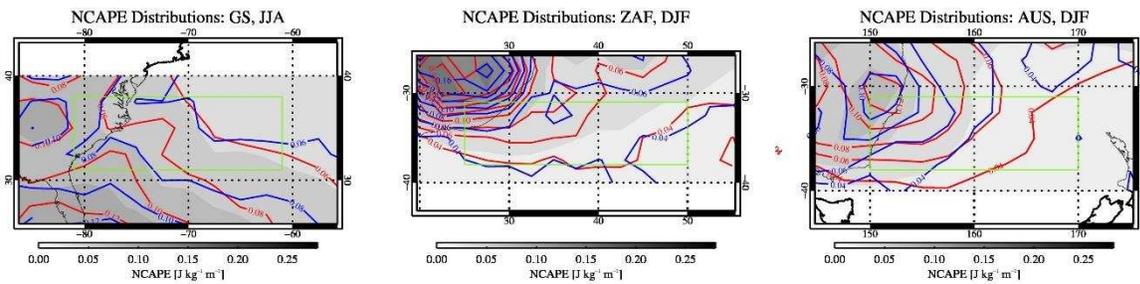
Fig. 2.17. Composites of NCAPE [J kg^{-1} m^{-1}] for low and high TLD over the Gulf Stream (GS; for the months June, July, and August), offshore of South Africa (ZAF; for the months December, January, and February), and offshore of eastern Australia (AUS; for the months December, January, and Feburary). NCAPE is contoured for days featuring the occurrence of TLD above the 90th percentile (red solid) and for days featuring TLD below the 10th percentile (blue solid; not including the occurrence of lightning density in the highest decile). The contour of mean NCAPE (shaded in grayscale) for all days where an LPCF was observed in the domain of interest (light green solid) for each season is shown for reference. Outlines of major continents are shown by thin, black solid lines.
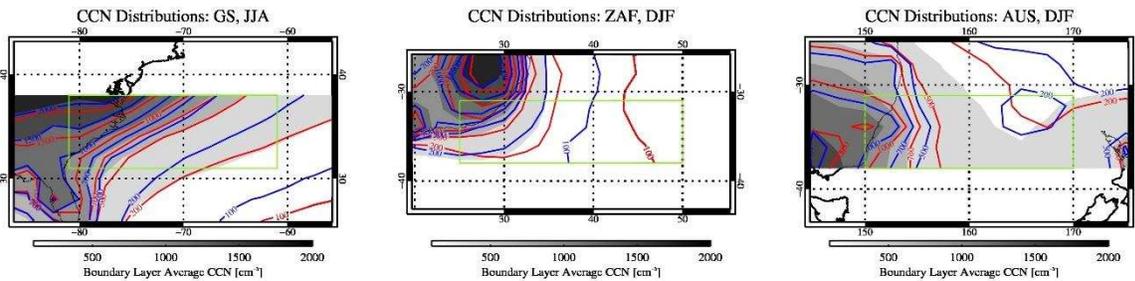
Fig. 2.18. As in Fig. 2.17, but for CCN [cm^{-3}].
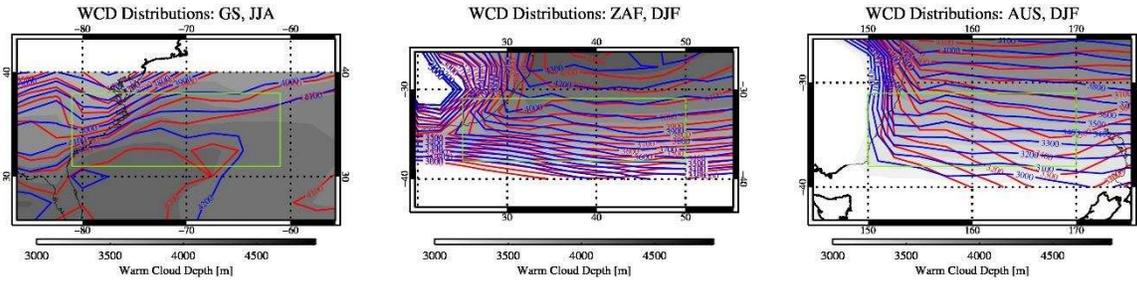
Fig. 2.19. As in Fig. 2.17, but for WCD [m].
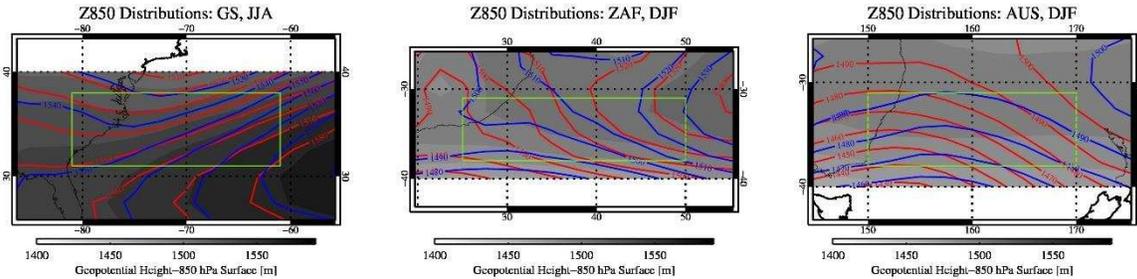
Fig. 2.20. As in Fig. 2.17, but for 850 hPa geopotential height [m].
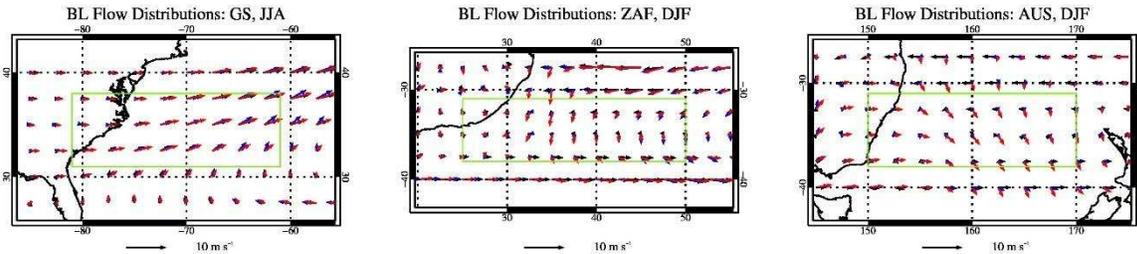
Fig. 2.21. As in Fig. 2.17, but for 1000-850 hPa average wind [m s^{-1}].

The aforementioned results from the initial cursory analysis communicate that in specific regions, the trends in convective characteristics with respect to NCAPE, CCN, and WCD are qualitatively consistent with the results from the global scale investigation. However, as exemplified for oceanic regions downstream of major continents, more unstable environments can occur in unison with stronger aerosol transport and/or greater CCN concentrations and hence major difficulty is encountered when attempting to decouple aerosol microphysical/dynamical impacts from thermodynamic impacts [e.g., *Koren et al.*, 2005,2010; *Altaratz et al.*, 2014]. The

fact that correlations between thermodynamic and aerosol quantities were shown to exist calls

for an approach to isolate the effects of each predictor on convective properties using statistical

methods.

2.5.2 Inferences from simple linear models

Multiple linear regression is a statistical technique that seeks to account for the maximum

portion of variance in a quantity of interest (i.e., the predictand) by minimizing the difference

between the observed value of the predictand and a value of the predictand that is estimated

using a linear combination of the chosen independent variables [*Wilks*, 2011]. Provided a

number of relevant assumptions are valid in context, the sensitivity of the predictand to each of

the independent variables can be quantified in isolation [*Montgomery et al.*, 2012]. As a

generalization of the one-dimensional case, the multiple linear regression model is specified as

follows:

$$Y_n = \beta_0 + \beta_1 X_{1,n} + \beta_2 X_{2,n} + \cdots + \beta_p X_{p,n} + \varepsilon \qquad \text{(Eq. 2.1)}$$

Where Y_n is the predictand, β_0 is the regression constant, β_p are coefficients for each of the X_p

independent predictor variables, and ε are errors between the fitted value of the predictand and Y_n

(for the set of *n* observations). *The statistical integrity of the model is commonly tested against*

the null hypothesis that the chosen set of independent variables exhibit no predictive skill for the

predictand, i.e., $\beta_0 \ldots \beta_p = 0.$

The discussion of uncertainty attendant to the chosen methodology in Sec. 2.2 above

highlights the need to institute an averaging scheme in order to investigate the central behavior

of the data [*Altaratz et al.*, 2014]. In line with the above analysis that analyzed two-dimensional

histograms based on the independent variables, the current approach is to use a three-

dimensional histogram of NCAPE, CCN, and WCD in order to incorporate potential influences

of all predictors simultaneously. The means of these predictor variables within each subset

(cube) of the parameter space are assumed to accurately represent true population values

attributed to individual CFs. Next, the mean of the predictand within each subset of the

parameter space was used to define simultaneous observations accordingly (e.g., Y_n, $X_{1,n}$, $X_{2,n}$,

$X_{3,n}$, ..., $X_{p,n}$, for p predictors for each of the n observations).

The explicit goals of conducting this type of statistical regression analysis are generally

1) to investigate whether the set of independent variables offer predictive capability for the

dependent variable of interest, 2) whether linear combinations of the independent variables

"explain" more of the variance in the dependent variable compared to each of the independent

variables in isolation and 3) to determine the relative contributions of each independent variable

to the variance in the dependent variable, controlling for other inputs. According to the latter

goal, standard anomalies, i.e., the quantity defined by the difference between a value and its

sample mean, divided by its sample standard deviation, were computed for each of the variables

of interest. Thus, regression of standard anomalies of the predictand onto standard anomalies of

the independent variables permits direct comparisons of the relative importance of each input.

An initial form of the multiple linear regression model follows from the above result (i.e., Sec.

2.4), that stronger convection occurs in environments with higher NCAPE, higher CCN

concentrations, and shallower WCD. The results from the global CF study above suggest

possible first-order linear dependence of convective intensity proxies on NCAPE, CCN and

WCD. However, several previous studies have noted potential for higher order dependencies of

the predictand on the chosen set of independent variables.

To reiterate, R08 suggest a "saturation/optimum" effect for aerosols, such that initial

increases in aerosol concentrations can invigorate convection, yet increasing the aerosol load

beyond some optimum value leads to a reduction in short-wave fluxes reaching the surface and subsequent depletion of surface-based instability (contributing a simultaneous direct radiative influence on thermodynamics). Alternatively, increasing the concentration of CCN in parcels near cloud base leads to progressively smaller droplets, as an increased number of condensation centers compete for the available water vapor; in this view, droplets may become so small that the collision/riming efficiency for small droplets and riming particles diminishes appreciably. Subsequently, the development of high concentrations of graupel above the freezing level and the occurrence of rebounding collisions between precipitation ice and smaller ice particles becomes less likely, in which case lightning and radar reflectivity in the mixed-phase region may decrease in frequency and magnitude, respectively. Lastly, it is possible that there could be a simultaneous quadratic/logarithmic dependence on WCD as this parameter influences the warm-phase residence time of ascending parcels. For progressively shallower WCD, the time for CCN or condensational growth processes to impact warm-rain microphysics decreases. So WCD plays the same forcing role as increasing CCN. When considering simultaneous influences of other predictors, is there a limit beyond which shallowing the WCD leads to no further changes in the response?

It is often the case that environmental variables are log-normally distributed and therefore these variables may be transformed into logarithmic form in practice [e.g., *Stolz et al.*, 2014]. To rectify log-normal behavior of the predictand or individual predictors in the current statistical investigation, multiplicative relationships are also relevant. For example, multiplicative models of the dependent variable, e.g., lightning, can be re-expressed as linear models via the following development:

$$Y = \beta_0 \left[\frac{X_1^{\beta_1} \cdot X_2^{\beta_2}}{X_3^{\beta_3}} \right] \qquad \text{(Eq. 2.2a)}$$

$$log_{10}(Y) = log_{10}\left(\beta_0\left[\frac{X_1^{\beta_1} \cdot X_2^{\beta_2}}{X_3^{\beta_3}}\right]\right) \quad \text{(Eq. 2.2b)}$$

$$log_{10}(Y) = log_{10}(\beta_0) + \beta_1 \cdot log_{10}(X_1) + \beta_2 \cdot log_{10}(X_2) - \beta_3 \cdot log_{10}(X_3) \quad \text{(Eq. 2.2c)}$$

Note that the preceding development is relevant in a multiple linear regression framework since the hypothesized model form remains linear in the regression parameters, β_p (i.e., the model form does not contain products of regression coefficients). Table 2.4 describes the experimental regression trials used in this study. The linearized form of the above multiplicative model will be referred to as the "logarithmic" form hereafter so that it may be differentiated from pure linear representations of Y and the predictor set alike.

Table 2.4. Linear models of varying complexity[d]

Model	Description	General Form
1	Single X_1	$Y \sim X_1$ ($log_{10}Y \sim log_{10}X_1$)
2	Single X_2	$Y \sim X_2$ ($log_{10}Y \sim log_{10}X_2$)
3	Single X_3	$Y \sim X_3$ ($log_{10}Y \sim log_{10}X_3$)
4	Linear	$(log_{10}Y)$ $Y \sim X_p$
5	Interaction	$(log_{10}Y)$ $Y \sim X_p \cdot X_q$
6	Quadratic	$(log_{10}Y)$ $Y \sim X_p^{\,2}$
7	Logarithmic	$(log_{10}Y)$ $Y \sim log_{10}X_p$
8	Linear, Interaction	$(log_{10}Y)$ $Y \sim X_p, X_p \cdot X_q$
9	Linear, Quadratic	$(log_{10}Y)$ $Y \sim X_p, X_p^{\,2}$
10	Quadratic, Interaction	$(log_{10}Y)$ $Y \sim X_p^{\,2}, X_p \cdot X_q$
11	Linear, Interaction, Quadratic	$(log_{10}Y)$ $Y \sim X_p, X_p \cdot X_q, X_p^{\,2}$
12	Linear, Interaction, Quadratic, Logarithmic	$(log_{10}Y)$ $Y \sim X_p, X_p \cdot X_q, X_p^{\,2}, log_{10}X_p$
13	Optimal	$(log_{10}Y)$ $Y \sim X_p, X_p^{\,2}, log_{10}X_p$ (three-term model)

[d]Models for the response variable, Y, versus linear representations of NCAPE (X_1), CCN (X_2), and WCD (X_3) employed in the statistical decomposition. Model #13 represents an experimental form that employs three independent variable terms of any transformation (linear, quadratic, and logarithmic allowed; one each for X_p), seeking the maximum strength of regression.

In the multiple regression analysis, the investigator is burdened with proving that the null hypothesis (i.e., $\beta_0...\beta_p = 0$) is false and thus is obliged to demonstrate that assumptions of Gaussian residuals, homoscedasticity (constant residual variance), and linear independence for the set of predictors are valid. Therefore, the multiple correlation for the regression, the strength/significance of the multiple regression, covariability/multicollinearity of the independent variables, distribution of regression residuals/residual variance, and standard error

of each β_p (to determine whether the individual parameter estimates are significantly different from zero) were examined in detail.

The linear correlation between a predictor and the rest of the independent variables in each model form, i.e., multicollinearity, was computed in order to determine the degree of linear dependence in the set of independent variables. From a statistical perspective, the ordinary least-squares matrix algebra solution for parameter estimates in the multiple linear regression framework depends on computation of the inverse of the product of the design matrix, X (i.e., the [n x p] matrix whose columns are vectors of length n, corresponding to n measurements of each predictor, p), multiplied by its transpose, i.e., $[X'X]^{-1}$. If one column (or more) of the design matrix can be rewritten as a linear combination of the remaining columns, the square matrix $[X'X]$ is said to be singular and therefore not invertible. Conceptually, if there is significant multicollinearity in the predictor set, two or more independent variables are highly correlated with each other. If there is then significant correlation between the set of predictors and the *response variable*, it becomes difficult (or impossible) to assess the relative influence of each predictor on the response variable in isolation from other independent variables.

The predictive strength and significance of each model was assessed using multiple-correlation coefficients both for the regression and for the regression adjusted according to the number of degrees of freedom. The F-statistic, defined by the ratio of variance of the regression to the variance of the residuals (assuming Gaussian residual behavior), was used to test the null hypothesis that at least one of the regression parameter estimates was significantly different from zero. Standard-error estimates for each of the regression parameters were used to assess regression coefficient stability via the appropriate t-statistic (i.e., the ratio of the parameter's coefficient estimate to its sample standard deviation), again assuming Gaussian residuals. These

quantities were computed for specific regions and seasons; the focus now shifts to a discussion

of results from the multiple regression analysis and assessment of the statistical validity of these

findings.

The extended set of results for each regression model (e.g., models 1-13 in Table 2.4), for

each region/season are provided in Appendix A.2, but several aspects of the collection of

experimental regression trials for all regions and seasons studied may be generalized (though

regions lacking sufficient data points were omitted). Importantly, the multiple linear model

including either linear or logarithmic representations of the convective intensity proxies and

predictors (i.e., NCAPE, CCN, and WCD) for almost every region analyzed illustrated that the

hypothesized first-order dependence was consistent with findings from the global scale

(reference Sec. 2.3.2 above; to be discussed in more detail below). For example, the multiple

correlation coefficients (R^2), an indicator of the model's overall predictive skill, for models

involving combinations of the predictors was larger than the R^2 for one-dimensional

linear/logarithmic relationships between TLD/AVGHT30 and each predictor separately. The

degree of linear dependence in the predictor set, i.e., multicollinearity, was small for

linear/logarithmic combinations of the predictors; approximately twenty percent or less of the

variability in one predictor was explained by linear combinations of other predictors in the set for

most regions annually (according to *Draper and Smith* [1998], these values are within standard

tolerance levels for multicollinearity in multiple linear regression).

The F-statistic for each model was statistically significant ($p < 0.01$), favoring rejection

of the null hypothesis, such that at least one of the $\beta_p \neq 0$ (an indication that the ratio of variance

of the regression to the variance of the residuals is large; possibly a consequence of the large

sample sizes involved). For many of the experimental trials involving simple linear/logarithmic

representations of the predictors, individual regression parameter estimates, i.e., the b_p, were found to be significant at the p < 0.05 level (two-tailed probability). Stated another way, statistical significance is implied by a 95% chance that the absolute value of a random t-variate is less than or equal to the t-statistic for the parameter estimate in question. For higher-order models that included multiple representations of NCAPE, CCN, and WCD in the predictor set, the multiple correlation was in some instances higher than that for lower-order models, but multicollinearity in the predictor set for higher-order models increased sufficiently to render individual parameter estimates generally insignificant from a statistical perspective. Note that the R^2 values for linear/logarithmic representations were comparable or only slightly smaller than the R^2 for higher-order models considered. Hence, the rest of the discussion will focus on linear/logarithmic representations, i.e., experimental model forms 4 and 7 from Table 2.4.

The above discussion of general results for individual regions contributes 1) that simple linear/logarithmic models are consistent with first-order dependence found on the global scale; 2) that these model forms capture a significant portion of the variance of the response; and 3) that the parameter estimates for each model form are unlikely to result from mere random chance. Linear and logarithmic model results will be studied next to assess the validity of the assumptions of the multiple linear regression framework; it is important to investigate regression residual behavior to determine whether parameter estimates remain unbiased [*Wilks*, 2011], and therefore applicable across the domain of the independent parameter space. Figs. 2.22-2.26 exemplify the distribution of the residuals with respect to each predicted value from the linear model (left panel) and logarithmic model (right panel) for TLD (AVGHT30 not shown) over the globe, continents, oceans, and different regions of interest for all months in the eight-year sample. The residual variance for linear models in the left panel suggests Gaussian residual

behavior (randomly distributed about a mean of 0 for a given predicted Y value), but heteroscedastic behavior is apparent, as the variance increases across the domain of predicted Y. Residuals for the logarithmic model form in the right panel are approximately Gaussian and the variance of the residuals is more uniform across the domain of predicted Y, especially over oceans, the Amazon (AMZ), and the Gulf Stream (GS) domains, for example.

The latter behavior evident for the logarithmic model form is more in line with the assumptions of the multiple linear regression framework. Although the logarithmic representation of TLD (and AVGHT30) over continents appears to be subject to "missing predictors" [e.g., *Wilks*, 2011; pp. 227] the mean-square-error (MSE) statistics, shown in the upper left of each plot, suggest that the logarithmic model is a better fit to the data as evidenced by smaller MSE values overall. Using Tables 2.5-2.6 as an additional reference for regression statistics, it is understood that the logarithmic model form consistently outperforms the linear model form as shown by the multiple correlation coefficient.
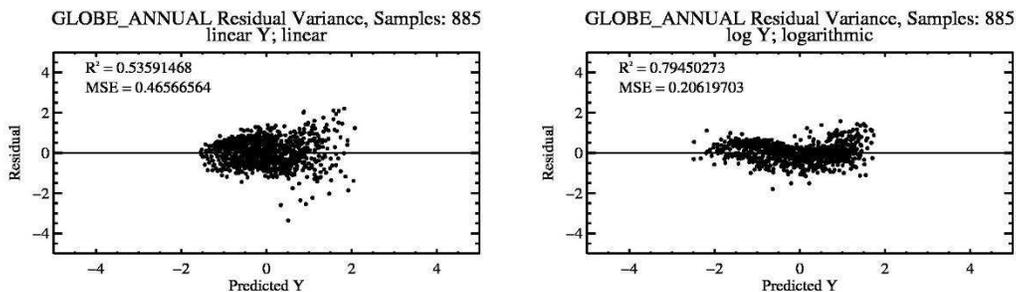
Fig. 2.22. The distribution of regression residuals (i.e., the difference between the predicted value of dependent variable, Y, and the observed value of Y) versus the predicted value of Y for the Global dataset for all months for the years 2004-2011. The number of samples, form of Y, and general form of the regression model parameters are provided in the title; linear models (left) are compared to logarithmic models (right). Units on each axis are representative of standardized anomalies. The multiple correlation (R^2) and mean square error (MSE) for each regression are provided in the upper left corner of each plot.
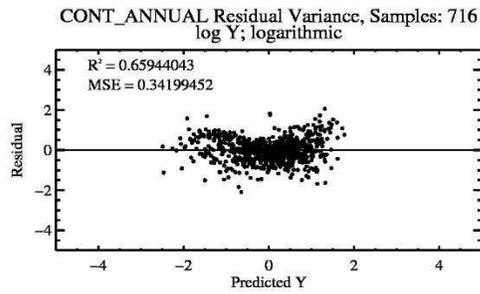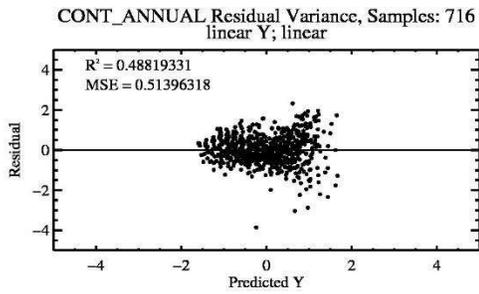
Fig. 2.23. As in Fig. 2.22, but for continents (CONT).
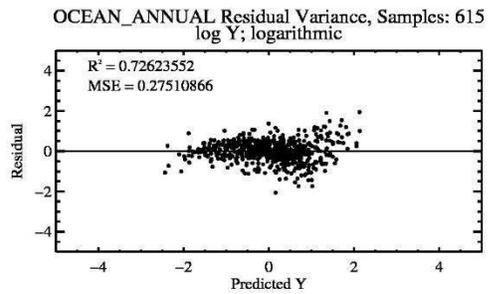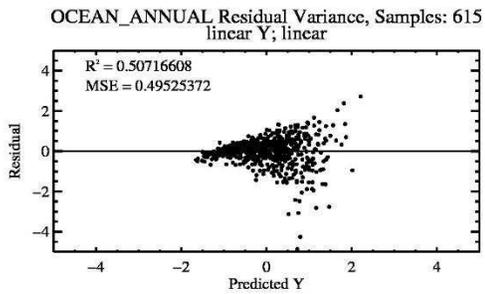
Fig. 2.24. As in Fig. 2.22, but for oceans (OCEAN).
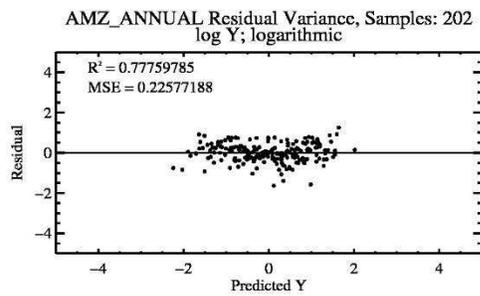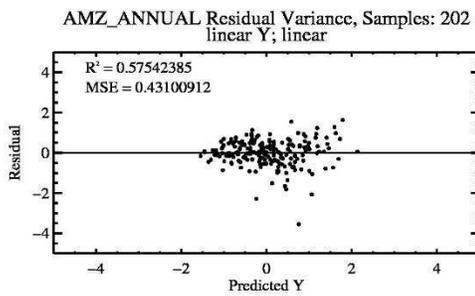
Fig. 2.25. As in Fig. 2.22, but for the Amazon (AMZ).
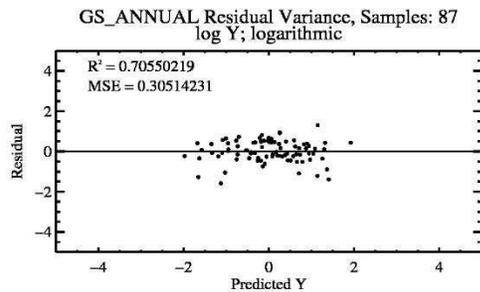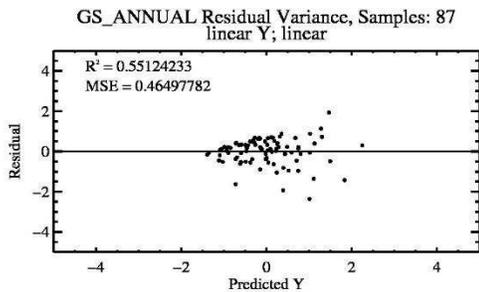
Fig. 2.26. As in Fig. 2.22, but for the Gulf Stream (GS).

Over the globe, continents, oceans, and select smaller-scale regions of interest, the logarithmic model exhibits higher multiple correlation with the convective intensity proxies (R^2~0.36-0.79 for TLD; R^2~0.27-0.73 for AVGHT30) compared to the linear model (R^2~0.04-0.61 for TLD; R^2~0.29-0.61 for AVGHT30). The improvement for logarithmic models of AVGHT30 vs. linear models of AVGHT30 based on the multiple correlation coefficients are generally small (in some cases, the linear model outperforms the logarithmic model). Total regression (via the F-statistic) and individual parameter estimates (via the t-statistic) were statistically significant at the $p < 0.05$ level over the GS, CIO, US, and AMZ for TLD and over the GS, SAM, US, AMZ, and CNG for AVGHT30 annually. Provided that the aforementioned discussion validates the logarithmic model (using NCAPE, CCN, and WCD as inputs) for convective intensity, a critical final consideration in the regression analysis is the comparison of the magnitudes of the weights applied to each predictor.

Table 2.5. Statistics of regression model forms 4 (linear) and 7 (logarithmic) for TLD[e]

Region	Season	Multicollinearity	Multiple R^2	Individual R^2	Estimates, β_p	Significance, β_p
GLOBE	Annual	0.12, 0.00, 0.13	0.54 (0.79)	0.18,0.25,0.03	0.53,0.49,-0.37	1.00,1.00,1.00
(N=885)		(0.12,0.03,0.10)		(0.13,0.53, 0.02)	(0.32,0.75,-0.31)	(1.00,1.00,1.00)
CONT	Annual	0.10,0.00,0.10	0.49 (0.66)	0.12,0.15,0.13	0.51,0.36,-0.49	1.00,1.00,1.00
(N=716)		(0.09,0.07,0.14)		(0.08,0.45,0.19)	(0.38,0.56,-0.41)	(1.00,1.00,1.00)
OCEAN	Annual	0.14,0.00,0.14	0.51 (0.73)	0.09,0.18,0.13	0.49,0.41,-0.53	1.00,1.00,1.00
(N=615)		(0.12,0.01,0.12)		(0.06,0.39,0.23)	(0.40,0.56,-0.58)	(1.00,1.00,1.00)
GS	Annual	0.12,0.01,0.12	0.55 (0.71)	0.17,0.38,0.01	0.44,0.58,-0.24	0.99,0.99,0.96
(N=87)		(0.12,0.00,0.12)		(0.19,0.50,0.01)	(0.48,0.67,-0.24)	(0.99,1.00,0.96)
GG	Annual	0.09,0.17,0.09	0.61 (0.61)	0.45,0.32,0.00	0.54,0.45,-0.14	1.00,1.00,0.70
(N=58)		(0.10,0.18,0.10)		(0.39,0.35,0.00)	(0.48,0.51,-0.21)	(1.00,1.00,0.87)
SAM	Annual	0.23,0.01,0.23	0.34 (0.43)	0.04,0.30,0.03	0.11,0.54,0.14	0.39,0.99,0.48
(N=29)		(0.31,0.01,0.31)		(0.00,0.41,0.01)	(0.11,0.65,-0.18)	(0.38,1.00,0.57)
ZAF	Annual	0.04,0.52,0.55	0.50 (0.72)	0.07,0.48,0.26	0.14,0.63,-0.05	0.47,0.94,0.13
(N=22)		(0.06,0.60,0.59)		(0.05,0.69,0.56)	(0.02,0.61,-0.27)	(0.09,0.91,0.56)
CA	Annual	0.12,0.14,0.03	0.47 (0.67)	0.43,0.16,0.00	0.59,0.20,0.00	1.00,0.84,0.02
(N=58)		(0.20,0.24,0.06)		(0.61,0.26,0.03)	(0.71,0.17,-0.16)	(1.00,0.73,0.76)
AUS	Annual	NaN (NaN)	NaN (NaN)	NaN (NaN)	NaN (NaN)	NaN (NaN)
(N=NaN)						
WP	Annual	0.03,0.05,0.05	0.04 (0.42)	0.02,0.00,0.01	0.03,0.08,-0.16	0.16,0.43,0.75
(N=54)		(0.02,0.06,0.05)		(0.36,0.09,0.03)	(0.57,0.22,-0.05)	(1.00,0.87,0.27)
CIO	Annual	0.10,0.06,0.07	0.35 (0.59)	0.15,0.18,0.02	0.38,0.38,-0.29	0.99,0.99,0.95
(N=51)		(0.15,0.10,0.06)		(0.31,0.28,0.05)	(0.53,0.40,-0.39)	(1.00,0.99,0.99)
US	Annual	0.21,0.10,0.27	0.25 (0.38)	0.10,0.06,0.02	0.49,0.23,-0.29	0.99,0.98,0.99
(N=128)		(0.21,0.09,0.26)		(0.05,0.22,0.05)	(0.44,0.46,-0.29)	(0.99,0.99,0.99)
AMZ	Annual	0.06,0.15,0.15	0.58 (0.78)	0.18,0.32,0.22	0.42,0.37,-0.40	1.00,0.99,0.99
(N=202)		(0.05,0.18,0.17)		(0.15,0.63,0.26)	(0.32,0.63,-0.29)	(0.99,1.00,0.99)

Region	Season	Multicollinearity	Multiple R²	Individual R²	Estimates, β_p	Significance, β_p
CNG (N=137)	Annual	0.12,0.02,0.13 (0.11,0.04,0.14)	0.32 (0.36)	0.14,0.16,0.00 (0.13,0.24,0.00)	0.42,0.41,-0.08 (0.38,0.46,-0.10)	0.99,0.99,0.59 (0.99,0.99,0.73)
IND (N=25)	Annual	0.26,0.51,0.39 (0.25,0.45,0.30)	0.60 (0.74)	0.17,0.60,0.19 (0.17,0.74,0.26)	0.05,0.77,0.04 (0.01,0.81,-0.07)	0.16,0.98,0.12 (0.03,0.99,0.24)

[e]Experimental regression model statistics for TLD in each region and all months in the eight-year sample. Linear and logarithmic model forms are detailed (output from logarithmic model forms are provided in parentheses). Multicollinearity is the linear multiple regression coefficient between each predictor and the rest of the independent set. Individual R^2 accounts for relationships between individual predictors to the response variable, Y. Multiple R^2 illustrates strength of the relationship between each combination of predictors and the response variable, Y. Significance of each parameter estimate is given in terms of a two-tailed probability that the absolute value of a random t-variate is less than or equal to the parameter estimate in question. Data triplets in certain categories are indicative of the values assigned to representations (either linear or logarithmic) of NCAPE, CCN, and WCD respectively. The number of observations is provided under each region in parentheses (regions with less than 20 data points were omitted from this analysis).

Table 2.6. Statistics of regression model forms 4 (linear) and 7 (logarithmic) for AVGHT30[f]

Region	Season	Multicollinearity	Multiple R²	Individual R²	Estimates, β_p	Significance, β_p
GLOBE (N=1052)	Annual	0.11,0.00,0.11 (0.12,0.04,0.10)	0.61 (0.73)	0.29,0.20,0.03 (0.23,0.54,0.03)	0.65,0.43,-0.39 (0.44,0.65,-0.25)	1.00,1.00,1.00 (1.00,1.00,1.00)
CONT (N=844)	Annual	0.09,0.00,0.09 (0.09,0.06,0.13)	0.57 (0.63)	0.24,0.08,0.11 (0.19,0.33,0.12)	0.65,0.27,-0.52 (0.52,0.47,-0.39)	1.00,1.00,1.00 (1.00,1.00,1.00)
OCEAN (N=779)	Annual	0.12,0.00,0.12 (0.10,0.02,0.10)	0.58 (0.63)	0.17,0.13,0.12 (0.15,0.34,0.11)	0.61,0.39,-0.56 (0.49,0.51,-0.44)	1.00,1.00,1.00 (1.00,1.00,1.00)
GS (N=197)	Annual	0.09,0.02,0.10 (0.08,0.02,0.10)	0.47 (0.48)	0.07,0.31,0.07 (0.08,0.31,0.08)	0.36,0.52,-0.31 (0.36,0.52,-0.31)	1.00,1.00,1.00 (1.00,1.00,1.00)
GG (N=113)	Annual	0.07,0.07,0.01 (0.06,0.07,0.01)	0.45 (0.49)	0.39,0.13,0.01 (0.34,0.23,0.01)	0.58,0.23,-0.14 (0.50,0.37,-0.17)	1.00,0.98,0.87 (1.00,1.00,0.93)
SAM (N=126)	Annual	0.08,0.01,0.07 (0.09,0.03,0.08)	0.36 (0.40)	0.07,0.16,0.10 (0.05,0.25,0.10)	0.33,0.37,-0.40 (0.28,0.44,-0.36)	1.00,1.00,1.00 (1.00,1.00,1.00)
ZAF (N=93)	Annual	0.07,0.31,0.30 (0.04,0.33,0.34)	0.41 (0.50)	0.02,0.30,0.30 (0.01,0.44,0.30)	0.13,0.33,-0.40 (0.11,0.51,-0.28)	0.78,0.99,1.00 (0.71,1.00,0.97)
CA (N=159)	Annual	0.13,0.15,0.04 (0.18,0.20,0.03)	0.63 (0.60)	0.59,0.19,0.00 (0.46,0.38,0.01)	0.70,0.19,-0.07 (0.52,0.40,-0.01)	1.00,0.97,0.61 (1.00,1.00,0.18)
AUS (N=34)	Annual	0.15,0.13,0.21 (0.12,0.10,0.17)	0.29 (0.27)	0.16,0.06,0.11 (0.09,0.07,0.12)	0.25,0.30,0.34 (0.16,0.35,0.38)	0.81,0.88,0.91 (0.59,0.94,0.95)
WP (N=161)	Annual	0.00,0.00,0.00 (0.00,0.01,0.02)	0.59 (0.59)	0.39,0.19,0.00 (0.35,0.23,0.00)	0.64,0.45,0.02 (0.59,0.49,0.07)	1.00,1.00,0.15 (1.00,1.00,0.14)
CIO (N=150)	Annual	0.03,0.02,0.04 (0.04,0.05,0.06)	0.58 (0.59)	0.38,0.22,0.00 (0.35,0.31,0.00)	0.60,0.42,-0.08 (0.54,0.49,-0.05)	1.00,1.00,0.67 (1.00,1.00,0.44)
US (N=180)	Annual	0.14,0.12,0.21 (0.16,0.11,0.23)	0.44 (0.38)	0.12,0.04,0.16 (0.07,0.07,0.15)	0.57,0.11,-0.57 (0.50,0.18,-0.52)	1.00,0.81,1.00 (1.00,0.97,1.00)
AMZ (N=278)	Annual	0.10,0.15,0.18 (0.10,0.21,0.23)	0.63 (0.71)	0.16,0.46,0.14 (0.16,0.60,0.14)	0.40,0.54,-0.28 (0.35,0.65,-0.19)	1.00,1.00,1.00 (1.00,1.00,0.99)
CNG (N=167)	Annual	0.12,0.02,0.14 (0.12,0.04,0.15)	0.41 (0.44)	0.08,0.25,0.04 (0.04,0.37,0.05)	0.39,0.48,-0.27 (0.27,0.57,-0.21)	1.00,1.00,1.00 (1.00,1.00,0.98)
IND (N=75)	Annual	0.19,0.26,0.14 (0.17,0.15,0.12)	0.32 (0.32)	0.17,0.25,0.00 (0.13,0.24,0.00)	0.25,0.43,0.07 (0.27,0.46,0.06)	0.94,1.00,0.43 (0.97,1.00,0.40)

[f]As in Table 2.5 except for AVGHT30.

In order to assess the relative importance of each predictor as they contribute to the variability in convective intensity proxies, a compilation of the parameter estimates for regions where the logarithmic model forms produced statistically significant output is presented for both

TLD and AVGHT30 (Figs. 2.27-2.28; annual results). Recall that the relative weight assigned to each predictor accounts for changes in the response, holding all other inputs fixed, and the results are directly comparable since all variables have been standardized prior to computing the regression output (these values may be interpreted as dependence "per unit" input of each predictor). For TLD, the logarithmic model in these regions explains between 38-79% of the variance of the predictand and the relative weight on CCN is greater than the relative weight on NCAPE in all regressions except for the CIO (Fig. 2.27); for the global dataset, the relative dependence on CCN is more than double that for NCAPE or WCD. Over continents, CCN is still the leading component explaining the variability in the response according to the logarithmic model but the relative weight is comparable to the other two inputs, NCAPE and WCD. Over oceans, CCN and WCD are approximately equally weighted, while the weight assigned to NCAPE is about 32% smaller than either CCN or WCD. Over the CNG, US, and CIO, the contributions of NCAPE and CCN are roughly equal (and positive) while the dependence on WCD is negative and appears to be of secondary importance.
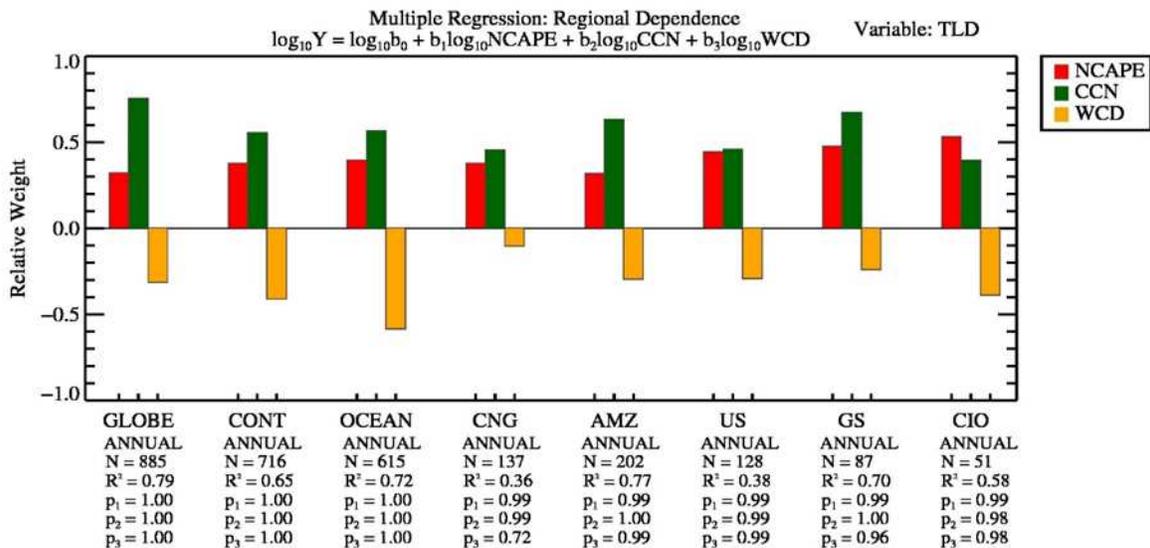
Fig. 2.27. Estimates of the relative weight assigned to each independent parameter resulting from the multiple linear regression using logarithmic model form #7 (*see Table 2.4*) for TLD for different regions of interest (shown in the

legend) for all months during the years 2004-2011. The number of samples (N), multiple correlation (R^2), and significance of the qth predictor (p_q; *see text for definition of statistical significance*) are provided for reference along the abscissa.

Multiple Regression: Regional Dependence
$$\log_{10}Y = \log_{10}b_0 + b_1\log_{10}NCAPE + b_2\log_{10}CCN + b_3\log_{10}WCD$$
Variable: AVGHT30
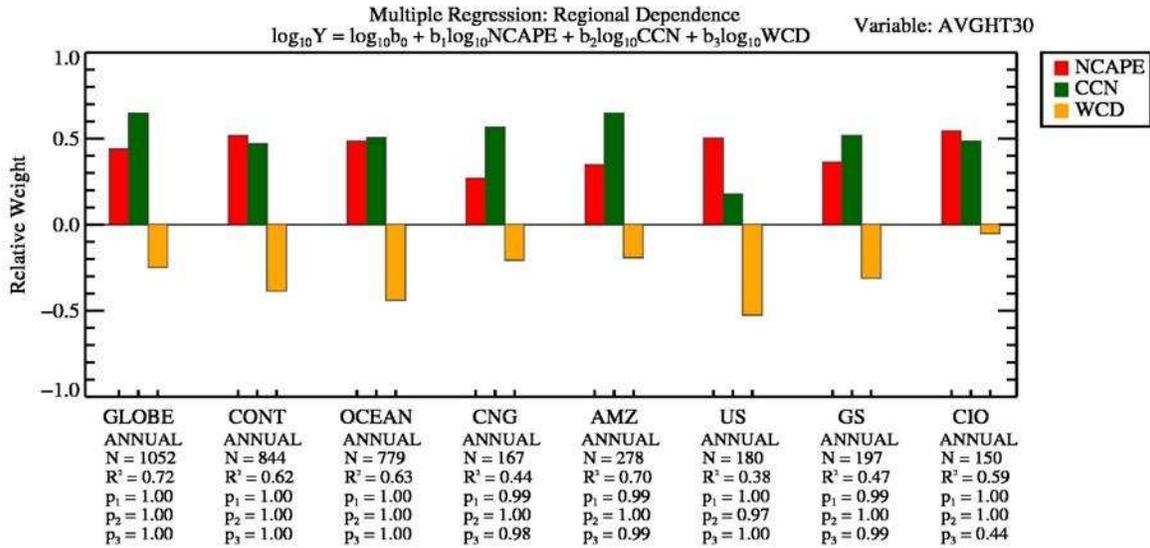
Fig. 2.28. As in Fig. 2.27, but for AVGHT30.

In general, the logarithmic regressions show higher AVGHT30 for higher NCAPE,

higher CCN, and shallower WCD for individual regions as well as on the global scale annually.

The logarithmic model analyzed based on the three predictors in the selected regions offers a

range of predictive skill as it captures about 38-72% of the variance in the predictand. Like the

results for TLD, the regional dependence of AVGHT30 on NCAPE, CCN, and WCD again

appears to be quite variable. Furthermore, comparing the results for TLD and AVGHT30 *in the*

same region shows that the sensitivity of the each convective intensity proxy to NCAPE, CCN,

and WCD is similar, corroborating the strong correlation between TLD and AVGHT30 shown in

Sec. 2.3.5.

In order to address potential differences in seasonal sensitivity of convective intensity

proxies to NCAPE, CCN, and WCD in the logarithmic model, broad hemispheric study areas

had to be tested to account for the decrease in sample size that resulted from limiting the data

subsets to individual seasons (Fig 2.28-2.29; note that concurrent regional and seasonal subsets

were studied, but the parameter estimates were statistically unstable, perhaps due to prohibitively small sample sizes). Hemispheric subsets were analyzed for individual warm seasons (JJA for the Northern Hemisphere and DJF for the Southern Hemisphere) to reduce the potential complication that may arise from including tropical (e.g., upright, warm-season type) and subtropical/extratropical (i.e., convective clouds governed by midlatitude dynamics) features in the same sample. Results for the logarithmic model regression for individual seasons in each hemisphere are in close agreement with the annual regression output presented above (note annual results in individual hemispheres are also shown in Figs. 2.29-2.30 for validation purposes), again emphasizing the consistency of the current statistical results across multiple scales of interest.
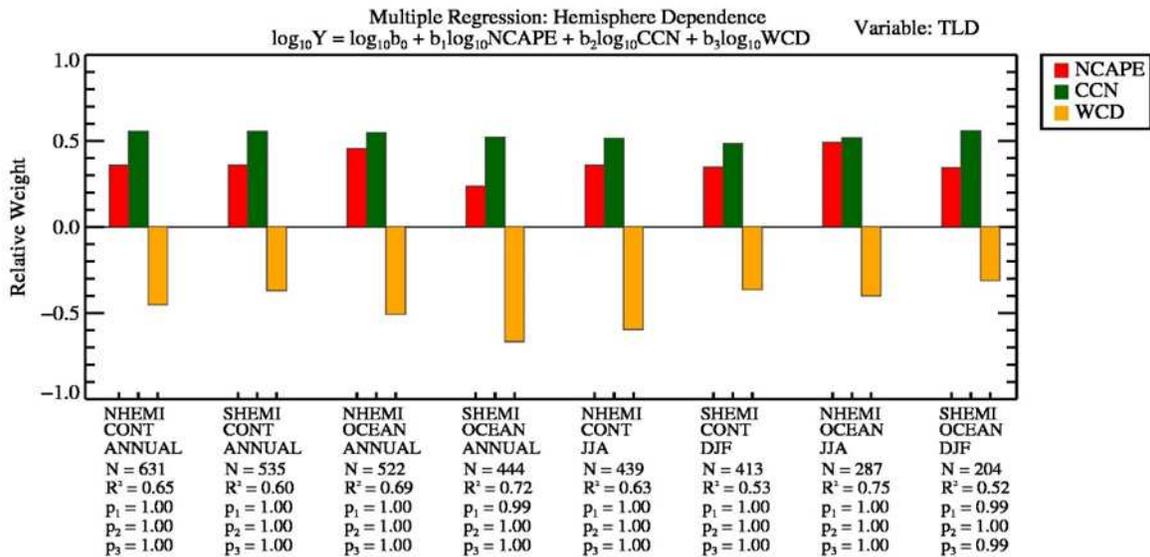
Fig. 2.29. As in Fig. 2.27, but for TLD in individual hemispheres (seasonal vs. annual datasets are identified along the abscissa for each hemispheric data subset).

76

Multiple Regression: Hemisphere Dependence

$$\log_{10}Y = \log_{10}b_0 + b_1\log_{10}NCAPE + b_2\log_{10}CCN + b_3\log_{10}WCD$$

Variable: AVGHT30
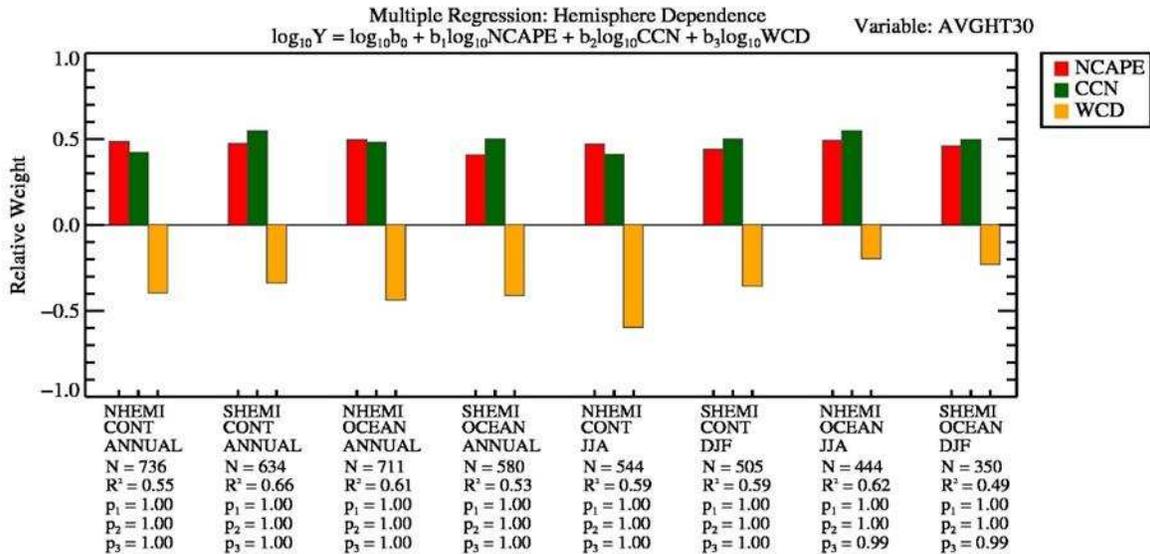
Fig. 2.30. As in Fig. 2.27, but for AVGHT30 in individual hemispheres (seasonal vs. annual datasets are identified along the abscissa for each hemispheric data subset).

A tangential study of the applicability of the statistical approach to global chemical transport modeling was motivated by previous studies that attempted to parameterize lightning in order to more accurately resolve and model chemical constituents of the atmosphere [*Pickering et al.*, 1990,1998]. For example, the frequency and vertical distribution of lightning is theoretically linked to the production of nitrogen oxides and therefore lightning has been implicated for altering subsequent distributions ozone, an important green-house gas [*Schumann and Huntrieser*, 2007]. *Allen and Pickering* [2002] developed parameterizations for satellite observations of lightning flash rates using the vertical mass flux, convective precipitation rate, and cloud top height but found considerable disagreement between their regressions and observations, particularly over tropical oceans, citing bias in their modeled predictor set. However, it is well-known the convective cloud tops in the Tropics may reach the tropopause in the absence of strong updrafts ($w > 5$-8 m s^{-1}) [*Zipser*, 2003] or appreciable vertical reflectivity column development and lightning [*Zipser and Lutz*, 1994; *DeMott and Rutledge*, 1998]. It is possible that weak relationships between parameterized and observed lightning flash rates in

previous global modeling studies [e.g., *Allen and Pickering*, 2002] resulted from focusing on variables that were too indirectly linked to the charge separation process.

The foregoing results support the notion that NCAPE, CCN, and WCD in combination are important drivers of the observed variability in lightning on the global scale. Accordingly, "real" values (as opposed to standard anomalies) of lightning, NCAPE, CCN, and WCD were used as inputs for the computation of ordinary least-squares regression parameters assuming the logarithmic model form (using all available data). The multiple correlation for the real-valued, logarithmic regression was $R^2=0.79$, the F-statistic for the regression was significant at the $p < 0.01$ level, and individual t-statistics for parameter estimates were significant at the $p < 0.10$ level (see Fig. 2.31). While it is beyond the scope of the current research, tests of the sensitivity of NO_x in global chemical transport models to lightning parameterized by NCAPE, CCN, and WCD following the real-valued regression analysis conducted here are required before implementation. As new emissions inventories become available for years following 2011, cross-validation of the empirical model developed here using independent data sets will be possible.

As mentioned above, multiple noise sources are likely to contribute to the observed spread/variance of the predictand for each NCAPE, CCN, and WCD bin. The sensitivity of the results of these statistical analyses (e.g., regression strength and parameter estimates for the various model forms) to changes in bin increments for averaging was examined by adjusting the bin increment from fine to coarse, according to multiples of 0.10 and 0.50. To illustrate, the above analysis used a multiple of 0.25 such that bin increments were 0.025 J kg^{-1} m^{-1} for NCAPE, 0.25 for CCN (logarithmic units, cm^{-3}), and 250 m for WCD. Regressions tended to improve as the number of bins decreased (i.e., for coarse bin increments); simply put, for a

smaller number of bins more noise in the data is smoothed out. However, as the binning became

coarser, the sample size decreased in some regions to the point where the possibility of

overfitting in the regression became a concern (see *Draper and Smith* [1998], *Wilks* [2011], and

other texts on applied statistics).

GLOBE_ANNUAL REGRESSION
Samples:885

$R^2 = 0.79450273$

$\log_{10}(Y) = \log_{10}(\beta_0) + \beta_1 * \log_{10}(X1) + \beta_2 * \log_{10}(X2) + \beta_3 * \log_{10}(X3)$

$\log_{10}(\beta_0) = -0.48819761$
X1 = NCAPE [J kg^{-1} m^{-1}]; $\beta_1 = 0.17677160$; $p = 0.063567758$
X2 = CCN [cm^{-3}]; $\beta_2 = 0.22952378$; $p = 5.3644180e{-}06$
X3 = WCD [m]; $\beta_3 = -0.65417974$; $p = 0.068196177$
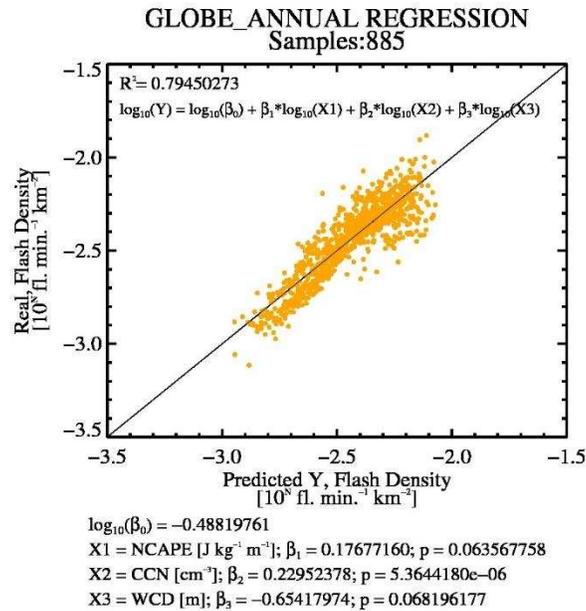
Fig. 2.31. Observed TLD versus predicted TLD based on multiple linear regression for model #7 (model form and multiple correlation are shown in the upper left) using real variables. The parameter estimates and their associated p-values have been included below the main figure.

It is important to remember that the set of predictors examined here was chosen

according to dominant hypotheses in the literature, i.e., the thermodynamic hypothesis and the

aerosol hypothesis. However, previous research points to the importance of low- or mid-level

shear [*Fan et al.*, 2009,2013] as well as environmental humidity [*Grant and van den Heever*,

2014,2015] in determining the behavior of deep convective clouds. From this perspective, the

fact that the multiple correlation coefficients for linear models of convective intensity proxies

based on NCAPE, CCN, and WCD alone are in some cases well below 1.0 is somewhat

unsurprising. For example, *Grant and van den Heever* [2015] found that precipitation intensity

in multi-cellular convection was sensitive to increasing aerosols when considering the evaporation rate of cloud droplets with respect to changing the height and ambient humidity of environmental dry layers; higher evaporation rates and stronger cold-pool forcing of secondary convection resulted from dry layers located near cloud base, leading to an overall increase in rainfall. Though not explicitly addressed in the aforementioned studies, modeled changes in secondary-forcing resulting from aerosol-evaporation-dynamical feedbacks within deep convective clouds would likely be manifest by variations in mixed-phase radar reflectivity and lightning frequency in observations via the implication of stronger subsequent updrafts. Hence, future studies that seek to extend the simultaneous observation approach implemented in the current investigations may benefit by including environmental factors that were not treated here.

2.6 Summary of findings on global and regional scales

The variability of radar reflectivity and lightning in convective features observed by TRMM's precipitation radar and lightning imaging sensor in the Tropics for 2004-2011 in response to thermodynamics and CCN was investigated. The thermodynamic environment and CCN concentrations in the vicinity of deep convective clouds were characterized using reanalysis and a chemical transport model, GEOS-Chem. The advantage of the strategy arises from the continuity of reanalysis and model output as it allows for a large number of convective features in remote areas across the global Tropics and Subtropics lacking in-situ observations to be included in the data sample. The objective was to use observations to validate a hypothesis put forth by Rosenfeld et al. [2008] - that aerosols modulate the release of conditional instability such that convective updrafts, radar reflectivity, and lightning are enhanced during the mature phase of convective cloud development.

In the global sample of deep convective features observed by the TRMM satellite, we found that TLD (AVGHT30) increased by more than 600% (2-3 km) as a function of both CCN concentrations and thermodynamic instability simultaneously when WCD was held constant. The trends observed for the global populations of CFs and LPCFs were also found in the subsets of these populations over continents and oceans. Importantly, the changes in TLD and AVGHT30 with respect to NCAPE or CCN separately (holding other independent variables fixed) were comparable in many of the cases examined here. The partial sensitivities of TLD and AVGHT30 accounted for only a fraction of the total range of variability in the global populations of CFs and LPCFs when considering simultaneous changes with respect to both thermodynamics and aerosols. Meanwhile, TLD (AVGHT30) was shown to increase by up to 91% (1.25 km) between shallower and deeper WCD for a given combination of NCAPE and CCN, with greater (higher) values shown in environments with shallower WCD. *Thus, total lightning density and the average height of 30 dBZ echoes were maximized for high NCAPE, shallower WCD, and high CCN for the global population of deep convective CFs as well as over both continental and oceanic regions.*

The relationship between TLD and AVGHT30 was shown to be strongly positive (r=0.95-0.98), but variable as a function of background CCN; the results suggest that clouds developing in more pristine environments must have 30 dBZ echoes reaching to higher altitudes in order to produce similar lightning rates when compared to more-polluted clouds. The mean reflectivity in the mixed-phase region of deep convective clouds was shown to be up to 5.6 dB higher in the most-polluted environment relative to pristine environments. Finally, the analysis of VPRR illustrated that the differences in radar reflectivity at a given height in pristine and polluted clouds diminishes (the differences in reflectivity range from 5.0-5.4 dB) when NCAPE

is above the global median value. Based on the results presented, one may speculate that the impact of aerosols on the development of deep convection becomes less pronounced when NCAPE is high and WCD becomes progressively shallower (note that the 35° inclination angle of the TRMM satellite meant that the results apply to deep convective clouds in the Tropics and Subtropics only, i.e., clouds with intermediate to deeper WCD).

This research represents the first effort to use a chemical transport model to more accurately attribute CCN number concentrations to individual deep convective features over a range of thermodynamic environments in order to investigate their simultaneous influence on the variability of radar reflectivity and lightning between continental and oceanic regions on the global scale. The results provide observational evidence supporting a merged hypothesis for the impacts of thermodynamics and aerosols on deep convective clouds in the Tropics and Subtropics. Meanwhile, the results also contribute insight about the simultaneous influence of warm cloud depth.

Importantly, the results from the multiple linear regression analysis provide quantitative support for the simultaneous hypothesis for the role of thermodynamics and aerosols as they may influence the development of convective clouds. Roughly 50-75% of the variance in convective intensity proxies is explained by simple linear models based on NCAPE, CCN, and WCD. Convective intensity proxies were found to be directly proportional to both NCAPE and CCN and inversely proportional to WCD consistently in different regions annually as well as in individual seasons. Noting the magnitude and sign of the parameter estimates and inverting the linearized logarithmic representation of the model back to the multiplicative form (see Eq. 2.2), the maximum sensitivity of the convective intensity proxies is theoretically found for smaller

(shallower) values of the independent variables; in other words, the largest changes in TLD/AVGHT30 are found when NCAPE and CCN are small and WCD is shallower.

Despite similar general sensitivity of TLD and AVGHT30 to NCAPE, CCN, and WCD in the multiple regression framework, there are still appreciable regional differences in the relative weight assigned to NCAPE, CCN, and WCD for both annual and seasonal data subsets. It is intriguing that the linear models' predictive skill varies between regions that are apparently similar from a climatological perspective. For example, *Williams and Sátori* [2004] investigated environmental and atmospheric thermodynamic characteristics of two tropical chimneys, the Amazon and Congo river basins, to understand differences in lightning climatology; although *Williams and Sátori* [2004] found only modest thermodynamic differences between the Amazon and Congo to explain differences in lightning density, the results of the current study suggest that a linear model of TLD based on NCAPE, CCN, and WCD performs reasonably well over the Amazon ($R^2 = 0.77$) and poorly over the Congo ($R^2 = 0.36$). Hence, the need to investigate the relative importance of the chosen predictors in different environments in order to link the observed statistical behavior to physical meteorological and aerosol variability is highlighted. Next, the focus of the investigation shifts to a small-scale region that has received comparatively little attention to date, the central equatorial Indian Ocean (CIO).

Historically, simultaneous observations of thermodynamics, aerosols, and the characteristics of deep convective clouds have been difficult to obtain over of the CIO due to the remoteness of this region and limited availability of surface observation sites. However, temporal and spatial continuity of the aerosol data derived from GEOS-Chem and thermodynamic data from global reanalysis readily lend themselves to addressing the current

uncertainty surrounding the interaction between thermodynamics, aerosols, and convective

clouds over the CIO.

CHAPTER 3

Aerosol-cloud-MJO interactions

3.1 Background

The discovery of 40-50 day oscillations in the zonal winds at 850 and 150 hPa in the tropical atmosphere over 45 years ago has incited a rich interest in researching and characterizing the Madden-Julian Oscillation [MJO; *Madden and Julian,* 1971,1972]. The MJO influences patterns of variability of lower and upper tropospheric winds, humidity, and temperature on 30-90 day timescales, and anomalies in the large-scale circulation associated with the MJO have been shown to circumnavigate the global Tropics [*Madden and Julian*, 1994; *Zhang*, 2005]. Furthermore, the MJO has been implicated as a modulator of regional monsoon patterns, tropical cyclones, tornadoes, lightning, and extratropical weather/climate [*Zhang*, 2013]. The Dynamics of the Madden-Julian Oscillation (DYNAMO) field campaign [*Yoneyama et al*., 2013] took place during boreal fall and winter of 2011-2012 in the CIO. One of the overarching objectives of the field campaign was to improve understanding of the mechanisms that govern the onset and propagation of the MJO, such as the transition from shallow to deep convection [e.g., *Stephens et al.*, 2004; *Yoneyama et al.*, 2013].

In recent years, the impact of the MJO on atmospheric composition has received increasing attention, as the literature over multiple decades advocates a strong link between atmospheric aerosols and the microphysical and dynamical evolution of convective clouds [e.g., *Tao et al.*, 2012; *Rosenfeld et al.*, 2014; and many others]. Observations from the Indian Ocean Experiment [INDOEX; *Ramanathan et al.*, 2001] identified a strong meridional gradient in aerosol concentration in the CIO [e.g., *Satheesh et al.*, 1998; *Moorthy and Saha*, 2000;

Ramanathan et al., 2001] and a substantial seasonal cycle related to the changing large-scale flow patterns of the Asian monsoon [*Li and Ramanathan*, 2002]. These plumes of continental aerosols have been linked to parts of Southern Asia and the Indian Subcontinent in particular [*Krishnamurti et al.*, 1998; *Bates et al.*, 2002].

Previous researchers have demonstrated distinct patterns of variability in the distribution of atmospheric aerosols in association with the MJO globally, over major ocean basins, and at individual observation sites [*Tian et al.*, 2008,2011; *Beegum et al.*, 2009; *Guo et al.*, 2013; *Langley DeWitt et al.*, 2013]. Key findings from these earlier studies include a robust inverse-relationship between local aerosol concentrations and rainfall, presumably due to wet scavenging, as well as notable changes in aerosol concentration, size distributions, and/or composition [e.g., *Tian et al.*, 2008; *Langley DeWitt et al.*, 2013] as a function of MJO phase [e.g., *Wheeler and Hendon*, 2004]. Specifically, *Langley DeWitt et al.* [2013] analyzed data collected from the R/V *Roger Revelle* during DYNAMO and found significant variations in the concentrations of sea-salt and anthropogenic aerosols over the CIO before, during, and after peak convective activity associated with the October, November, and December 2011 MJO episodes. Sea-salt aerosols were shown to increase during westerly wind burst events in the wake of active MJO convection owing to disturbed sea states and wave breaking. *Langley DeWitt et al.* [2013] also documented several instances of aerosol concentrations reminiscent of continental conditions along the Equator at 80.5°E longitude.

There is outstanding uncertainty as to 1) the origin of these episodic "outbreaks" of high aerosol concentrations on intraseasonal timescales over the remote reaches of the CIO and 2) whether any discernible signature of the impact of elevated aerosol concentrations on convection associated with the MJO exists. The seminal paper on the subject of aerosol-cloud interactions

by *Rosenfeld et al.* [2008] details a hypothesis regarding convective invigoration and the transition from shallow to deep convection based on local variations in aerosol concentrations in the subcloud layer. In a recent study of individual convective features across the global Tropics, *Stolz et al.* [2015] noted stronger convection with increasing aerosol concentrations over oceanic regions for fixed thermodynamic conditions. Non-negligible differences in radar-observed precipitation structure occurred for relatively small increases in aerosol concentrations (100-200 cm^{-3}) in pristine background environments, in agreement with previous investigations [e.g., *Storer and van den Heever*, 2013; *Storer et al.*, 2014; *Sheffield et al.*, 2015]. Thus, it is possible that even modest changes in aerosol concentrations could impact deep convective clouds associated with the initiation of MJOs over the CIO. A pressing question is whether the interaction between clouds and aerosols in the CIO is bi-directional. Previous research corroborates that the MJO's large-scale wind anomalies affect local concentrations of atmospheric aerosols over the CIO [e.g., *Langley DeWitt et al.*, 2013], but do these changes in atmospheric aerosol concentrations above the background pristine state influence the population of convective clouds on the basin scale and subsequently feedback onto the MJO itself?

The goal of this study is to build on the findings of previous research by investigating the sources, transport, and impact of atmospheric aerosols over the CIO during three specific MJO episodes observed in DYNAMO. We will leverage a set of satellite and ground-based observations to study the distribution of convective clouds, their radar reflectivity structure, and lightning characteristics as they may relate to changes in atmospheric composition. The CIO represents an ideal natural laboratory to explore the potential influence of varying aerosol concentrations on deep convection [e.g., *May et al.*, 2009] associated with the MJO phenomenon

since thermodynamic conditions over the CIO are not expected to vary significantly in space over the region (Fig. 3.1).
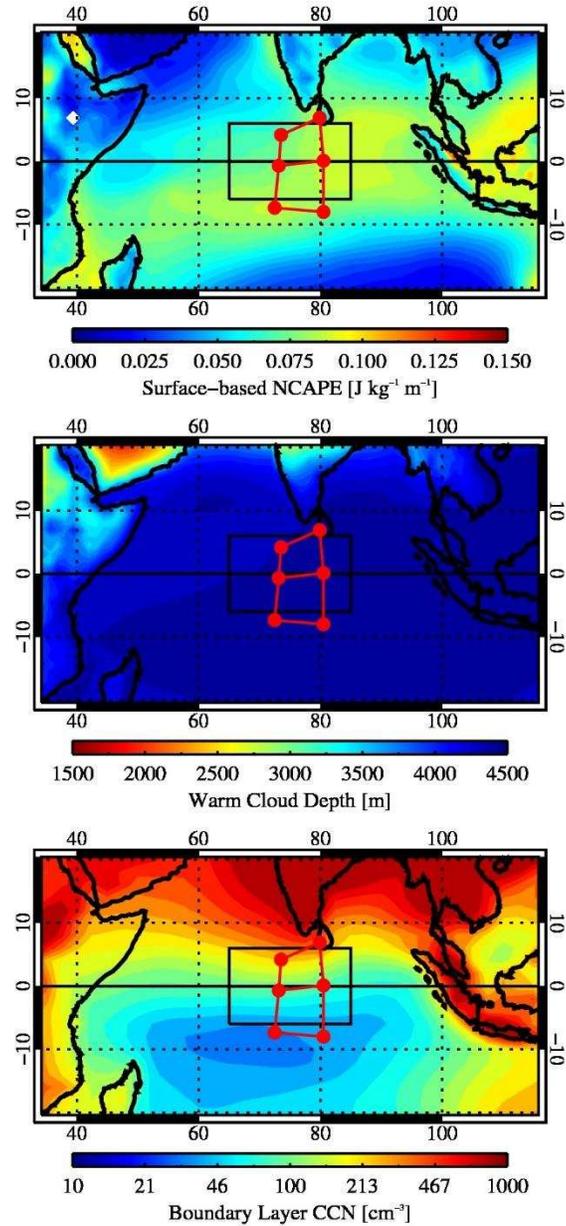
Fig. 3.1. Daily-average values of Surface-based NCAPE (top), WCD (middle), and CCN (bottom) for boreal cold seasons during the eight-year period between 2004-2011 over the CIO domain (20°S-20°N, 35-115°E).

3.2 Data, methods, and results

DYNAMO intensive observation periods during October-December 2011 allowed for an unprecedented view of multiple MJO episodes using a specialized suite of radar, satellite, and ground-based instrumentation in addition to upper air observations at strategic positions in an organized quadrilateral array over the CIO [e.g., *Johnson and Ciesielski*, 2013; *Yoneyama et al.*, 2013; *Xu and Rutledge*, 2014]. For the current study, the geographical extent of these northern and southern domains has been expanded relative to the sounding array dimensions from earlier studies to accommodate spatial and temporal sampling considerations of the Tropical Rainfall Measuring Mission (TRMM) satellite (6^0S-6^0N, 65-85^{0}E; Fig. 3.2). Specifically, geostationary infrared satellite imagery (METEOSAT-7), cloud-to-ground (CG) lightning from Vaisala's Global Lightning Dataset (GLD360), space-borne radar observations of basin-wide rainfall characteristics, and vertical reflectivity structure from the TRMM precipitation radar (PR) [*Kummerow et al.*, 1998; *Iguchi et al.*, 2000] 3B42 and 2A25 products, respectively, were analyzed in the context of background thermodynamics and winds from the Modern Era Retrospective Analysis for Research and Applications (MERRA; http://gmao.gsfc.nasa.gov/merra/). Note that Vaisala's GLD360 network detected approximately 50-60% of CG strokes exceeding 10 kA in peak current during the time period of interest over the CIO [R. Said, *personal communication*].
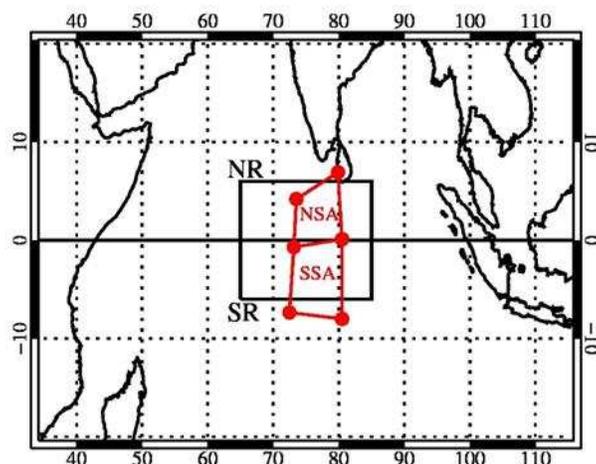
Fig. 3.2. The geographical domain of the central equatorial Indian Ocean spanning 20°S-20°N, 35-115°E. The northern (NSA) and southern sounding arrays (SSA) from DYNAMO (red solid/dots), wider northern (NR) and southern (SR) study regions (6°S-6°N, 65-85°E; black solid boxes), and outlines of major coastlines are also shown.

This study utilizes data from a chemical transport model, GEOS-Chem (www.geos-chem.org) with the online TOMAS aerosol microphysics module [*Adams and Seinfeld*, 2002; *Pierce and Adams*, 2009; *D'Andrea et al.*, 2013, *Pierce et al.*, 2013], to simulate the number concentration of aerosols with diameters greater than or equal to 40 nm (N40) across a domain spanning the Indian Ocean and portions of adjacent continents (20⁰S-20⁰N, 35-115⁰E; see *Stolz et al.*, [2015] and references therein for a detailed discussion of model design and uncertainty). N40 is assumed to be an appropriate proxy for the number concentration of cloud condensation nuclei [CCN; *Dusek et al.*, 2006]. In addition, we assume that the spatial and temporal variability of CCN in the lower troposphere is adequately characterized by relatively coarse resolution used in the model (roughly 200 km in the horizontal, 10 vertical levels between 1000 and 850 hPa). Plumes of continental aerosols have been shown to exist primarily in layers extending up to 400-3000 m above the ocean surface in remote regions of the Arabian Sea, Bay

of Bengal, and the tropical Indian Ocean [*Satheesh et al.*, 1999; *Ramanthan et al.*, 2001]; hence, we represent CCN for our study by the boundary layer average for 1000-850 hPa.

The temporal resolution of individual datasets varied, therefore, we chose to assimilate each data source to a six-hour interval to capture salient aspects of variability in the convective cloud population as well as variability on longer timescales (e.g., by analyzing time series of relevant quantities). CCN concentrations, the horizontal components of the 925 hPa wind, and surface-based pseudoadiabatic normalized convective available potential energy (NCAPE) were analyzed at 0, 6, 12, and 18 Z from 1 October – 31 December 2011 over the CIO domain. CG lightning, the distribution of cold cloud features (CCFs; defined by contiguous areas of brightness temperature, $T_b < 208$ K in IR channel 4), and rainfall rate from the TRMM 3B42 product were compiled over the ensuing six-hour period prior to the next model time step. Note that for each interval, the distribution of NCAPE and CAPE were computed over the same study region to track potential changes in background thermodynamics; however, the mean values of CAPE (NCAPE) between the northern and southern hemisphere domains of interest differed by approximately 200 J kg^{-1} (0.02 J kg^{-1} m^{-1}) or less during the three-month study period.

We carefully analyzed the simultaneous evolution of rainfall, CCFs, CCN, CG lightning, and NCAPE for the three MJO episodes observed during DYNAMO. The results are succinctly summarized in Fig. 3.3, which shows times series of TRMM 3B42 rainfall, the aggregate area of METEOSAT7 CCFs (< 208K), CG lightning stroke rate densities, boundary layer average CCN, and surface-based NCAPE over both the northern and southern regions. The raw time series (colors) have been smoothed using a 29-time step (1 week) boxcar running mean (thick, black solid) to gain insight about variations on longer timescales.

MJO episodes are clearly evident (hereafter labeled MJOs 1, 2, and 3, for the October, November, and December events respectively) in both the northern and southern regions according to variations in TRMM 3B42 rainfall; the same signal is also apparent in the CCF observations in both regions. CG lightning occurs more frequently in the northern hemisphere compared to the southern hemisphere and it is evident that peaks in CG lightning density occur just prior to peaks in heavy rainfall (especially in the northern region). Boundary layer CCN concentrations vary on MJO timescales (perhaps in relation to the MJO itself) and CCN concentrations are markedly higher in the northern region compared to the southern region. Peak CCN concentrations in the northern region commonly exceed 500 cm^{-3}, while peak CCN concentrations in the southern region rarely exceed 100 cm^{-3}. We note that the aggregate area of the cold cloud feature populations (second panels in Fig. 3.3) are very similar between the northern and southern regions over the course of the DYNAMO-observed MJOs, yet the bulk of the lightning occurs in the northern region. Furthermore, there is little evidence of MJO modulation of NCAPE in either the two study regions with the exception of increasing NCAPE associated with increasing sea surface temperature prior to heavy rain periods [*Xu and Rutledge*, 2014].
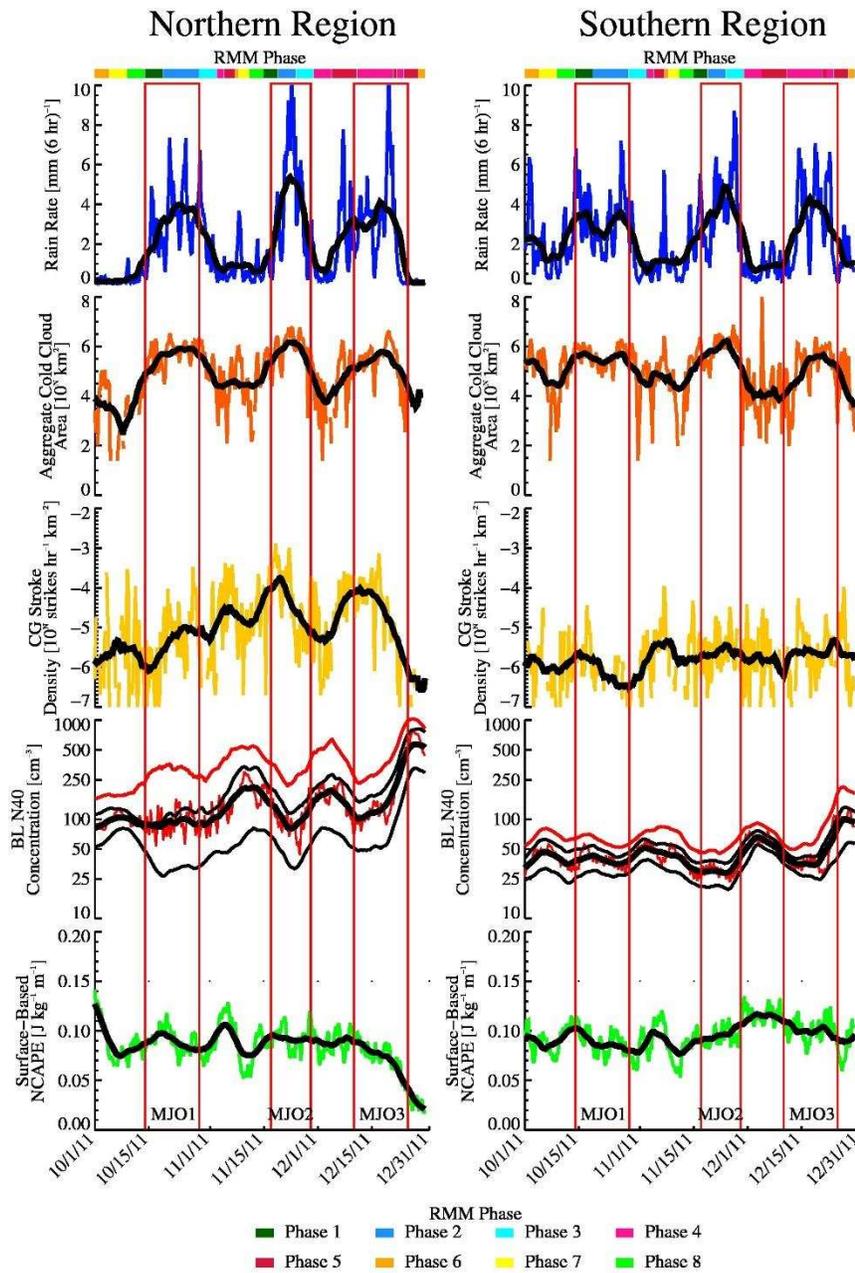
Fig. 3.3. Time series of TRMM 3B42 rain rate, the aggregate area of cold cloud features, cloud-to-ground lightning flash density, CCN (25th, 75th , and 90th percentile values at each time step have been overlaid) and surface-based normalized convective available potential energy for 3 MJO periods documented in DYNAMO (vertical red solid bars) over the northern (left panel) and southern (right panel) study regions. Thick black, solid lines in each tile represent the 29-time step (1 week) boxcar average. The real-time multivariate MJO index is delineated at the top of each panel for reference.

There is clear evidence that aerosols are rapidly depleted due to washout (note the quadrature relationship of the rainfall and CCN time series in Fig. 3.3 in both the northern and

southern regions) as the heavy MJO rain sets in, consistent with *Langley Dewitt et al.* [2013]. A subtle, yet important, aspect of the covariability between rainfall, lightning, NCAPE, and CCN, is that CCN concentrations are diminishing (in the domain-average sense) as lightning activity increases, presumably associated with isolated, deep convection, prior to the onset of widespread rainfall. However, CCN concentrations north of the equator during these periods of enhanced lightning still tend to be greater by a factor of two or more than what is typically observed in pristine oceanic regions [e.g., *Heintzenberg et al.*, 2000; *Spracklen et al.*, 2011]. Hence, we suggest that the temporal intersection of increasing NCAPE, sufficient for the development of deep convection, and higher CCN concentrations observed in the northern region (compared to the southern region) are perhaps acting to enhance convection locally via the invigoration mechanism proposed by *Rosenfeld et al.* [2008].
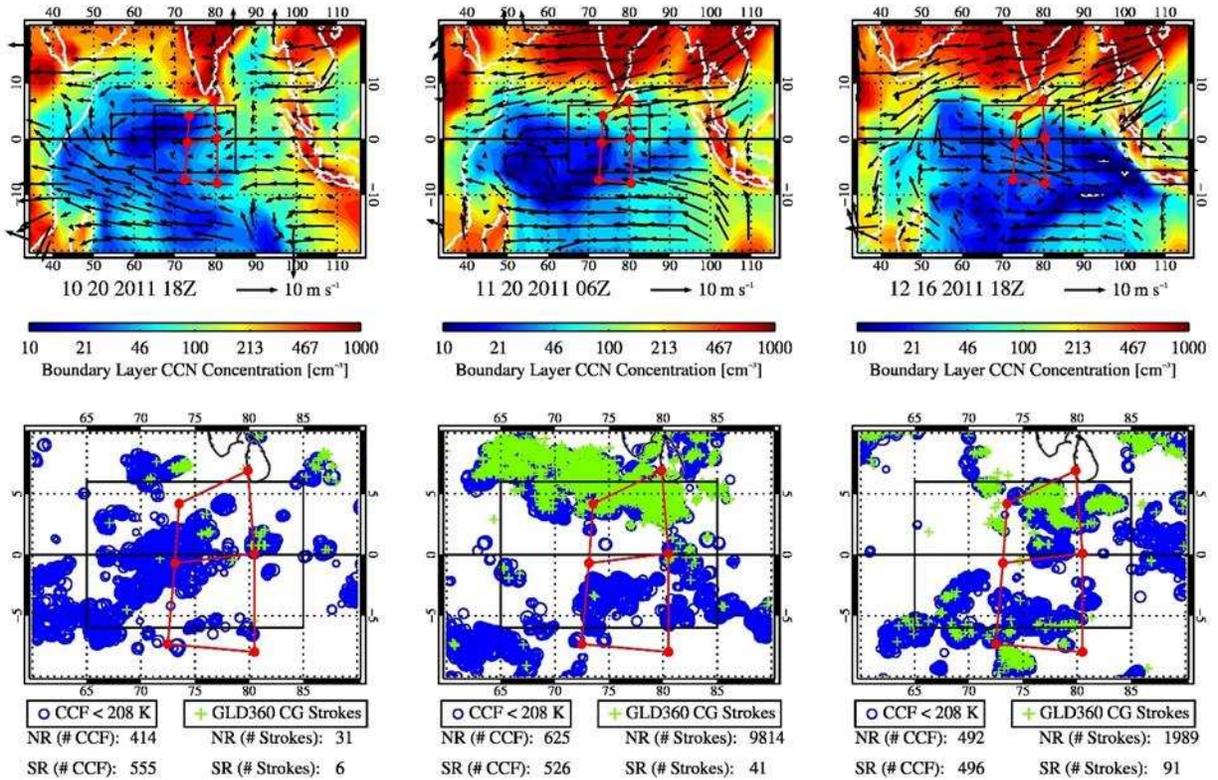
Fig. 3.4. Six-hour aggregate observations of (upper panel) boundary layer average CCN (shading; logarithmic scale) and 925 hPa wind (black vectors) versus (lower panel) cold cloud features (CCFs; blue circles, size of circle is proportional to $\log_{10}$ of the CCF area [km^2]) and CG lightning (green plus signs) over the central equatorial Indian Ocean leading up to MJO 1 (October; left), MJO 2 (November; center), and MJO 3 (December; right). The number of CCFs and CG lightning strokes have been tabulated for the northern and southern regions respectively in the lower panel. Red dots/red solid lines denote the northern and southern sounding arrays from *Johnson and Ciesielski* [2013] and black boxes denote the northern and southern regions utilized for the current study. Outlines of major coastlines are shown for reference, but note change of geographical area between the upper (20°S-20°N, 35-115°E) and lower (10°S-10°N, 60-90°E) panels.

Cecil et al., [2014] illustrated a distinct north-south gradient in total lightning flash rate

climatology over the CIO using observations from multiple satellite lightning detectors.

Therefore, this motivated us to explore the geographical dependence of the relationships

apparent in the previous time series analysis of convective characteristics, rainfall, NCAPE,

CCN, and lightning. For example, in the periods leading up to peak rainfall in MJOs 2 and 3,

similar populations of CCFs were present in each region, yet CG lightning was found to occur

almost exclusively in the northern region – where CCN concentrations were appreciably higher (Fig. 3.4).

In any given 6-hour interval, the CG lightning density in the northern region vs. the southern region differed by about an order of magnitude on average, leading us to suggest that higher CCN concentrations in the northern region are acting to invigorate convection following the mechanism discussed by *Rosenfeld et al.*, [2008]. More-active mixed-phase processes owing to enhanced CCN concentrations then act to produce stronger electrification and more lightning [*Saunders*, 1993]. Note that for the period leading up to MJO1, the difference in lightning between the northern and southern region is apparent, but the rates in both regions are small compared to observations in subsequent MJO episodes (i.e., MJO2 and MJO3). Comparison of the large-scale aerosol transport in the upper panel of Fig. 3.4 illustrates that the low-level circulation associated with the winter monsoon was not yet established prior to MJO1. In contrast, during the periods leading up to MJOs 2 and 3 northeasterly flow prevails in the northern region leading to higher CCN concentrations locally.

We consulted the TRMM database of vertical reflectivity profiles (2A25 product) for overpasses in each study region [*Iguchi et al.*, 2000; convective rain profiles with 30 dBZ echo top > 5 km] during suppressed periods (i.e., RMM phases 4-1, when CCN concentrations are generally higher than typically observed over pristine oceanic regions) to investigate potential differences in vertical precipitation structure with respect to gradients in CCN concentrations (Fig. 3.5). Each convective pixel was matched to the nearest GEOS-Chem gridpoint and the GEOS-Chem data were linearly interpolated to the time of the TRMM overpass.

On average, the difference between more-polluted (CCN > 300 cm^{-3}) and pristine reflectivity profiles (CCN < 50 cm^{-3}) maximizes in the mixed-phase region (5-10 km). The

difference in reflectivity between polluted and pristine environments north and south of the equator, respectively, are most apparent above the freezing level and range from about 0.5 up to 2.8 dB. The reflectivity differences between polluted and pristine environments are appreciable in the northern region as they range from 1.0-2.5 dB (a subtle difference in radar reflectivity is also noted in the southern region). The sensitivity of these results to the chosen aerosol stratifications was investigated and the qualitative result was found to be roughly invariant (despite limitations of decreasing sample sizes for more-polluted conditions). These results suggest that deep convective clouds in the CIO are more intense when they are subject to increased levels of pollutants transported from nearby continents.
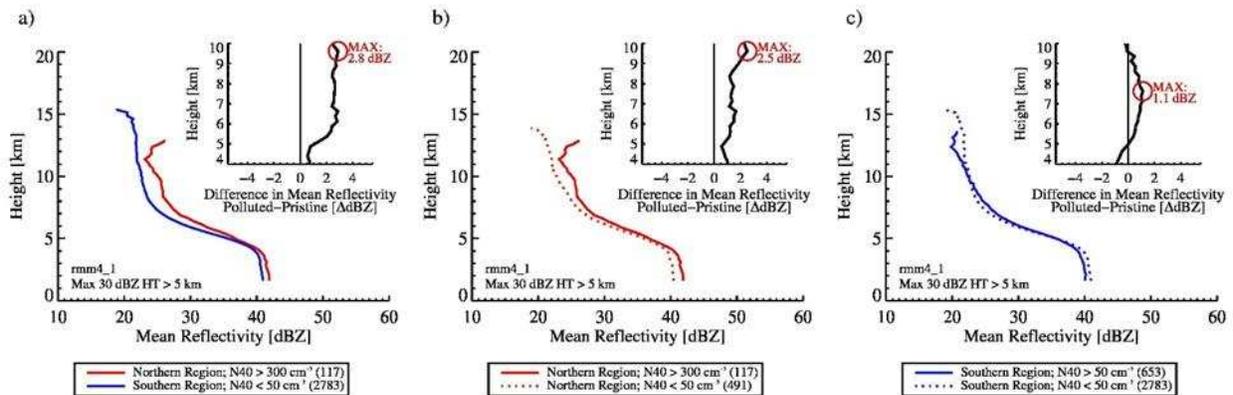
Fig. 3.5. The mean vertical profile of radar reflectivity for TRMM PR pixels with 30 dBZ echo top height greater than 5 km for a) polluted northern region vs. pristine southern region, b) polluted northern region vs. pristine northern region, and c) polluted southern region vs. pristine southern region (see legend in each panel) for real-time multivariate MJO (RMM) phases 4, 5, 6, 7, 8, and 1 where the RMM amplitude exceeded 1.0 during October-December 2011. The number of observations for each population is provided in parentheses in the legend of each panel. Differences in reflectivity in the mixed-phase region are plotted in each panel's inset; the difference is defined as the polluted reflectivity profile minus the pristine reflectivity profile in each case.

We now investigate physical mechanisms to explain the variability of CCN concentrations in the context of the three MJO events observed during DYNAMO, i.e, potential intrusions of horizontal aerosols from nearby landmasses [e.g., *Krishnamurti et al.*, 1998]. According to *Gill* [1980], for a heat source centered on the equator (i.e., heating associated with

convective precipitation during active MJO episodes), broad scale easterlies should develop along the equator while cyclonic gyres should be manifest to the north and south of the equator, west of the heat source, in an eastward-propagating Kelvin wave/westward-propagating Rossby wave response pattern. In fact, in the time following the peak of the heavy rainfall during MJOs 2 and 3 (see animations in supplemental material), cyclonic eddies embedded within the large-scale circulation are apparent in both hemispheres. The flow pattern in the northern hemisphere gyre advects CCN-rich air from primarily India, Sri Lanka, and Indonesia into the northern region of the DYNAMO domain. Meanwhile, the southern gyre draws very clean air from the southern reaches of the Indian Ocean towards the equator, thereby establishing a sharp meridional aerosol gradient prior to the next MJO cycle.

We note that CCN concentrations over southern India and Sri Lanka frequently exceed 500 cm^{-3}, such that northerly anomalies in the meridional flow leads to a dramatic increase in CCN concentrations primarily north of the equator within our study area of interest (e.g., episodes occurring 7 November, 5 December, and 15 December, 2011). Additionally, biomass burning practices throughout much of Sumatra and Kalimantan in Indonesia are known to contribute to significant increases in local concentrations of aerosols during July-November each year [*Field et al.*, 2009]. However, evidence for episodes of persistent (anomalous) easterly flow leading up to MJO onset, contributing to a meridional aerosol gradient between the northern and southern regions in the CIO, was lacking during our period of investigation.

According to *Langley DeWitt et al.* [2013] westerly wind bursts occurring in the wake of active MJO convection in DYNAMO contributed to appreciable increases in sea-salt aerosols. Although we did not quantify the contributions of individual constituent aerosol species to the total estimates of CCN in the current GEOS-Chem simulations, we do not expect that local

increases in sea-salt aerosols could account for the large overall increase in the total CCN over the CIO in the wake of active MJO convection; sea-salt mass is weighted towards larger particle sizes and contributes a proportionally smaller aerosol number per mass than anthropogenic aerosols. Rather, circulations induced by the MJO itself lead to increases in CCN north of the equator, with CCN having origins over nearby land masses.

Rosenfeld et al. [2002] argued that the introduction of a relatively small number of sea-salt particles within continental aerosol plumes advected offshore maintains a generally maritime-like population of convective clouds, where efficient warm-rain processes prevail as sea-salt aerosols activate and grow readily in the presence of a relatively large number of small droplets that result from higher-than-average CCN concentrations over ocean. Our results do not favor the *Rosenfeld et al*. [2002] hypothesis. Rather the enhanced lightning and stronger mixed-phase radar reflectivities north of the equator in more-polluted environments are more consistent with the invigoration hypothesis of *Rosenfeld et al*. [2008].

3.3 Discussion and summary of results from DYNAMO

Although the current study is limited to the three month period from October to December 2011 surrounding the DYNAMO field campaign in the CIO, the continuity of the GEOS-Chem aerosol transport model output and GLD360 lightning data has allowed for the compilation of a significant number of observations in a region that has received relatively little attention in context. Comparisons between two regions north and south of the equator were emphasized in this study and highlighted a number of important findings.

First, we have documented apparent trends in CCN concentrations that appear to be related to large-scale circulations associated with the MJO (e.g., Fig. 3.3) over the CIO. The

large-scale circulation's influence on regional differences in CCN concentrations over the CIO documented in this study (upper panel of Fig. 3.4) corroborates earlier research from INDOEX outlining the existence and seasonal variability of the large-scale hemispheric gradient in aerosol concentrations in this region. Secondly, we have presented considerable evidence that hemispheric differences in CCN concentrations impact deep convective clouds (i.e., more intense radar reflectivities aloft and more lightning) leading up to the heavy rainfall period associated with the MJO (Figs. 3.4-3.5). Thus, our results are in support of a two-way interaction between aerosols and convective clouds in the MJO – that the MJO is likely responsible for modulating regional CCN concentrations in the CIO and the impact of these CCN on subsequent cloud system development is non-negligible.

Is it possible that the implicit differences in the intensity of convection (e.g., heating and vertical mass flux) between the northern (polluted) and southern (pristine) regions could impact larger-scale circulation or propagation of the convective envelope in individual MJO episodes? Results from an independent thermodynamic analysis over the northern and southern sounding arrays observed during the DYNAMO field campaign indicate that tropospheric heating rates were greater to the north of the equator compared to south of the equator in suppressed and active periods of the MJO cycle [R. Johnson, *personal communication*]. Importantly, the large-scale circulation response to asymmetric heating about the equator [e.g., *Gill*, 1980] is hypothesized to increase the offshore transport of anthropogenic aerosols from northern land masses, in effect, leading to the persistence of the aforementioned two-way interaction.

In this chapter, a collection of satellite observations of convective clouds were compared with estimates of boundary layer CCN concentrations from the GEOS-Chem transport model throughout the period spanning October-December 2011 during the DYNAMO field campaign

over the CIO. We investigated differences in the distribution, vertical precipitation structure, and lightning characteristics of deep convective clouds in response to hemispheric differences and temporal variations of background boundary layer CCN concentrations related to the MJO evolution (while noting a general homogeneity in environmental thermodynamics between regions north and south of the equator using data from the MERRA reanalysis).

Our findings suggest that aerosols advected from nearby land masses may be responsible for invigorating convective clouds during certain phases of the MJO in the northern region of the CIO via the mechanism described by *Rosenfeld et al.*, [2008]. The results also suggest that synoptic scale, off-equatorial, cyclonic gyres develop in the wake of active MJO convection [e.g., *Gill*, 1980] over the CIO and they appear to be responsible for enhancing the background meridional aerosol gradient primarily during suppressed phases of the MJO (as the northern gyre advects continental air from southern Asia southward and the southern gyre transports pristine air from the southern Indian Ocean northward). Although rainfall and the total area of cold cloud features were found to be comparable between the northern and southern regions of interest for the three MJO episodes observed during DYNAMO, convection developing in more-polluted environments north of the equator produced up to an order of magnitude more CG lightning compared to convection in pristine environments south of the equator. Convection occurring in polluted environments also had higher radar reflectivity in the mixed-phase region (~1-3 dB differences at an altitude of approximately 5-10 km), consistent with theory for aerosol-induced convective invigoration [e.g., *Rosenfeld et al.*, 2008]. Here we emphasize that NCAPE did not vary appreciably between the northern and southern regions during DYNAMO (although thermodynamic characteristics of the environment are known to fluctuate temporally between phases of the MJO); thus, we ensure that spatial variability in the convective cloud

population with respect to changing CCN concentrations north and south of the equator occurs in a roughly homogenous thermodynamic background.

Though the relationships between aerosols and convective clouds observed were consistent for three distinct MJO episodes in our study period, it remains to be seen whether these patterns are a robust feature of MJO climatology for this region. The next section details an examination of a longer observation climatology that attempts to validate the findings from DYNAMO using a more complete sample of MJO events. Composite distributions of lower tropospheric wind, CCN, thermodynamic quantities, and convective intensity proxies relative to MJO evolution are presented next.

CHAPTER 4

Climatology of MJO-modulated aerosol variability and the response of deep convection

4.1 Background

After the inadvertent discovery of the MJO by *Madden and Julian* [1971], extensive investigations of surface and upper-air observations have documented 40-50 day periodicities in lower and upper tropospheric wind fields, temperature, and surface pressure and general broadband frequency behavior of this phenomenon [*Madden and Julian*, 1972,1994]. Coherent observations of zonal winds, temperature, and pressure between surface/upper-air stations across the equatorial Tropics lead to the establishment of the idea of (eastward-) propagating disturbances and global-scale baroclinic structures in the zonal winds associated with the MJO [*Madden and Julian*, 1972]. *Madden and Julian* [1972] identified important phase relationships between pressure, horizontal wind (i.e., divergence signatures) and temperature/moisture in the middle troposphere that prompted speculation about the possible importance of moist convective processes in addition. Early theories surrounding the existence of convectively-coupled equatorial waves with peak variance on intraseasonal time scales were substantiated by climatological studies of tropical clouds that contributed evidence of large (L ~ 2000-4000 km), eastward-propagating "super cloud clusters" that formed and subsequently intensified over the CIO [e.g., *Nakazawa*, 1988; *Wang and Rui*, 1990].

To present, the continued aggregation of observations and advances in computation capability have allowed scientists to develop new methods that aim to objectively identify convective anomalies associated with the MJO and track their propagation. Many investigators have attempted to isolate intraseasonal variability in the Tropics, i.e., variability potentially

associated with the MJO, by bandpass filtering atmospheric observables approximately in the 30-90 day spectral band prior to analysis [e.g., *Kiladis et al.*, 2005]. Subsequent steps in studies that employ spectral filtering methods can involve the computation of dominant spatial structures of variability via empirical orthogonal function (EOF) analysis. Conveniently, the EOF analysis seeks to define a collection of individually weighted independent vectors (or spatial structures) that collectively explain the variance associated with a quantity of interest [*Hannachi et al.*, 2007].

Citing early studies, *Wheeler and Hendon* [2004] noted that the leading two EOFs of bandpass-filtered single fields (i.e., outgoing longwave radiation or zonal wind) typically characterize eastward-propagating disturbances at tropical latitudes. However *Wheeler and Hendon* [2004] go on to demonstrate that the projection of global daily data onto the leading two *combined EOFs* of outgoing longwave radiation, zonal wind at 850 hPa, and zonal wind at 200 hPa enhances the signal-to-noise ratio of variance on intraseasonal timescales (compared to methods that use EOFs of single fields). Meanwhile, by their definition, it is possible to define a metric for determining the state of the MJO in real-time, i.e., the real-time multivariate MJO index (RMM index), using the principle component time series of the leading EOF modes.

Following the study of three distinct MJO events observed during DYNAMO (see Ch. 3), the aim of the next part of the analysis is to determine whether the covariability noted between aerosols, large-scale winds, and convective clouds associated with the MJO is a robust feature of climatology. Though no consensus currently exists for the optimal method to objectively define the MJO in climatology [*Straub*, 2013; *Kiladis et al.*, 2014], in the interest of investigating the climatological interaction of the large-scale flow, rainfall, CCN, and convective clouds, the

RMM index will be employed in a composite-based analysis. Data and methodology for the climatological analysis will be described next.

4.2 Data and Methodology

      The climatological analysis was based on data from multiple sources for the years 2004-2011 over the CIO domain (20°S-20°N, 35-115°E). Following the investigation presented in Ch. 3, the TRMM 3B42 rainfall product [*Huffman et al.*, 2007] is used to document large-scale rainfall variability over multiple years for this study. Note that the period of interest occurs exclusively after 2001, at which time the TRMM satellite underwent a boost maneuver to extend its observational lifetime. As a result, potential differences in spatial resolution of the TRMM PR footprint and sensitivity across multiple years are avoided. Data for the 3B42 product are available at 0, 6, 12, and 18Z at 0.25° resolution. Note that rainfall rate in the 3B42 product is representative of a 3-hourly average; a constant scaling factor was applied to the data to estimate an equivalent 6-hourly rainfall rate.

      The horizontal zonal ($\vec{u}$) and meridional ($\vec{v}$) components of the wind, relative humidity, and temperature at 0, 6, 12, and 18Z at a horizontal resolution of 1.25° (over a portion of the TRMM domain spanning the CIO; 38°S-38°N,35-115°E) in both the latitudinal and longitudinal dimensions for 25 pressure levels between 1000 and 100 hPa (at 25 hPa increments for 1000-700 hPa and 50 hPa incremements for 700-100 hPa) were obtained from the Modern Era Retrospective Analysis for Research and Applications (MERRA; via http://gmao.gsfc.nasa.gov/merra/). The vector average of the horizontal components of the wind was computed between 1000 and 850 hPa in order to obtain an estimate of the boundary layer average flow speed and direction. Surface-based CAPE (and NCAPE) was computed at all

gridpoints for each of the four model time steps over the domain of interest using the vertical

profiles of temperature and moisture. Likewise, warm cloud depth (assuming a surface-based

parcel) was computed for all model time steps at all gridpoints using the method described in

Sec. 2.2 above, following *Iribarne and Godson* [1981].

Note that quality-controlled global lightning data from GLD360 (Vaisala, Inc.) were

available only for the intensive observation periods during the DYNAMO field campaign

(October-December 2011). In order to obtain a climatological perspective of the lightning data

component, the current investigation appeals to the longer climatology of total lightning

provided by the TRMM lightning imaging sensor (LIS). Previous studies have shown large

regional differences in the ratio of intracloud to cloud-to-ground lightning flashes (i.e., the IC to

CG ratio) [*Boccippio et al.*, 2001] such that lightning characteristics observed by global very low

frequency (VLF) lightning networks (e.g., GLD360), which predominantly observe cloud-to-

ground discharges, may differ appreciably from lightning characteristics observed by space-

borne total lightning imagers or ground-based very high frequency (VHF) networks, which

capture the sum of intracloud and cloud-to-ground flashes. However, it is of interest to

document relative changes in lightning behavior between regions of the CIO along with the

temporal evolution of lightning during different phases of the MJO for comparison with the

earlier results from DYNAMO. Therefore, the TRMM LIS dataset adequately satisfies the

requirement for a climatological lightning dataset in the current study.

Vertical profiles of radar reflectivity (i.e., VPRR) were again taken from the TRMM

2A25 attenuation-corrected reflectivity profile product. Analogous to the methodology

presented in Sec. 3.2, convective profiles (determined using the 2A23 'raintype' parameter) with

30 dBZ reaching altitudes greater than 5 km (i.e., MAXHT30 > 5 km) were studied. To reiterate,

the purpose of this study is to investigate the variability of deep convective clouds with respect to environmental changes (i.e., in thermodynamics and aerosols) potentially occurring during the various phases of the MJO observed over multiple years over the CIO. Restricting the study of VPRR to those PR pixels with MAXHT30 > 5 km ensures that the sample of radar reflectivity structures is most closely associated with congestus or cumulonimbus convective modes, as it is unlikely that reflectivity from growing cumuli or predominantly warm-phase clouds achieves the aforementioned reflectivity altitude threshold.
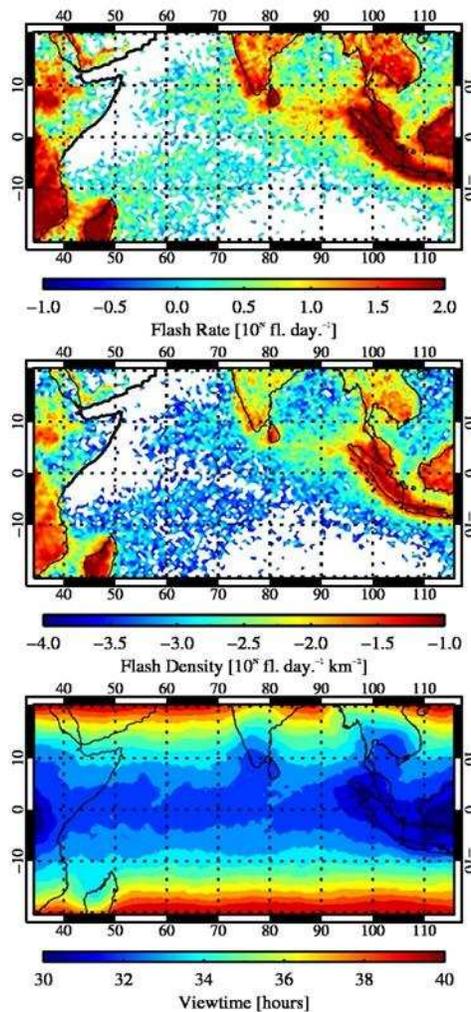
Fig. 4.1. Climatological distributions of total daily lightning flash rate (top), total daily lightning flash density (middle), and sensor viewtime as observed by the TRMM Lightning Imaging Sensor (LIS) for cold seasons in the years 2004-2011 over the CIO domain (20°S-20°N, 35-115°E).

107

Output from the GEOS-Chem transport model with the online TwO-Moment Aerosol Sectional microphysics module [TOMAS; *Adams and Seinfeld*, 2002] is used to define a proxy for the concentration of cloud condensation nuclei (CCN) according to the number of aerosols with diameters larger than or equal to 40 nm (N40). Data from the model were output at a horizontal resolution of 2.0 latitude by 2.5 longitude, four times daily (0, 6, 12, and 18Z), and at ten levels between 1000 and 850 hPa. The values of N40 at each gridpoint were averaged vertically across all levels to define the boundary layer average CCN proxy studied here [e.g., *Dusek et al.*, 2006]. Importantly, the wind fields in the GEOS-Chem model and MERRA reanalysis are derived from a similar source and therefore using these two separate data sources to characterize potential variability in boundary layer CCN concentrations due to advection is appropriate.

Data from both the MERRA reanalysis and the GEOS-Chem transport model are available in a native gridded data format at each output time, allowing for simple interpretation in the 6-hourly or daily-average analysis setting. In contrast, data from the TRMM satellite is available in an orbit-level granule (each orbit at 35° inclination lasts approximately 91 minutes), with sixteen files available each day of the year. The TRMM PR pixel data are geospatially tagged and time referenced, such that cross-referencing with other data sources (i.e., regularly gridded data) is trivial. To document the potential influence of regional variations in CCN on vertical reflectivity structure, each deep convective pixel (identified using the methodology described above) was matched to the nearest neighboring GEOS-Chem grid point before interpolating the chemical transport model output to the time of the TRMM satellite overpass.

The TRMM LIS instrument documents the location, radiant intensity, areal extent, and error flags for individual optical emission events that are then grouped into flashes. Lightning

flashes from each orbit were subjected to a quality-control filtering routine whereby a flash was eliminated from the daily count if the associated one-second data had any indication of fatal platform, instrument, or processing errors (see TRMM LIS documentation, available http://thunder.nsstc.nasa.gov/data/data_lis.html). The native resolution for the LIS observation strategy is 0.5° in latitude/longitude coordinates across all longitudes between 38°S and 38°N (the field of view of the LIS instrument is approximately 600 x 600 km, thus the field of view extends ~300 km beyond the maximum orbital latitude of the TRMM satellite). The collection of quality-controlled flashes in each orbit was then gridded at the native grid resolution of 0.5° and the counts were summed over each day.

The estimated amount of time that a 0.5° gridbox is in the field of view of the LIS instrument, i.e., the effective observation time, is termed the "viewtime". On average, the viewtime for individual gridboxes is approximately 90 s, though this value ranges from roughly 0-105 s in a single overpass. All orbits for each day were queried during the eight-year period and the total daily viewtime was computed by summing over all overpasses in a given day (typically 2-4 overpasses for any given gridbox). Note that the product of the number of overpasses and the average effective observation for each overpass accounts for a small fraction of the time elapsed in one day (roughly 3-7 minutes). Flash rate (flash rate density) was then computed by taking the quotient of the daily flash count and the daily total viewtime (and then normalizing by the area in each gridbox while accounting for the change in gridbox area with increasing latitude). Total flash rate and total flash rate density statistics and estimates of the total viewtime for the global TRMM domain are shown in Fig. 4.1; despite the fact that this study encompasses less than half of the TRMM lifetime, the quantitative estimates of lightning flash rate (top), flash density (middle) and sensor viewtime (bottom) over a given location

compare well with other climatological studies of the TRMM LIS and Optical Transient Detector

datasets [e.g., *Christian et al.*, 2003; *Cecil et al.*, 2014].
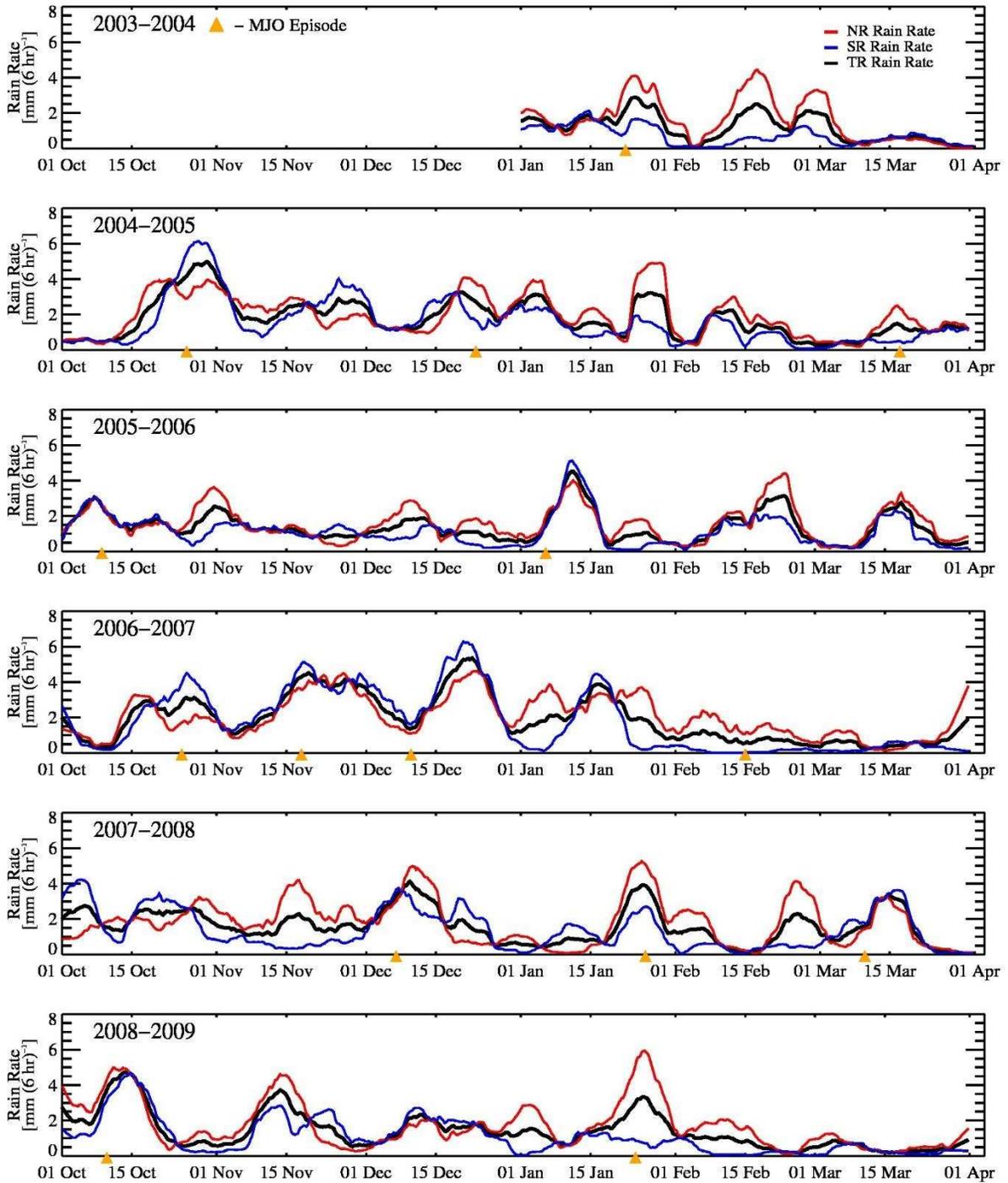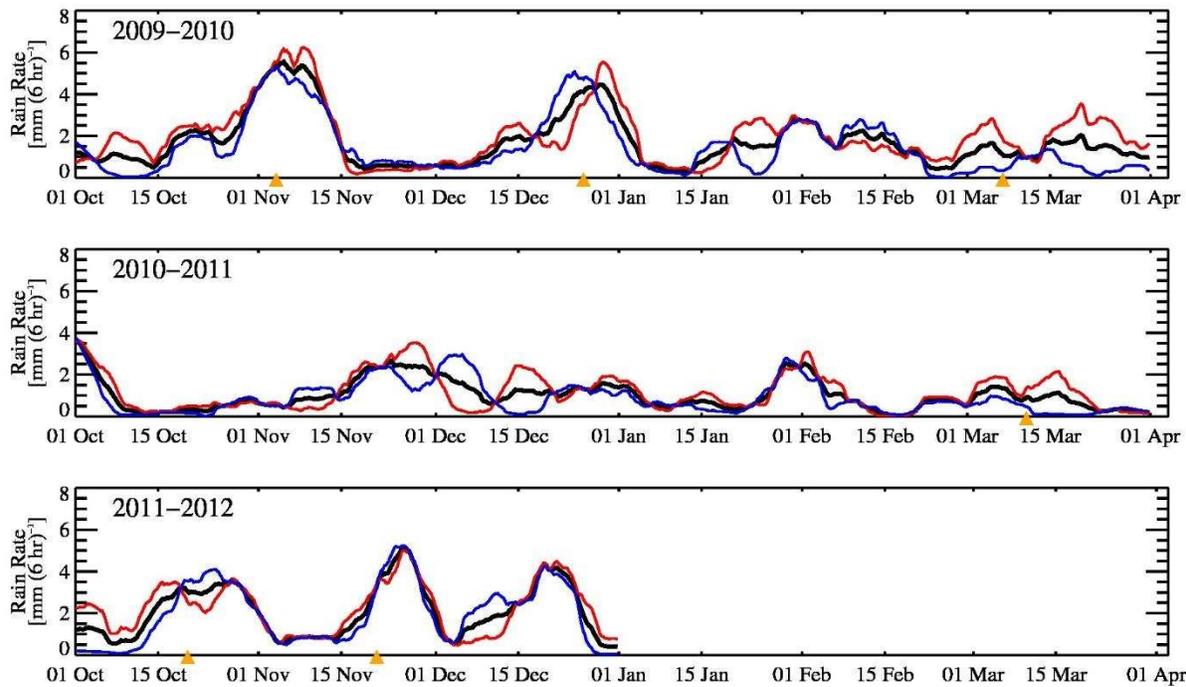
Fig. 4.2. Areal-average TRMM 3B42 rainfall per 6 hours over the northern region (red solid), southern region (blue solid), and total region (northern + southern regions; black solid) for nine cold seasons between January 2004 and December 2011. Rainfall time series have been smoothed using a 29-point (~1 week) boxcar moving average. The onset of RMM phase 2 for identified MJO episodes is shown along the abscissa (gold triangle).

The boreal cold season that spans time between October and March is coincident with the strongest MJO-like variability in the equatorial region of the Indian and West Pacific Ocean basins [*Madden*, 1986; *Zhang and Dong*, 2004]. The principle components of the leading two combined EOFs of outgoing longwave radiation, zonal wind at 850 hPa, and zonal wind at 200 hPa, i.e., RMM_1 and RMM_2 according to *Wheeler and Hendon* [2004], were used to assign an RMM phase (phases 1-8) to each day in the months January, February, March, October, November, and December for the years 2004-2011 (time series of RMM indices are freely available via http://cawcr.gov.au/staff/mwheeler/maproom/RMM/). As a continuation of the analysis in section 3, the two regions that were defined for the DYNAMO-centric analysis, i.e., the northern and southern regions, (see Fig. 3.2), are again used for assessing possible regional differences in the context of the MJO climatology. RMM phase composites were created by

111

averaging quantities of interest during each phase (for RMM phases 1-8) within the northern and

southern regions on each day in the cold season months where the combined RMM amplitude

(i.e., $RMM_1{}^2 + RMM_2{}^2$) exceeded a value of 1.0, as is common in practice [e.g., *Kiladis et al.*,

2014].

Rainfall and lightning were scaled by factors of 6.0 (dimensionless) and 86,400 s day^{-1},

respectively, prior to taking the unconditional mean (including values of values of 0 mm hr^{-1} or 0

fl. day^{-1}, where they existed within each study region) across the northern and southern regions

each day. The computation of the areal mean NCAPE and WCD for surface-based parcels was

conditional on the presence of non-zero NCAPE at a given reanalysis gridpoint for a given time

step. Values in the chemical transport model were generally non-zero at all gridpoints for all

time steps, so the areal mean computation for CCN was done without regard for potential

occurrences of zero values.

4.3 Results

4.3.1 RMM phase dependence

For the eight-year period between 2004-2011, individual MJO episodes were identified

by cataloging periods where the RMM phase index 1) was in phases 1-3 for a minimum of 10

days; 2) progressed from phase 6 through the peak active phases 2 and 3 (for the CIO) and then

onto suppressed phases 4 and 5; and 3) had an amplitude of 1.0 or greater for at least half of the

days in the cycle. Time series of the areal-average rain rates from the TRMM 3B42 product are

shown for the northern, southern, and total (northern region + southern region) regions in Fig.

4.2 for the boreal cold season, with the onset of phase 2 in each identified MJO episode

highlighted along the abscissa. In general, the onset of RMM phase 2 relative to the peaks in the

rainfall time series varies between episodes and the rainfall observed during MJO episodes can be rather asymmetric between the northern and southern regions. For several of the identified MJO events, the rainfall amounts do not appear to differ substantially from background quiescent periods (i.e., 15 February 2007, 7 March 2010, and 13 March 2011); *Kiladis et al.* [2014] noted that zonal wind signatures can strongly project onto the leading two combined EOF modes of equatorial variability identified by *Wheeler and Hendon* [2004], leading to an inflated RMM index amplitude, even in the absence of large-scale negative outgoing longwave radiation (i.e., convective) anomalies. Nonetheless, more than 20 individual events are identified using the RMM indexing method and thus there is high confidence that a sufficient sample of independent MJO episodes has been observed during eight-year period.

As a proof-of-concept, *Wheeler and Hendon* [2004] composited outgoing longwave radiation anomalies as well as upper and lower tropospheric winds across RMM phases 1-8 and found a large-scale $O(L{\sim}4000\text{ km})$ peak negative anomaly $(< 30\text{ W m}^{-2})$ that developed over the CIO near the equator during RMM phases 1-3 and subsequently decayed during eastward-propagation over the Maritime Continent and into the western Pacific Ocean. *Wheeler and Hendon* [2004] noted large-scale flow behavior consistent with previous diagnostic models for diabatic heat sources centered on the equator [e.g., *Gill*, 1980]. Following their results, at a given location in the CIO, the passage of an MJO episode is marked by the evolution from low-level easterly flow, isolated, deep convection and heavy rainfall, to the development of widespread, stratiform light precipitation under an expansive anvil shield, to low-level westerly flow and lower precipitation amounts. Here, the question is whether the phase composite evolution of proxies for deep convection are linked to variations in thermodynamics or aerosols

or some combination of the two, in a manner consistent with what was observed during

DYNAMO.
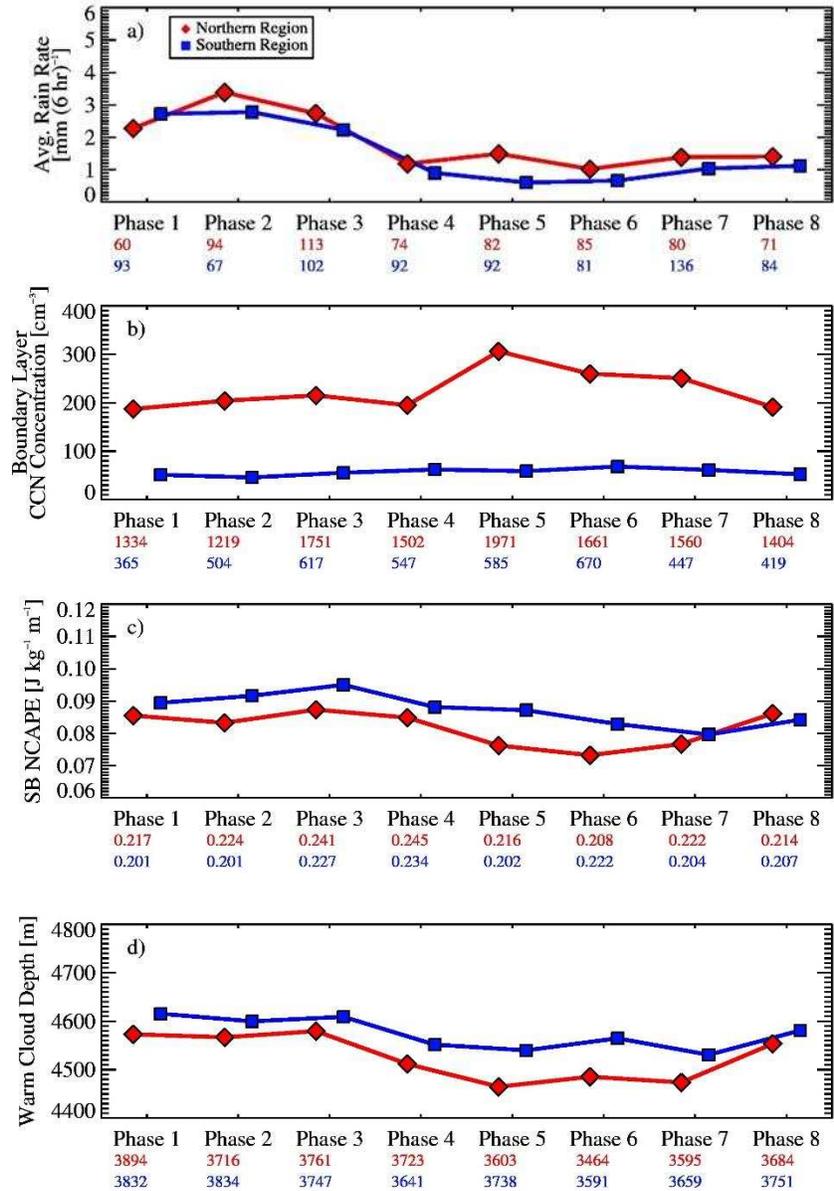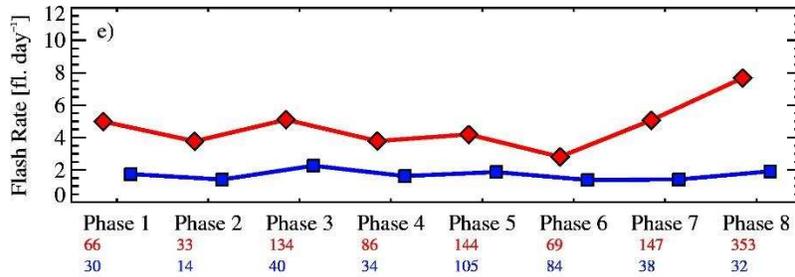
Fig. 4.3. Composite means a) TRMM 3B42 rainfall, b) CCN, c) surface-based NCAPE, d) WCD, and e) lightning flash rate for RMM phases 1-8 (x-axis) over the northern region (red solid/diamonds) and the southern region (blue solid/squares) for cold seasons during the years 2004-2011. The maximum value of each parameter anywhere within each study region during each RMM phase is shown by the numbers below each panel (in red for the northern region, in blue for the southern region) to communicate an estimate of the variance of each parameter.

RMM composites of areal-average rainfall, CCN, NCAPE, WCD, and lightning flash rates for the northern and southern regions over all MJO episodes observed for the years 2004-2011 are shown in Fig. 4.3. The phase-to-phase evolution of areal-average rainfall is similar between the northern and southern regions as daily rainfall peaks in RMM phase 2 over the CIO and is at a minimum throughout RMM phases 4-6 (Fig. 4.3a). The equivalent daily rainfall totals are comparable to those which were observed over the finer sounding array regions during DYNAMO [e.g., *Xu and Rutledge* 2014, 2015]. However, the rainfall rate in the northern region is larger than the rainfall rate in the southern region by 20-30% on average in individual RMM phases.

Consistent with observations from the smaller subset of MJO episodes from DYNAMO, CCN concentrations are greater by about a factor of four in the northern region compared to the southern region on average for the collection of MJO events identified here (Fig. 4.3b). In RMM phase 5, CCN concentrations in the northern region abruptly increase by more than 100 cm^{-3} on average while the CCN concentrations in the southern region increase only slightly (increases of 20-30 cm^{-3} are observed) and never exceed 100 cm^{-3}. At this point, the disparity in CCN concentrations between regions increases to a factor of six before the concentrations in the

115

northern region gradually decrease during RMM phases 6-8. Note that the absolute maximum CCN concentrations observed in both regions differ by an order of magnitude, as the northern region shows values in excess of 10^3 cm^{-3} while the southern region only reaches values of 10^2 cm^{-3}.

From the thermodynamic perspective, NCAPE and WCD between the northern region and the southern region appear are very similar (in contrast to the disparity in CCN concentrations between regions cited above). The evolution of surface-based NCAPE depicts a relative maximum in RMM phase 3 and relative minima in RMM phases 6 and 7 for the northern and southern regions, respectively, while the difference in NCAPE between regions rarely exceeds 0.02 J kg^{-1} m^{-1}. Aside, the magnitude of the conditional instability over both regions (CAPE ~ 800-1200 J kg^{-1}) nearly matches the observations from comprehensive studies of in-situ, upper-air observations over the sounding arrays during DYNAMO [cf., Table 6 in *Ciesielski et al.*, 2014]. The evolution of WCD by MJO phase depicts generally high values during periods of active convection in RMM phases 2-3 and slightly shallower values during the suppressed phases. Note however, that WCD values in both the northern and southern regions are in excess of the global median (WCD > 4200 m) and are among the deepest observed anywhere in the world [*Williams et al.*, 2005]. Overall, the phase composite behavior of NCAPE and WCD contributes to the ideal of homogeneous thermodynamics between regions of the CIO.

Fig 4.3e shows the phase-to-phase evolution of the daily total lightning flash rates in the northern and southern regions; flash rates (computed using an independent data set from the observations in DYNAMO) are systematically higher in the northern region compared to the southern region and the difference ranges from a factor of approximately 2 (in RMM phase 6) to 4 (in RMM phase 8). The evolution of total flash rate in the northern region is periodic with the

most frequent lightning occurring in RMM phase 8, prior to MJO onset over the CIO; there is little evidence of similar temporal behavior in the southern region. Lastly, the maximum daily flash rates in the northern region are often more than double those observed for the southern region (e.g., RMM phases 1, 2, and 4). Furthermore, the maximum areal-average daily flash rate in the northern region is more than ten times that in the southern region for RMM phase 8, when the areal-average values are maximized in both regions.
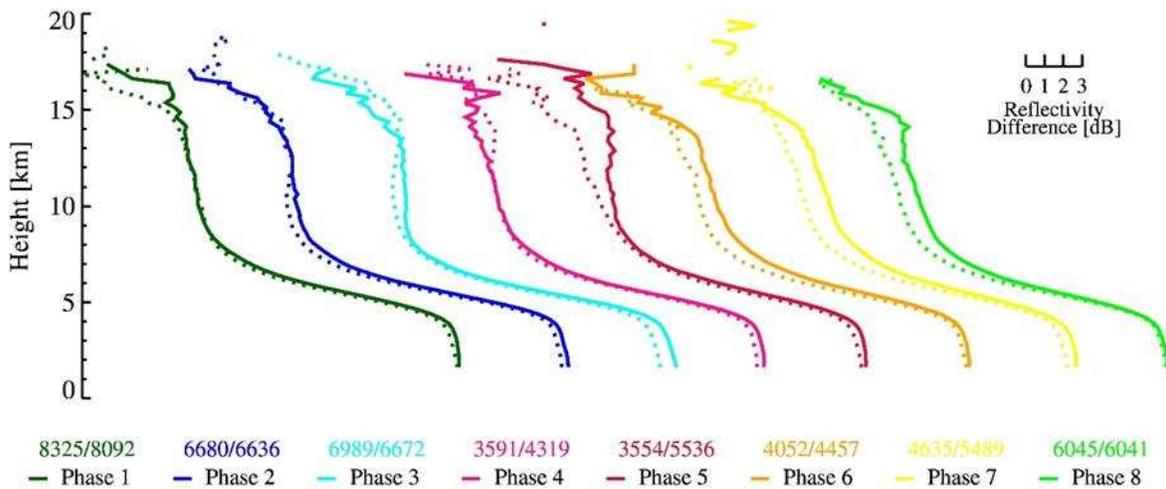
Fig. 4.4. Composite mean VPRR for RMM phases 1-8 (arbitrary x-axis) for the northern region (solid) and the southern region (dotted) for cold seasons during the years 2004-2011. A reflectivity scale to assess relative differences has been included in the upper right. RMM phases are differentiated by color and the number of vertical profiles in each region for a given phase is provided (northern region/southern region).

Fig. 4.4 shows the phase composites for VPRR in the northern and southern regions for the same period. The maximum differences in reflectivity are shown to occur in the mixed-phase region with the VPRR from the northern region (solid lines) exhibiting values of up to 2 dB greater at 5-10 km altitude compared to the southern region (dotted lines), primarily in RMM phases 6-8. Smaller reflectivity differences between the northern and southern regions are also noted throughout active convective phases, but the magnitude of the reflectivity in the mixed-phase region for deep convective pixels observed north of the equator is consistently greater than reflectivity at these altitudes in the southern region. In line with observations of typical maritime

deep convective clouds, reflectivity is found to decrease rapidly above the freezing level [e.g.,

Szoke et al., 1986; *Zipser and Lutz*, 1994], primarily during RMM phases 2 and 3, consistent

with the relative dearth of lightning observed in the CIO compared to other regions globally (the

equivalent annual flash rate densities range from about 1-5 fl. yr^{-1} km^{-2}, compare to *Cecil et al.*

[2014]). Small increases in near-surface reflectivity in the northern region relative to the

southern region are perhaps indicative of more-robust ice-based microphysical processes

translating to stronger rain near the surface [e.g., *Stolz et al.*, 2014].

Despite the apparent differences in the magnitude and evolution of environmental

(primarily CCN concentrations) and convective intensity proxies between the northern and

southern regions, there are still open questions. Given the areal-average approach, it is logical to

question whether greater lightning flash rates observed in the northern region compared to the

southern region result simply from differences in the frequency of deep convective clouds

between regions. However, the numbers of deep convective pixels observed by the TRMM PR

in each study region (shown for each RMM phase at the bottom of Fig. 4.4) over the eight-year

cold season climatology are very nearly the same for each RMM phase, suggesting that the

frequency of deep convective clouds and/or aggregate cold cloud area in both the northern and

the southern regions are approximately equal, as shown in observations from DYNAMO.

Hence, climatological observations of convective intensity proxies are consistent with the notion

that stronger convection is found systematically in the northern region compared to the southern

region. Are spatial and temporal variations of CCN, low-level wind, NCAPE, and WCD

consistent with the aforementioned result in mind?

In a study of global aerosol distributions, *Anderson et al.* [2003] found that point

observations of aerosol quantities were highly correlated over distances less than 400 km

(r~0.80). Thus to a good approximation, aerosol quantities (i.e., CCN concentrations) are roughly homogeneous on the mesoscale, while appreciable differences are likely to exist for larger spatial scales. Since the areal-averages of rainfall, CCN, NCAPE, WCD, and lightning quantities encompass roughly $O(10^6)$ km^2 in both the northern and southern regions, significant local variations could exist within each area of interest. Until now, an appreciable amount of evidence in favor of aerosol indirect effects on deep convective clouds in isolation from thermodynamics has been presented. Thus, it is possible that mesoscale variations in CCN concentrations may be associated with variations in convective intensity proxies on commensurate scales. To investigate the potential influence of mesoscale variability on the phase-to-phase co-evolution of environmental characteristics and convective parameters in the RMM phase composite framework, the geographical distributions of the mean values at individual gridpoints were computed for each RMM phase. These results are presented next.

4.3.2 Geographical composite variability

Figs. 4.5-4.6 show the mean geographical distribution of CCN concentrations for each of the eight RMM phases, with the anomalies of the horizontal wind components at 925 hPa (defined relative to the mean over all days in the cold seasons for 2004-2011) overlaid. The Student's t-test (assuming two-tailed probability) was used to assess whether the anomalous horizontal winds in each RMM phase were significantly different from the background climatology. The daily wind anomalies were assumed to be independent (i.e., the autocorrelation was sufficiently small) so that the number of degrees of freedom equaled the number of days that each phase was classified, less one. Anomalous winds that were statistically significant at the 95% level have been highlighted in gold.

In each RMM phase, a meridional gradient in CCN concentrations is apparent between the northern and southern regions, but intraseasonal modulation is also evident. In RMM phases 2 and 3, when peak convection occurs over the CIO according to RMM phase diagnostics, CCN concentrations are generally at a minimum value with widespread estimates of 100-200 cm^{-3} in the northern region and ~50 cm^{-3} in the southern region. Meanwhile, in the suppressed periods (RMM phases 4-5) an equatorward expansion of the area of polluted continental air is clearly evident, with CCN concentrations exceeding more than 500 cm^{-3} in the northern region and CCN concentrations of 100 cm^{-3} extending just south of the equator in CIO.

ANOMALY, u′
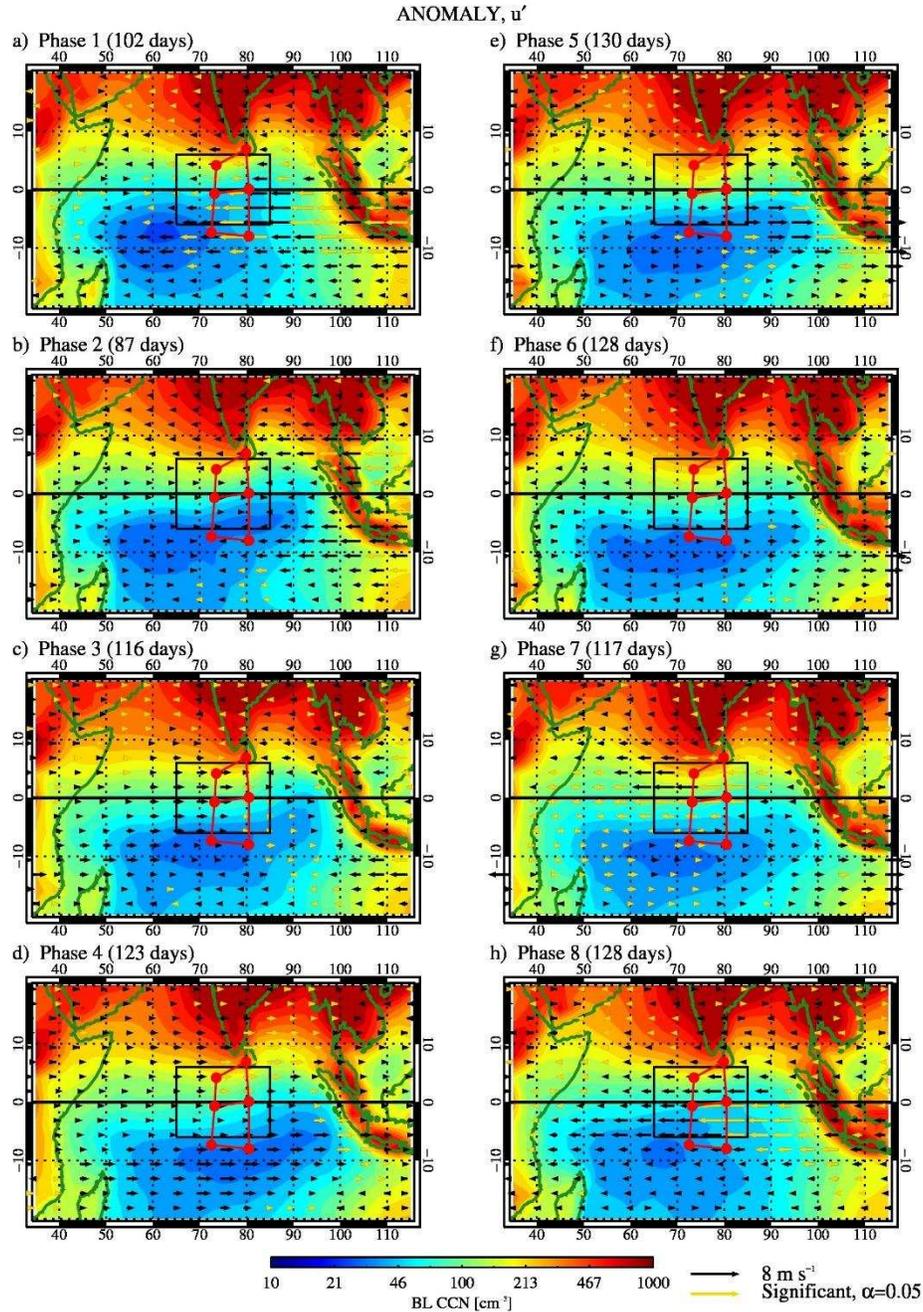
Fig. 4.5. Composite mean boundary layer CCN concentrations (shaded; logarithmic scale) as a function of RMM phases 1-8 (the number of days identified for each phase is included above each panel) for cold seasons during the years 2004-2011 over the CIO domain (20°S-20°N, 35-115°E). Anomalies in the zonal wind component at a given point have been overlaid, scaled by the 8 m s⁻¹ solid, black vector in the lower right. Wind anomalies that are statistically significant at the 95% level (compared to climatology at each gridpoint) have been colored gold. Outlines of major continents are shown by solid dark green lines. The northern and southern study regions are shown by black solid boxes and the northern and southern sounding arrays identified by *Johnson and Ciesielski* [2013] are shown by red solid lines with dotted vertices.

121
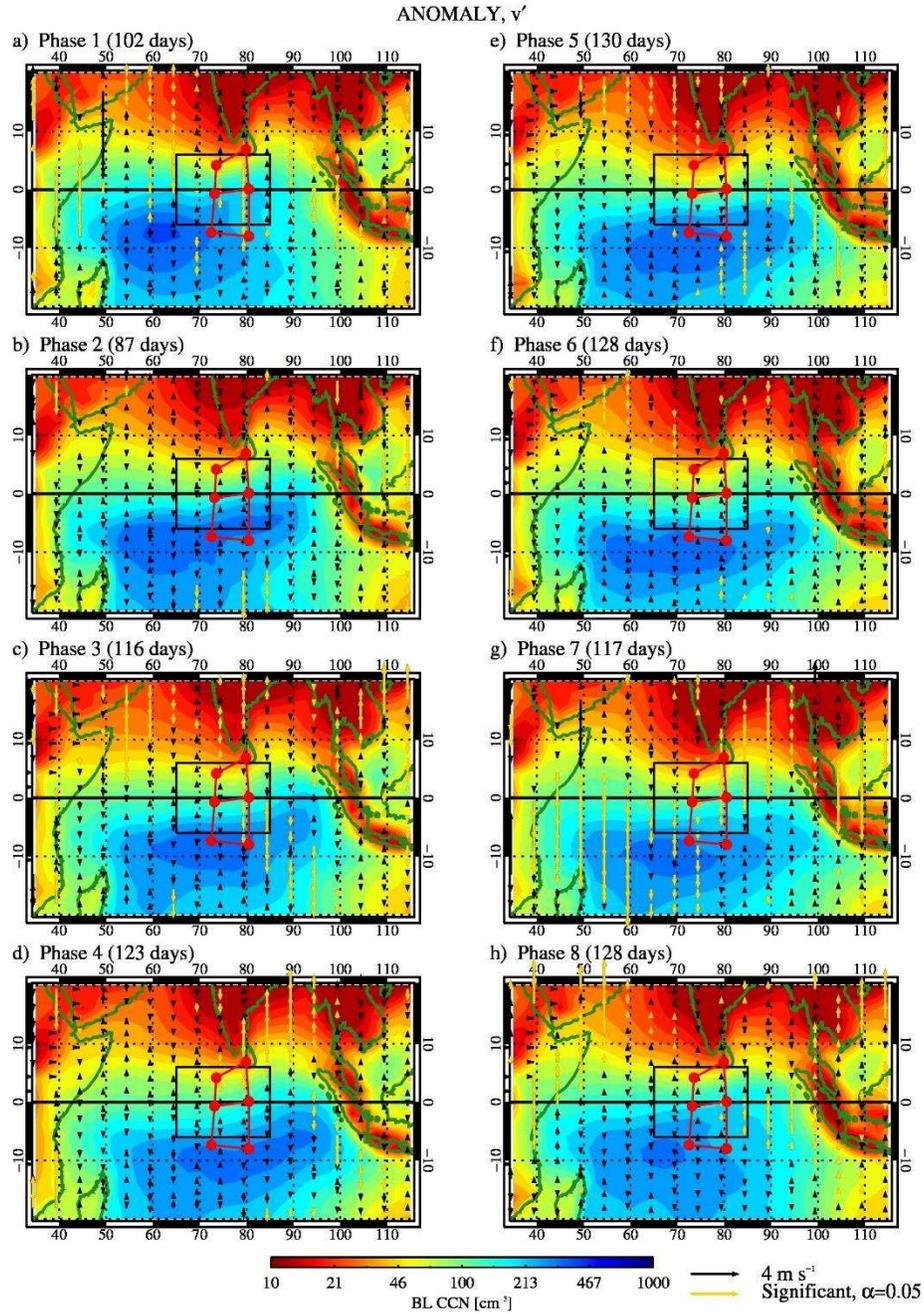
Fig. 4.6. As in Fig. 4.5, but anomalies of the meridional wind component are overlaid. Note the reference vector scale has been reduced by 50%, as shown in the lower right (4 m s^{-1} reference vector).
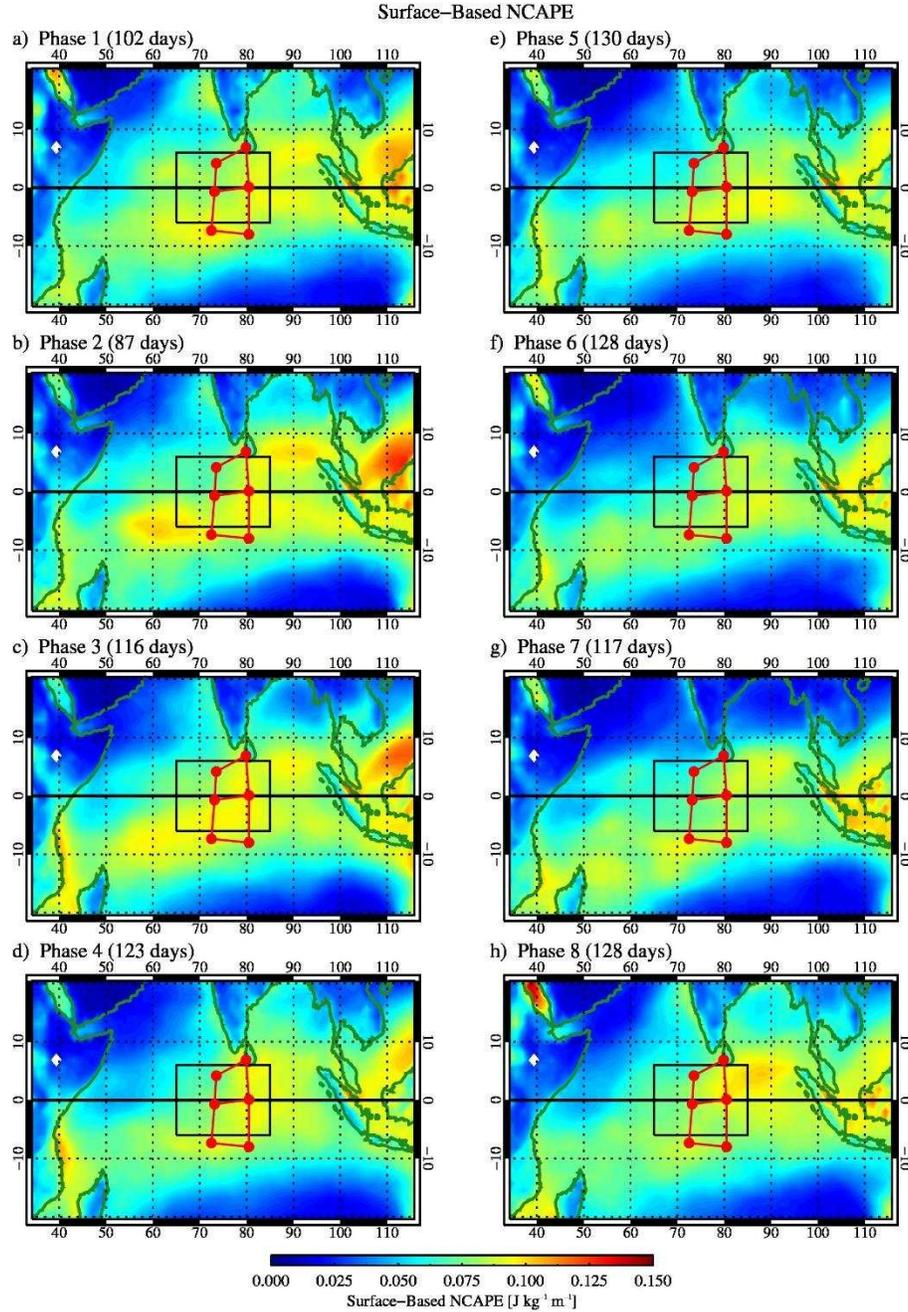
Fig. 4.7. Composite mean surface-based NCAPE (shaded; computed using the MERRA reanalysis) as a function of RMM phases 1-8 (the number of days identified for each phase is included above each panel) for cold seasons during the years 2004-2011 over the CIO domain (20°S-20°N, 35-115°E). The northern and southern study regions are shown by black solid boxes and the northern and southern sounding arrays identified by *Johnson and Ciesielski* [2013] are shown by red solid lines with dotted vertices.

123

Warm Cloud Depth
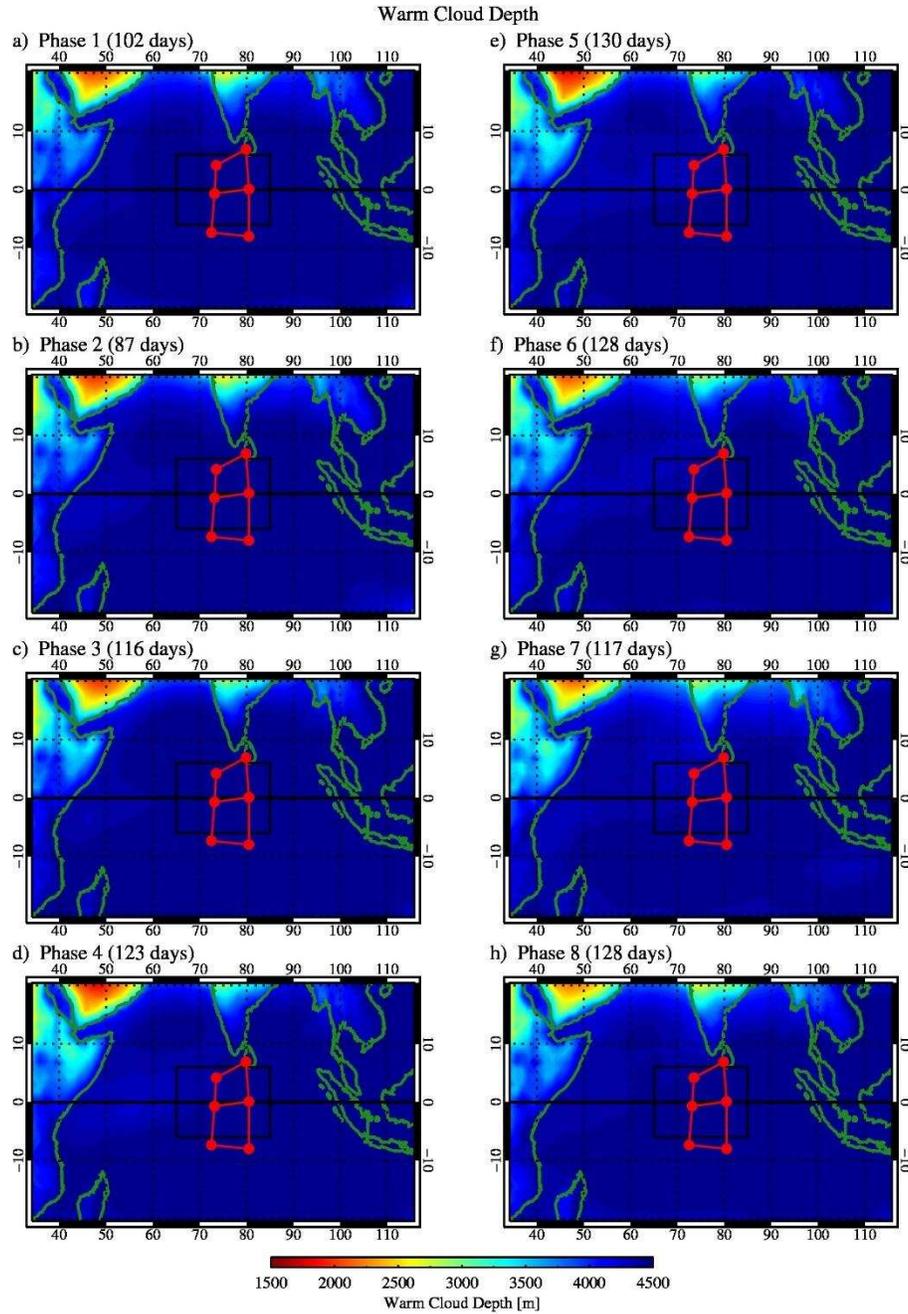
Fig. 4.8. As in Fig. 4.7, but for WCD computed according to the methods in *Iribarne and Godson* [1981] (using MERRA reanalysis).

LIS AREAL AVERAGE FLASH RATE
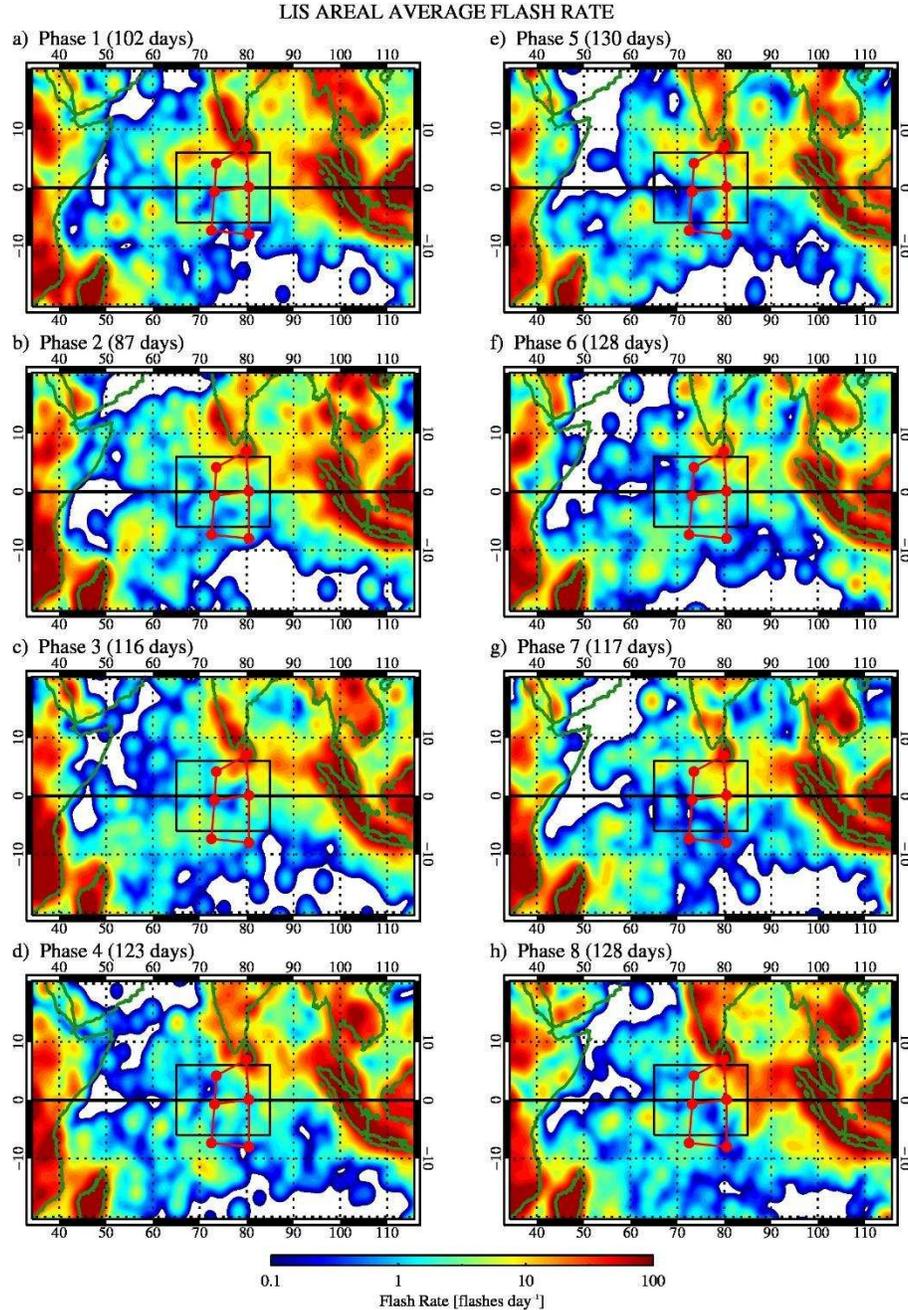
Fig. 4.9. As in Fig. 4.7, but for total lightning flash rate observed by TRMM LIS. Note that the flash rate has been plotted according to a logarithmic scale and the data were subsequently smoothed using a 7.5° Gaussian kernel in both longitude and latitude.

Zonal wind anomalies along the equator between both the northern and the southern

study regions are among the strongest observed anywhere throughout the CIO domain,

regardless of the RMM phase. The peak negative zonal wind anomalies (~ -4 m s^{-1}) are observed

in RMM phases 8 and 1, possibly indicative of a Kelvin wave response ahead of the main

125

convective envelope. Positive zonal wind anomalies on the order of 2 m s^{-1} are noted in RMM phases 3-5 along the equator, consistent with westerly wind burst episodes that are commonly observed in the wake of active MJO phases in the CIO. In the suppressed phases (especially RMM phases 4-5), equatorward meridional wind anomalies in both the northern and southern regions are noted over parts of the CIO.

In fact, inspection of the meridional wind anomalies in RMM phase 5 reveals southerly anomalies in the Bay of Bengal and northerly anomalies further west over the Arabian Sea, southern India, and Sri Lanka. In the southern hemisphere, northerly anomalies are present just west of Sumatra, Indonesia and southerly anomalies are observed near 75-85°E, generally poleward of 10°S latitude (the meridional wind speed anomalies are approximately 0.5-2 m s^{-1} in magnitude in both hemispheres). Importantly, the northerly anomalies to the south and west of India contribute to a net strengthening of the northerly component in the monsoon flow. In the presence of the background meridional CCN concentration gradient, stronger equatorward advection of continental aerosols is implied; the magnitude of the meridional CCN concentration gradient appears to be strongest during RMM phase 5 and this behavior is reflected in the areal-average RMM composites shown in Sec. 4.3.1. Note that in both hemispheres, the structure of the anomalous meridional winds is consistent with the presence of cyclonic eddies within the larger scale circulation in the wake of active convective episodes, perhaps evidence of off-equatorial Rossby wave activity noted by *Gill* [1980], *Yamagata and Hayashi* [1984], and more recently by *Johnson and Ciesielski* [2013].

The composites of areal-averaged NCAPE and WCD were very similar between the northern and southern regions (shown in Figs. 4.7-4.8), in support of hypothetical spatial similarity in thermodynamics across both study regions in the CIO. However, anomalous

northerly flows and equatorward excursions of continental CCN could also indicate cooling and drying in the low levels associated with periodic outbreaks of continental air that occur during the boreal cold season (not shown). In general the geographical distributions of NCAPE and WCD illustrate meridional homogeneity, with the maximum (minimum) surface-based instability generally found during RMM phase 3 (RMM phase 6). However, longitudinal differences within both the southern and especially the northern regions are clearly apparent during some phases. In the areal-average RMM phase composite view, NCAPE was found to be slightly smaller in the northern region compared to the southern region throughout much of the suppressed period (RMM phases 4-6) over the CIO. From the geographical distributions, erosion of surface-based instability is apparent in the northwestern reaches of both the northern and southern regions. In Fig. 4.8, WCD is generally greater than 4200 m over both the northern and southern regions in the CIO, but there was evidence that the surface dew-point depressions were larger on average in the northern region as WCD was slightly shallower there (accounting for a difference in WCD of roughly only 100-200 m).

For completeness, the composite geographical distributions of daily total lightning flash rates are presented for each RMM Phase. Following the methods outlined above, composites of the total lightning flash rate [flashes day^{-1}] were computed by cataloging all flashes observed during the aggregate time that each 0.5° gridbox was within the TRMM satellite's field of view during each RMM phase. The resulting distributions were generally noisy, so a Gaussian kernel (7.5-degree width) was used to smooth each composite field in the latitude and longitude dimensions. Immediately apparent is the overwhelming preponderance of lightning over land areas adjacent to the oceanic study regions. There is a subtle intraseasonal modulation (cf., Figs. 4.9a,f, and h) in the lightning signal as lightning flash rate is highest over the CIO in RMM

127

phases 7, 8, and 1 and high lightning rates extend further offshore south of India, Sri Lanka, and west of Sumatra during these times.

In contrast, high lightning rates are less apparent in the southern region during all RMM phases. The southern region is more than 600 km from the nearest appreciable landmass and therefore direct sources of CCN transport are not clear, however convection in the inter-tropical convergence zone, resulting from overturning in the Hadley cells, is probably a contributing factor for lightning observed in this region. Note that we have ignored the potential for local sources of water-soluble chemical species (i.e., dimethylsulfide, DMS) to contribute to variability in local CCN concentrations in this analysis. However, monthly climatology of ocean-atmosphere fluxes of DMS [e.g., *Lana et al.*, 2011] does not appear to explain the existence of or reinforce the background meridional gradient of CCN over the CIO shown here and in prior studies.

4.4 Discussion

Following the results presented in the global CF, regional statistical analysis, and DYNAMO-centric analysis (Secs. 2-3), it would seem that higher lightning rates observed in the northern region compared to the southern region could be attributed to systematic differences in aerosol concentrations between regions, rather than differences in the frequency of occurrence of strongly-electrified deep convective clouds between regions north and south of the equator during the eight-year climatology. Recall that lightning flash rates were approximately 2 times greater in the northern region compared to the southern region as a function of the MJO phase (e.g., Fig. 4.3e) and VPRR composites (Fig. 4.4) depicted greater reflectivity in the mixed-phase region (5-10 km altitude) in addition to similar numbers of deep convective pixels observed by

the TRMM PR, primarily in the suppressed period leading up to MJO onset. The differences in NCAPE and WCD between the northern and southern regions were generally not appreciable, but both thermodynamic quantities varied temporally as a function of MJO phase; there was evidence of the occurrence of periodic outbreaks of cool, dry, continental air near the surface, perhaps in association with anomalous northerly flows, during suppressed periods (RMM phases 4-6), that would act to stabilize the lower troposphere.

Another implication of the equatorward surges of continental air from southern Asia is that large numbers of CCN are transported to regions that are typically pristine in nature (i.e., CCN concentrations < 100 cm^{-3}) over the equatorial regions of the CIO. Furthermore, the results from the climatological composite analysis conducted here suggest that anomalous (meridional) flows in RMM phase 5 are concurrent with enhanced meridional aerosol gradients at the equator. In other words, anomalous wind patterns that occur during preferred times of the intraseasonal cycle sustain high CCN concentrations north of the equator and pristine conditions south of the equator.

Based on these findings, it is tempting to invoke the aerosol hypothesis for suppressed warm-rain processes and invigorated convection to explain the regional differences in convective spectra (e.g., enhanced lightning/vertical reflectivity columns within deep convective clouds) between the northern and southern regions of the CIO. However, close inspection of the RMM phase composites of areal-average values of CCN, lightning, and VPRR for the northern region reveals that temporal variations of lightning flash density in the northern region alone are quite significant in the pre-onset phases of the MJO (e.g., between RMM phases 8 and 1), but CCN concentrations are roughly equal (CCN ~ 200 cm^{-3}). Considering that the thermodynamic quantities studied here show little variation in magnitude in the average sense between RMM

phases 8 and 1 leading up to MJO initiation over the CIO (Fig. 4.3c,d), an explanation for the observed 60% increase in lightning between RMM phases 8 and 1 in the northern region is lacking.

Note that the results from the statistical investigation (presented in Sec. 2.5) illustrated that approximately 58% of the variability (composed of spatial and temporal contributions) in total lightning density over the CIO could be attributed to the 3-parameter multiple linear model consisting of NCAPE, CCN, and WCD. For the logarithmic form of the multiple linear regression model for lightning density over the CIO, the relative weight assigned to CCN was roughly 37% smaller than the relative weight for NCAPE and comparable to the weight assigned to WCD (e.g., Fig 2.27). The RMM phase areal-average composites (Fig. 4.3) illustrate that the range of the temporal variability in lightning between RMM phases is larger than the regional variability between the northern and the southern hemispheres in the CIO. In other words, lightning varies approximately by a factor of 4 from pre-onset to suppressed RMM phases within the northern region, while lightning varies approximately by a factor of 2 on average between the northern and southern regions. There is also evidence of coherent (in-phase) temporal variability between lightning, NCAPE, and WCD, whereas the temporal variations in lightning appear to be in quadrature with the variations in CCN. Is it possible then that CCN are primarily responsible for region-to-region differences in convective intensity over the CIO, while the larger temporal component of variability in convective intensity between different phases of the intraseasonal cycle is driven more by thermodynamics? Despite simultaneous behavioral tendencies for lightning, thermodynamics, and aerosols, as well as the strength of the multiple regression for the CIO on the annual basis (the multiple linear model contributes to explaining more than half of the lightning variance over this region during the eight-year period of interest), the statistics

favor the possibility that other mechanisms may be important for explaining the full variations of convective intensity on such fine scales of interest.

Specifically in the Tropics, the vertical wind shear in the lower and middle troposphere has been shown to strongly influence convective cloud morphology, dimension, and cloud lifetime, subsequently impacting the vertical mass flux, cold-pool dynamics, and secondary convective development [*LeMone et al.*, 1998; *Cetrone and Houze*, 2006; *Rowe and Houze*, 2015]. *Tompkins* [2001] posited that the spatial scales of secondary convection forming along cold-pool boundaries in the Tropics could be attributed in part to the magnitude of surface fluxes as they modify the cold-pool characteristics (e.g., $\Delta\theta_e$ and propagation speed of the cold-pool boundary); the spatial dimensions and vigor of convective updrafts have been cited as key cloud attributes for determining to what extent entrainment and mixing process can impact a cloudy parcel's buoyancy [e.g., *Williams and Stanfill*, 2002]. The efficacy of entrainment/mixing processes to dilute convective updrafts is also likely to be sensitive to the ambient humidity of the free troposphere. In turn, both shear and free-tropospheric relative humidity have been shown to vary considerably during various phases of the MJO [e.g., *Johnson and Ciesielski*, 2013; *Xu and Rutledge*, 2014; *Rowe and Houze*, 2015]. Hence, incorporating the effects of shear and free-tropospheric humidity could lead to improvements when accounting for the spatial and temporal variance of lightning and radar reflectivity over the CIO, but the inclusion of two additional independent variables introduces new potential sources of uncertainty and greater complexity to an already difficult problem.

In this study, the behavior of lightning and radar reflectivity with respect to changes in environmental thermodynamics and aerosols across various phases of the intraseasonal cycle were found to agree with the results from a shorter-duration, yet more comprehensive,

assessment of observations/model analyses from DYNAMO (e.g., Ch. 3), suggesting that the

apparent trends are robust in climatology over the CIO. The results support 1) the idea that

anomalous flows embedded within the larger-scale monsoon flow over southern Asia and the

CIO basin during suppressed convective periods associated with the MJO maintain the

meridional gradient of CCN concentrations in the boundary layer over this region and 2) that

deep convection is stronger north of the equator according to observed regional differences in

lightning and vertical reflectivity structure. The simultaneous importance of both

thermodynamics and aerosols in explaining the observed climatological variability of lightning

and radar quantities is again invoked. Next, these results from the multi-scale investigations

presented here are synthesized and concluding remarks follow thereafter.

CHAPTER 5

Synthesis and conclusion

The results and discussions presented in this dissertation advocate for the simultaneous

hypothesis, i.e., that both aerosols and thermodynamics are important modulators of deep

convective clouds and lightning in the Tropics [e.g., R08; *Koren et al.*, 2010,2012; *Storer et al.*,

2014]. In light of the findings from the global-scale, "simultaneous-type" analysis conducted

here, the difficulty of isolating the impacts of thermodynamics and aerosols on deep convection

apparently stems from the high likelihood of significant correlations amongst environmental

factors on regional scales of interest. In the statistical component of this investigation (e.g., Sec.

2.5), reducing the spatial scale of the observations to individual regions in some cases restricted

the dynamic range of both the predictand and chosen predictors sufficiently to render equivocal

findings, perhaps due to the aforementioned potential flaw in experimental design. For example,

the "heuristic" composite-based analysis of lightning variability over regions offshore and

downstream from major continents (e.g., Sec. 2.5.1) showed evidence for the existence of high

lightning rates in association with stronger low-level pressure troughs that enhanced offshore

transport of continental pollutants ahead of the trough axis; meanwhile, stronger low-level

troughs promote warm-air advection in the lower troposphere beneath cooler air aloft associated

with the mid-level (e.g., at 500 hPa) trough to the west, effectively contributing to increased

NCAPE (stronger convective inhibition for surface-based parcels was also noted).

In other words, environmental variability leading to superlative thermodynamic

conditions favorable for deep convection [e.g., *Williams et al.*, 2005] can also lead to increases in

CCN concentrations. In this scenario, observations of associated convective spectra in the multi-

dimensional parameter space incorporating thermodynamics and aerosols tend to amass along the diagonal plane that bisects the axial directions for individual predictors. Variability in the predictand is then confined strictly to the space where predictors vary strongly together, thereby precluding unambiguous separation of the relative contributions of individual factors.

The success of the global approach depends on observing many convective features across a wide range of the multi-dimensional parameter space, potentially both along diagonal planes between predictors and in regions in the off-diagonal space (refer to Figs. 2.3-2.8). In this way, partial sensitivities can be assessed (as was done here using explicit stratifications first in Sec. 2.3 and using a more-rigorous multiple regression approach assuming independence in the predictor set in Sec. 2.5). As a result of these experiments, estimates of the isolated influence of each independent predictor on the response were compiled. On the global scale, the most influential element was CCN, but in general, independence in the predictor set was apparent for individual regions and the magnitude of the relative weight assigned to each predictor in the statistical models were often found to be comparable on large scales. It is important to emphasize the lack of multicollinearity found for the predictor set consisting of NCAPE, CCN, and WCD when studied on large scales (e.g., hemispheres and regions generally larger than $O(10^6 \text{ km}^2)$, as it validates a schematic depiction to summarize the main finding from this research (Fig. 5.1), i.e., *that convective intensity increases with increasing NCAPE, increasing CCN concentrations, and decreasing WCD*.

Polarizing examples of deep convective clouds that span the continuum of the observations studied here, along with their associated lightning and radar reflectivity characteristics, are shown in Fig. 5.1. The exemplary cloud shown in the left portion of the diagram is meant to depict deep convective clouds that occur in environments that favor

infrequent lightning and insignificant development in the vertical reflectivity structure, i.e., environments with deep WCD, generally low CCN concentrations, and small values of NCAPE; these environments are typically found over oceanic regions (e.g., Fig. 2.5-2.6). Hence, updrafts are shown to be strongest at or below the freezing level [e.g., Zipser, 1994] and the horizontal dimensions of the main convective core and radar reflectivity are relatively narrow [e.g., *Williams and Stanfill*, 2002]. Generally low values of reflectivity between about 0°C and -40°C in the mixed-phase region could be a consequence of weak updrafts that have been observed at these altitudes within deep convective clouds over tropical oceanic locations [e.g., *Heymsfield et al.*, 2010] and may be responsible for the overall lack of lightning associated with this population of deep convective clouds when invoking the theory for thunderstorm charging [e.g., *Takahashi*, 1978; *Williams et al.*, 1991; *Saunders*, 1993].

In contrast, convective features that exhibit high lightning flash rate density and higher AVGHT30 were found to occur in environments where NCAPE was high, CCN concentrations approached 500-1000 cm^{-3}, and WCD was shallower than about 4200 m. Following previous hypotheses [e.g., *Williams and Stanfill*, 2002; *Williams et al.*, 2005; *Carey and Buffalo*, 2007], clouds with shallower WCD are shown to have wider horizontal core dimensions. The highest AVGHT30 was found to be in the vicinity 8 km on average (i.e., altitudes where the temperature is close to -20°C assuming a moist adiabatic lapse rate), and these CFs were generally associated with the highest lightning flash density, as exemplified on the right side of Fig. 5.1. Note that the storm echo top (~17-20 dBZ as resolved by the TRMM PR) is similar in both cases, as reflectivity echo tops in deep convective clouds in the Tropics commonly reach upwards of 15 km, while potentially large differences in the vertical development of the reflectivity column at lower altitudes can still be apparent [*DeMott and Rutledge*, 1998; *Carey and Rutledge*, 2000].

Note that the trend from low (small) to high (large) AVGHT30 (flash density) for increasing NCAPE, increasing CCN, and shallowing WCD was shown for both continental and oceanic subsets in Sec. 2.3, but continental CFs produced high lightning density with high AVGHT30 over a variety of land-surface types, where CCN concentrations exceeded the pristine baseline observed over oceans. Interior land areas were generally the most polluted (e.g., Fig. 2.1) and have therefore been portrayed accordingly in Fig. 5.1; in addition, the differing depictions of aerosols over continental and oceanic regions implies that the source, species, and vertical distribution of aerosols can vary substantially between regions. This aspect of the problem has not been examined explicitly in this analysis, though aerosol speciation could be an important aspect to consider in a future study as discussed above in Sec. 2.4. The difference in aerosol type and relative concentrations of CCN versus ice nuclei at a given location may have implications for differences in the observed radar reflectivity characteristics (beyond those that were characterized here); the magnitude of the difference in near-surface reflectivity between weaker (oceanic) and stronger (continental) CFs on average was found to be small, but exceptions to this rule are apparent in the data studied here.
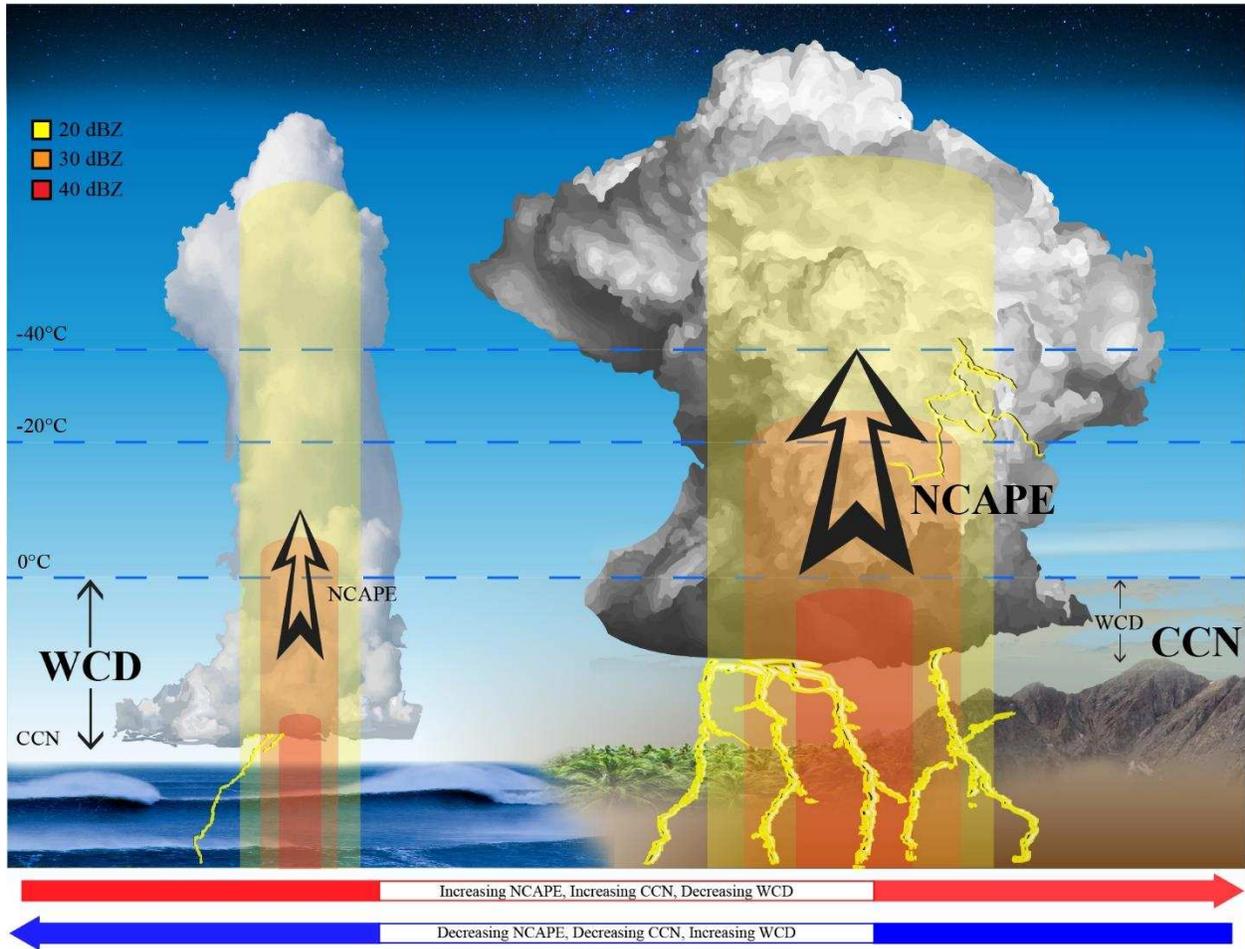
Fig. 5.1. A schematic depiction of the variability of deep convection on the global scale, summarizing the primary finding from the current set of investigations.

Based on the above discussion, "singular hypothesis" perspectives would appear to be inadequate for investigations that attempt to account for the full variability of the population of deep convective clouds in the Tropics; this research and the results from previous studies in the literature contribute evidence to rule out the possibility of aerosols *or* thermodynamics acting as a lone driver of the observed variability of deep convective cloud properties on the global scale. Chs. 2-4 also illustrated that multiple parameters (e.g., NCAPE, CCN, and WCD in addition to other possible factors) likely influence the temporal evolution and spatial variability of deep convective clouds over individual, fine-scale regions.

Previous studies point to the possible importance of other factors that were not considered here, e.g., environmental relative humidity and wind shear, but as mentioned above in Chs. 2 and 4, inclusion of additional independent variables would likely lead to greater complexity in the analysis. If the objective of a future investigation is to strictly reproduce the observed variability in deep convective clouds based on environmental parameters, then adding additional parameters may be appropriate in a forward-type regression analysis. Alternatively, if the goal of a future investigation is to determine the relative contributions of environmental factors to regional variations in convective intensity, then the covariance of additional variables from the physical atmospheric system (e.g., wind shear and relative humidity) with the three predictors studied here (NCAPE, CCN, and WCD) must be understood prior to developing a hypothesis for simultaneous influences of these other factors on convective intensity. Henceforth, it is the author's recommendation that efforts on this front be directed toward understanding why region-to-region differences in the relative weights assigned to individual predictors exist. Are there regions and seasons where the background environment dictates that the influence of certain independent variables becomes either dominant or insignificant? It is of interest to know whether these behaviors are intrinsic to certain regions and/or seasons, as appreciable annual variability was noted in a related analysis of seasonal mean variable behavior (not shown). The discussion now turns to a summary of the research and conclusions.

Why does precipitation intensity, internal structure, and lightning occurrence within deep convective clouds vary throughout the Tropics and Subtropics? This research was motivated in part by the author's anecdotal experience observing tropical thunderstorms and squall lines between the jungles of Sumatra, Panamá, and infrequently in Hawaii. On the other hand, an

extensive body of observations over multiple decades in the literature illustrates notable differences in rainfall intensity and lightning frequency regionally in the Tropics and Subtropics.

Two hypotheses, the thermodynamic and aerosol hypotheses, stand out prominently as they attempt to explain the observed regional differences in convective intensity throughout the Tropics. Early investigators posited a dominance of either thermodynamics *or* aerosols in accounting for the regional variability of convective cloud characteristics (e.g., rainfall intensity) but more recent studies have demonstrated near-orthogonal or the independent influence of both thermodynamics and aerosols on deep convective clouds. Thus, the objective of the research was *to determine the relative contributions of thermodynamics and aerosols to the observed variability of deep convective clouds and lightning in the Tropics.* A simultaneous hypothesis that incorporated the salient aspects of *both* thermodynamic and aerosol paradigms was developed and subsequently tested.

The dissertation was initially envisioned as a four-part collection (parts 1a, 1b, 2a, and 2b) detailing the findings surrounding the variability of deep convective clouds and lightning using a multi-scale experimental approach. A study of the global behavior of convective features and lightning-producing convective features in response to changes in thermodynamic and aerosol characteristics of their immediate environments was undertaken to determine large-scale, "bulk" patterns of variability related to changes in NCAPE, CCN concentrations, and WCD (part 1a; Ch. 2 herein). In the global-scale analysis, an elementary method of data stratification contributed a convincing set of evidence in support of the simultaneous hypothesis for explaining the variability of deep convection on large scales in the Tropics, i.e., between continents and oceans. A natural follow-on to global-scale study was a more rigorous statistical decomposition of the relative importance of the independent variables over individual regions (part 1b; Ch. 2,

139

Sec. 5 herein); the results of the statistical investigation over individual regions (and seasons) were consistent with the findings on the global scale, yet the relative weight assigned to individual independent parameters varied between regions and therefore these results motivated further study to link statistical output to the background meteorological context.

In a shift of focus from global, to regional, to individual basin-scale analyses, data collected between October-December 2011 during the DYNAMO field campaign over the CIO was used to investigate the inter-basin differences in convective spectra with respect to intraseasonal changes (associated with the MJO) in aerosols in the lower troposphere in an otherwise spatially homogeneous thermodynamic background (part 2a; Ch. 3 herein). Coherent variations in the large-scale circulation, convective clouds, and aerosols were noted, consistent with relevant theory for equatorial heating modes/wave disturbances and observations from the INDOEX field campaign that took place more than a decade prior over the CIO. The relative brevity of the DYNAMO data record prompted the author to appeal to the longer climatology of MJO events occurring in previous boreal cold seasons to investigate whether the patterns of covariability noted during DYNAMO were robust (part 2b; Ch. 4 herein). Bearing in the mind the spatial and temporal patterns of variability noted, the results from Chs. 3-4 again pointed to the importance of all three independent variables studied here, highlighting the applicability of the simultaneous hypothesis for convective cloud variability in individual regions/basins (i.e., over finer scales).

The four parts of the research have been combined into three manuscripts for publication in the peer-reviewed literature. The first of which is entitled, "Simultaneous influences of thermodynamics and aerosols on deep convection and lightning in the Tropics" and was accepted for publication in the Journal of Geophysical Research – Atmospheres (JGR-A) on 3 June 2015.

Ch. 2, Sec. 5 will be revised and prepared as a stand-alone manuscript to follow the aforementioned article from JGR-A; the questions posed at the end of Ch. 2, Sec. 4 have been addressed accordingly in Ch. 2, Sec. 5 and it is the view of the author that these results warrant publication in light of continuity and to guide on-going science pursuits. Ch. 3 of the dissertation was originally submitted to Geophysical Research Letters (GRL) and received generally favorable reviews in the editor's response. After much thought and deliberation, the author and his coauthors have opted to combine Ch. 3 and Ch. 4 into a cohesive manuscript that more thoroughly addresses the criticism put forth by reviewers from GRL. The primary results from the collection of manuscripts can be summarized as follows:

- In a study of more than 1.4 million CFs/260,000 LPCFs in the TRMM satellite observation domain between the years 2004-2011 (Ch. 2), *convective intensity was shown to increase with increasing NCAPE, increasing CCN, and decreasing WCD over both continents and oceans.*
    - New observational evidence for the simultaneous influence of WCD on deep convection was presented; results agree with previous findings
    - Differences in the VPRR for polluted environments (CCN > 500 cm^{-3}) depicted ~5.0 dB enhancements at 5-10 km altitude compared to pristine environments (CCN < 100 cm^{-3}) holding thermodynamic quantities constant
    - Total lightning density (TLD) and the average height of 30 dBZ (AVGHT30) echoes were strongly correlated in the average sense; the slope of this relationship was sensitive to the background aerosol concentration

- A statistical decomposition (Ch. 2) of the relative influence of NCAPE, CCN, and WCD on TLD and AVGHT30 illustrated that simple linear models of the three independent variables accounted for up to approximately 78% of the variance in the lightning and radar reflectivity on the global-scale
  - In the multiple regression statistical output, the relative weight assigned to CCN was double the relative weight assigned to NCAPE and WCD, respectively, on the global scale; in general, the relative weight applied to each independent variable had the same sign regardless of the individual region studied, but the magnitudes of regression coefficients varied regionally
  - The global observation strategy contributed to sufficient independence in the predictor set to ensure robust quantification of the influence of each variable in isolation
  - The multiple correlation and general invariance noted in the magnitude and sign of regression parameters *on the large-scale* suggest that these results may be applied in efforts to improve global chemical transport modeling (e.g., LNO_x generation and dispersion)
- A fine-scale analysis of the relative contribution of CCN to the observed variability of convective clouds over the CIO (Ch. 3) suggested that systematic differences in CCN concentrations in the lower troposphere between the northern and southern regions of the CIO contribute to more lightning (up to a factor of 10) and greater radar reflectivity (~2-3 dB; maximized in the mixed phase region) north of the equator compared to south of the equator.

- o Variability in the large-scale circulation on intraseasonal timescales appears to maintain the meridional gradient of lower-tropospheric aerosols; implied stronger convective heating in polluted regions may lead to enhanced aerosol-cloud-circulation feedbacks in association with the MJO
- A climatological study of the covariability between rainfall, CCN, NCAPE, WCD, and lightning over the same region again illustrated systematic differences in CCN, similar thermodynamics and rainfall, and higher lightning rates north of the equator compared to south of the equator in the CIO (for approximately more than 20 MJO episodes during the years 2004-2011; Ch. 4)
  - o Climatological behavior was found to be similar to that observed during DYNAMO, emphasizing the robust nature of the patterns of covariability found between rainfall, CCN, thermodynamics, and lightning over the CIO

The primary findings of this dissertation advocate for the simultaneous hypothesis that aerosols modulate the amount of available thermodynamic potential energy realized throughout the lifecycle of a tropical convective cloud (also noting sensitivity of the response to the depth of the cloud's warm phase). As demonstrated here, a considerable amount of the variability in the deep convective cloud population can be accounted for when studying the response of convective intensity to the combination of NCAPE, CCN, and WCD on large spatial scales (following previous hypotheses in the literature); but, as the time and space scales of interest become progressively finer, the results indicate that other factors may be important for explaining observed tendencies in lightning and radar reflectivity characteristics.

The next generation of observational platforms is in development (or is in the process of being deployed) and the outlook for passive and active remote-sensing strategies to address some of the continued uncertainties surrounding simultaneous thermodynamic and aerosol influences on deep convection is promising in light of on-going and planned missions [*Goodman et al.*, 2013; *Heymsfield et al.*, 2013; *Rennó et al.*, 2013; *Hou et al.*, 2014; *Rosenfeld et al.*, 2014]. It is the hope of the author that this research provides the impetus for future investigations that extend this analysis to extratropical latitudes while utilizing high-resolution data from the next generations of geostationary satellites and space-borne radar (e.g., GOES-R and the GPM satellite constellation), lightning detection systems, and chemical transport models.

REFERENCES

Adams, P. J. and J. H. Seinfeld (2002), Predicting global aerosol size distributions in general circulation models. *J. Geophys. Res.*, 107(D19), 4370, doi:10.1029/2001JD001010.

Albrecht, R. I., C. A. Morales, and M. A. F. Silva Dias (2011), Electrification of precipitating systems over the Amazon: Physical processes of thunderstorm development, *J. Geophys. Res.*, 116, D08209, doi:10.1029/2010JD014756.

Allen, D. J. and K. E. Pickering (2002), Evaluating lightning flash rate parameterizations for use in a global chemical transport model. *J. Geophys. Res.*, 107(D23), 4711, doi:10.1029/2002JD002066.

Altaratz, O., I. Koren, Y. Yair, and C. Price (2010), Lightning response to smoke Amazonian fires. *Geophys. Res. Lett.*, 37, L07801, doi:10.1029/2010GL042679.

Altaratz, O., I. Koren, L. A. Remer, and E. Hirsch (2014), Review: Cloud invigoration by aerosols–coupling between microphysics and dynamics. *Atmos. Res.*, 140, 38-60, doi: dx.doi.org/10.1016/j.atmosres.2014.01.009.

Anderson, T. L., R. J. Charlson, D. M. Winker, J. A. Ogren, and K. Holmén (2003), Mesoscale variations of tropospheric aerosols. *J. Atmos. Sci.*, 60, 119–136.

Andreae, M. O. and D. Rosenfeld (2008), Aerosol-cloud-precipitation interactions part 1: The nature and sources of cloud-active aerosol. *Earth-Science Reviews*, 89, 13–41.

Andreae, M. O., D. Rosenfeld, P. Artaxo, A. A. Costa, G. P. Frank, K. M. Longo, and M. A. F. Silva-Dias (2004), Smoking rain clouds over the Amazon. *Science*, 303, 1337–1342. doi: 10.1126/science.1092779.

Bates, T. S., D. J. Coffman, D. S. Covert, and P. K. Quinn (2002), Regional marine boundary
layer aerosol size distributions in the Indian, Atlantic, and Pacific Oceans: A comparison
of INDOEX measurements with ACE-1, ACE-2, and Aerosols99, *J. Geophys. Res.*,
107(D19), 8026, doi:10.1029/2001JD001174.

Baker, B., M. B. Baker, E. R. Jayaratne, J. Latham, and C. P. R. Saunders (1987), The influence
of diffusional growth rates on the charge transfer accompanying rebounding collisions
between ice crystals and soft hailstones. *Q. J. Roy. Meteor. Soc.*, 113, 1193–1215.

Baker M. B. and J. G. Dash (1994), Mechanism of charging between colliding ice particles in
thunderstorms. *J. Geophys. Res.*, 99(D5), 10621–10626, doi: 10.1029/93JD01633.

Beegum, S. N., K. Krishna Moorthy, S. S. Babu, R. R. Reddy, and K. R. Gopal (2009), Large
scale modulations of spectral aerosol optical depths by atmospheric planetary waves,
Geophys. Res. Lett., 36, L03810, doi: 10.1029/2008GL036509.

Bell, T. L., D. Rosenfeld, K.-M. Kim, J.-M. Yoo, M.-I. Lee, and M. Hahnenberger (2008),
Midweek increase in U.S. summer rain and storm heights suggests air pollution
invigorates rainstorms, *J. Geophys. Res.*, 113, D02209, doi:10.1029/2007JD008623.

Benedetti, A., et al. (2009), Aerosol analysis and forecast in the European Centre for Medium-
Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, *J. Geophys.
Res.*, 114, D13205, doi:10.1029/2008JD011115.

Blanchard, D. O. (1998), Notes and correspondence: Assessing the vertical distribution of
convective available potential energy. *Wea. Forecast.*, 13, 870–877.

Boccippio, D. J., S. J. Goodman, and S. Heckman (2000), Regional differences in tropical
lightning distributions. *J. App. Met.*, 39, 2231–2248.

Boccippio, D. J., K. L. Cummins, H. J. Christian, and S. J. Goodman (2001), Combined satellite-

Bates, T. S., D. J. Coffman, D. S. Covert, and P. K. Quinn (2002), Regional marine boundary
layer aerosol size distributions in the Indian, Atlantic, and Pacific Oceans: A comparison
of INDOEX measurements with ACE-1, ACE-2, and Aerosols99, *J. Geophys. Res.*,
107(D19), 8026, doi:10.1029/2001JD001174.

Baker, B., M. B. Baker, E. R. Jayaratne, J. Latham, and C. P. R. Saunders (1987), The influence
of diffusional growth rates on the charge transfer accompanying rebounding collisions
between ice crystals and soft hailstones. *Q. J. Roy. Meteor. Soc.*, 113, 1193–1215.

Baker M. B. and J. G. Dash (1994), Mechanism of charging between colliding ice particles in
thunderstorms. *J. Geophys. Res.*, 99(D5), 10621–10626, doi: 10.1029/93JD01633.

Beegum, S. N., K. Krishna Moorthy, S. S. Babu, R. R. Reddy, and K. R. Gopal (2009), Large
scale modulations of spectral aerosol optical depths by atmospheric planetary waves,
Geophys. Res. Lett., 36, L03810, doi: 10.1029/2008GL036509.

Bell, T. L., D. Rosenfeld, K.-M. Kim, J.-M. Yoo, M.-I. Lee, and M. Hahnenberger (2008),
Midweek increase in U.S. summer rain and storm heights suggests air pollution
invigorates rainstorms, *J. Geophys. Res.*, 113, D02209, doi:10.1029/2007JD008623.

Benedetti, A., et al. (2009), Aerosol analysis and forecast in the European Centre for Medium-
Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, *J. Geophys.
Res.*, 114, D13205, doi:10.1029/2008JD011115.

Blanchard, D. O. (1998), Notes and correspondence: Assessing the vertical distribution of
convective available potential energy. *Wea. Forecast.*, 13, 870–877.

Boccippio, D. J., S. J. Goodman, and S. Heckman (2000), Regional differences in tropical
lightning distributions. *J. App. Met.*, 39, 2231–2248.

Boccippio, D. J., K. L. Cummins, H. J. Christian, and S. J. Goodman (2001), Combined satellite-

and surface-based estimation of the intracloud–cloud-to-ground lightning ratio over the continental United States. *Mon. Wea. Rev.*, 129, 108–122, doi: http://dx.doi.org/10.1175/1520-0493(2001)129<0108:CSASBE>2.0.CO;2.

Boccippio, D. J., W. A. Petersen, and D. J. Cecil (2005), The tropical convective spectrum. Part I: Archetypal vertical structures. *J. Climate*, 18, 2744–2769.

Burrows, S. M., C. Hoose, U. Pöschl, and M. G. Lawrence (2013), Ice nuclei in marine air: biogenic particles or dust?, *Atmos. Chem. Phys.*, 13, 245-267, doi:10.5194/acp-13-245-2013.

Byers, H. R. and R. R. Braham (1948), Thunderstorm structure and circulation. *J. Meteor.*, 5(3), 71–86.

Carey, L. D., and S. A. Rutledge, (2000), The relationship between precipitation and lightning in tropical island convection: A C-band polarimetric radar study. *Mon. Wea. Rev.*, 128, 2687–2710.

Carey, L. D., and K. M. Buffalo (2007), Environmental control of cloud-to-ground lightning polarity in severe storms. *Mon. Wea. Rev.*, 135, 1327–1353.

Cecil, D. J., S. J. Goodman, D. J. Boccippio, E. J. Zipser, and S. W. Nesbitt (2005), Three years of TRMM precipitation features. Part I: Radar, radiometric, and lightning characteristics. *Mon. Wea. Rev.*, 133, 543–566.

Cecil, D. J., D. E. Buechler, and R. J. Blakeslee (2014), Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. *Atmos. Res.*, 135–136, 404–414, doi:10.1016/j.atmosres.2012.06.028.

Cetrone, J. and R. A. Houze Jr. (2006), Characteristics of tropical convection over the ocean near Kwajalein. *Mon. Wea. Rev.*, 134, 834–853.

Christian, H. J., R. J. Blakeslee, D. J. Boccippio, W. L. Boeck, D. E. Buechler, K. T. Driscoll, S. J. Goodman, J. M. Hall, W. J. Koshak, D. M. Mach, and M. F. Stewart (2003), Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *J. Geophys. Res.*, 108(D1), doi:10.1029/2002JD002347.

Ciesielski, P. E., and coauthors (2014), Quality-controlled upper-air sounding dataset for DYNAMO/CINDY/AMIE: Development and corrections. *J. Atmos. Ocean. Tech.*, 31, 741–764, doi: 10.1175/JTECH-D-13-00165.1.

Cifelli R., W. A. Petersen, L. D. Carey, S. A. Rutledge, and M. A. F. da Silva Dias (2002), Radar observations of the kinematic, microphysical, and precipitation characteristics of two MCSs in TRMM LBA, *J. Geophys. Res.*, 107(D20), 8077–8093, doi:10.1029/2000JD000264.

D'Andrea, S. D., S. A. K. Häkkinen, D. M. Westervelt, C. Kuang, E. J. T. Levin, V. P. Kanawade, R. Leaitch, D. V. Spracklen, I. Riipinen, and J. R. Pierce (2013), Understanding global secondary organic aerosol amount and size-resolved condensational behavior. *Atmos. Chem. Phys.*, 13, 11519–11534, doi:10.5194/acp-13-11519-2013.

Dee, D. P. et al., (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. Roy. Met. Soc.*, 137, 533–597.

DeMott, C. A., S. A. Rutledge (1998), The vertical structure of TOGA COARE convection. Part I: radar echo distributions. *J. Atmos. Sci.*, 55, 2730–2747.

Draper, N. R. and H. Smith (1998), *Applied Regression Analysis*, 3rd ed., Wiley-Interscience, New York, 736 pp.

Dusek, U., G. P. Frank, L. Hildebrandt, J. Curtius, J. Schneider, S. Walter, D. Chand,

F. Drewnick, S. Hings, D. Jung, S. Borrmann, and M. O. Andreae (2006), Size matters more than chemistry for cloud-nucleating ability of aerosol particles. *Science*, 312, 1375–1378, doi: 10.1126/science.1125261.

Dye, J. E., W. P. Winn, J. J. Jones, and D. W. Breed (1989), The electrification of New Mexico thunderstorms 1. Relationship between precipitation development and the onset of electrification, *J. Geophys. Res.*, 94(D6), 8643–8656, doi: 10.1029/JD094iD06p08643.

Fan, J., L. R. Leung, D. Rosenfeld, Q. Chen, Z. Li, J. Zhang, and H. Yan (2013), Microphysical effects determine macrophysical response for aerosol impacts on deep convective clouds. *Proc. Nat. Acad. Sci.*, 110(48), E4581–E4590, doi: 10.1073/pnas.1316830110.

Fan, J., R. Zhang, G. Li, and W.-K. Tao (2007), Effects of aerosols and relative humidity on cumulus clouds, *J. Geophys. Res.*, 112, D14204, doi:10.1029/2006JD008136.

Fan, J., T. Yuan, J. M. Comstock, S. Ghan, A. Khain, L. R. Leung, Z. Li, V. J. Martins, and M. Ovchinnikov (2009), Dominant role by vertical wind shear in regulating aerosol effects on deep convective clouds, *J. Geophys. Res.*, 114, D22206, doi:10.1029/2009JD012352.

Field, R. D., G. R. van der Werf, and S. S. P. Shen (2009), Human amplification of drought-induced biomass burning in Indonesia since 1960. *Nature Geosci.*, 2, 185–188, doi:10.1038/NGEO443.

Gill, A. E. (1980), Some simple solutions for heat-induced tropical circulation. *Quart. J. Roy. Meteor. Soc.*,106, 447–462.

Guo, Y., B. Tian, R. A. Kahn, O. Kalashnikova, S. Wong, and D. E. Waliser (2013), Tropical Atlantic dust and smoke aerosol variations related to the Madden-Julian Oscillation in MODIS and MISR observations. *J. Geophys. Res. Atmos.*, 118, 4947–4963, doi:10.1002/jgrd.50409.

Graf, H. F. (2004), The complex interaction of aerosols and clouds. *Science*, 303, 1309–1311, doi: 10.1126/science.1094411.

Grant, L. D., and S. C. van den Heever (2014), Microphysical and dynamical characteristics of low-precipitation and classic supercells. *J. Atmos. Sci.*, 71, 2604–2624, doi:10.1175/JAS-D-13-0261.1.

Grant, L. D. and S. C. van den Heever, (2015), Cold pool and precipitation responses to aerosol loading: Modulation by dry layers. *J. Atmos. Sci.*, 72, 1398–1408, doi: http://dx.doi.org/10.1175/JAS-D-14-0260.1.

Halverson, J. B., T. Rickenbach, B. Roy, H. Pierce, and E. R. Williams (2002), Environmental characteristics of convective systems during TRMM-LBA. *Mon. Wea. Rev.*, 130, 1493–1509.

Hannachi, A., I. T. Jolliffe, and D. B. Stevenson (2007), Empirical orthogonal functions and related techniques in atmospheric science: A review. *Intl. J. Climatol.*, 27, 1119–1152, doi: 10.1002/joc.1499.

Heiblum, R. H., I. Koren, and O. Altaratz (2012), New evidence of cloud invigoration from TRMM measurements of rain center of gravity, *Geophys. Res. Lett.*, 39, L08803, doi:10.1029/2012GL051158.

Heintzenberg, J., D. C. Covert, and R. Vandingenen (2000), Size distribution and chemical composition of marine aerosols: A compilation and review. *Tellus*, 52B, 1104–1122.

Heymsfield, G. M., L. Tian, A. J. Heymsfield, L. Li, and S. Guimond (2010), Characteristics of deep tropical and subtropical convection from nadir-viewing high-altitude airborne doppler radar. *J. Atmos. Sci.*, 67, 285–308, doi: 10.1175/2009JAS3132.1.

Heymsfield, G. M., L. Tian, L. Li, M. McLinden, and J. I. Cervantes (2013), Airborne radar

observations of severe hailstorms: Implications for future spaceborne radar. *J. App.*

   Meteor. Clim., 52, 1851–1867, doi: 10.1175/JAMC-D-12-0144.1.

Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K.

   Nakamura, and T. Iguchi (2014), The global precipitation measurement mission. *Bull.*

   Amer. Meteor. Soc., 95, 701–722, doi: http://dx.doi.org/10.1175/BAMS-D-13-00164.1.

Houze Jr., R. A., and A. K. Betts (1981), Convection in GATE, *Rev. Geophys.*, 19(4), 541–576,

   doi:10.1029/RG019i004p00541.

Hudson, J. G., and S. Mishra (2007), Relationships between CCN and cloud microphysics

   variations in clean maritime air, *Geophys. Res. Lett.*, 34, L16804,

   doi:10.1029/2007GL030044.

Huffman, G. J., and Coauthors, (2007), The TRMM Multisatellite Precipitation Analysis

   (TMPA): Quasi-global, multiyear, combined sensor precipitation estimates at fine scales.

   J. Hydrometeor., 8, 38–55, doi:10.1175/JHM560.1.

Iguchi, T., T. Kozu., R. Meneghini, J. Awaka, and K. Okamoto (2000), Rain-profiling algorithm

   for the TRMM precipitation radar. *J. Appl. Meteor.*, 39, 2038–2052, doi:

   http://dx.doi.org/10.1175/1520-0450(2001)040<2038:RPAFTT>2.0.CO;2.

Iribarne, J. V. and W. L. Godson (1981), *Atmospheric thermodynamics*, 2nd ed., D. Reidel

   Publishing Co., Dordrecht, Netherlands.

Johnson R. H. and P. E. Ciesielski (2013), Structure and properties of Madden–Julian

   Oscillations deduced from DYNAMO sounding arrays. *J. Atmos. Sci.*, 70, 3157–3179.

   doi: http://dx.doi.org/10.1175/JAS-D-13-065.1.

Jorgensen, D. P. and M. A. LeMone (1989), Vertical velocity characteristics of oceanic

convection. *J. Atmos. Sci.*, 46, 621–640, doi: http://dx.doi.org/10.1175/1520-0469(1989)046<0621:VVCOOC>2.0.CO;2.

Kelley, O. A., J. Stout, M. Summers, and E. J. Zipser (2010), Do the tallest convective cells over the tropical ocean have slow updrafts? *Mon. Wea. Rev.*, 138, 1651–1672.

Khain, A. P., D. Rosenfeld, and A. Pokrovsky (2005), Aerosol impact on the dynamics and microphysics of deep convective clouds. *Q. J. Roy. Met. Soc.*, 131, 2639–2663. doi: 10.1256/qj.04.62.

Khain, A. P., T. Prabha, N. Benmoshe, G. Pandithurai, and M. Ovchinnikov (2013), The mechanism of first raindrops formation in deep convective clouds. *J. Geophys. Res. Atmos.*, 118, 9123–9140, doi:10.1002/jgrd.50641.

Khain, A. P., V. Phillips, N. Benmoshe, and A. Pokrovsky (2012), The role of small soluble aerosols in the microphysics of deep maritime clouds. *J. Atmos. Sci.*, 69, 2787–2805.

Kiladis, G. N., K. H. Straub, and P. T. Haertel (2005), Zonal and Vertical Structure of the Madden–Julian Oscillation. *J. Atmos. Sci.*, 62, 2790–2809, doi: http://dx.doi.org/10.1175/JAS3520.1.

Kiladis, G. N., J. Dias, K. H. Straub, M. C. Wheeler, S. N. Tulich, K. Kikuchi, K. M. Weickmann, M. J. Ventrice (2014), A comparison of OLR and circulation-based indices for tracking the MJO. *Mon. Wea. Rev.*, 142, 1697–1715, doi: 10.1175/MWR-D-13-00301.1.

Koren, I., Y. J. Kaufman, D. Rosenfeld, L. A. Remer, and Y. Rudich (2005), Aerosol invigoration and restructuring of Atlantic convective clouds. *Geophys. Res. Lett.*, 32, L14828, doi:10.1029/2005GL023187.

Koren, I., G. Feingold, and L. A. Remer (2010), The invigoration of deep convective clouds over

the Atlantic: aerosol effect, meteorology or retrieval artifact? *Atmos. Chem. Phys.*, 10, 8855–8872, doi:10.5194/acp-10-8855-2010.

Koren, I., O. Altaratz, L. A. Remer, G. Feingold, J. Vanderlei Martins, and R. H. Heiblum (2012), Aerosol-induced intensification of rain from the Tropics to the midlatitudes. *Nature Geosci.*, 5, 118–122. doi:10.1038/ngeo1364.

Krishnamurti, T. N., B. Jha, J. Prospero, A. Jayaraman, and V. Ramanathan (1998), Aerosol and pollutant transport and their impact on radiative forcing over the tropical Indian Ocean during the January–February 1996 pre-INDOEX cruise. *Tellus*, 50B, 521–542, doi: 10.1034/j.1600-0889.1998.00009.x.

Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson (1998), The Tropical Rainfall Measuring Mission (TRMM PR) sensor package, *J. Atmos. Oceanic Technol.*, 15, 809-817.

Lana, A., et al. (2011), An updated climatology of surface dimethlysulfide concentrations and emission fluxes in the global ocean, *Global Biogeochem. Cycles*, 25, GB1004, doi:10.1029/2010GB003850.

Langley DeWitt, H., D. J. Coffman, K. J. Schulz, W. Alan Brewer, T. S. Bates, and P. K. Quinn (2013), Atmospheric aerosol properties over the equatorial Indian Ocean and the impact of the Madden-Julian Oscillation. *J. Geophys. Res. Atmos.*, 118, doi:10.1002/jgrd.50419.

Lau, K. H. and D. E. Waliser (2011), *Intraseasonal Variability in the Atmosphere-Ocean Climate System*. Praxis, 646 pp.

Lebo, Z. J. and H. Morrison (2014), Dynamical effects of aerosol perturbations on simulated idealized squall lines. *Mon. Wea. Rev.*, 142, 991–1009, doi: 10.1175/MWR-D-13-00156.1.

Lee, S. S., L. J. Donner, V. T. J. Phillips, and Y. Ming (2008), The dependence of aerosol effects on clouds and precipitation on cloud-system organization, shear and stability, *J. Geophys. Res.*, 113, D16202, doi:10.1029/2007JD009224.

Lee, S. S. (2012), Effect of aerosol on circulations and precipitation in deep convective clouds. *J. Atmos. Sci.*, 69, 1957 – 1974, doi: 10.1175/JAS-D-11-0111.1.

LeMone, M. A. and E. J. Zipser (1980), Cumulonimbus vertical velocity events during GATE. Part I: Diameter, intensity, and mass flux. *J. Atmos. Sci.*, 37, 2444–2457, doi: http://dx.doi.org/10.1175/1520-0469(1980)037<2444:CVVEIG>2.0.CO;2.

LeMone, M. A., E. J. Zipser, and S. B. Trier (1998), The role of environmental shear and thermodynamic conditions in determining the structure and evolution of mesoscale convective systems during TOGA COARE. *J. Atmos. Sci.*, 55, 3493–3518.

Li, F. and V. Ramanathan (2002), Winter to summer monsoon variation of aerosol optical depth over the tropical Indian Ocean. *J. Geophys. Res.*, 107(D16), 4284, doi: 10.1029/2001JD000949.

Li, X., W-K. Tao, H. Masunaga, G. Gu, X. Zeng (2013), Aerosol effects on cumulus congestus population over the Tropical Pacfic: A cloud-resolving modeling study. *J. Meteor. Soc. Japan*. 91(6), 817–833. doi:10.2151/jmsj.2013-607.

Li, Z., F. Niu, J. Fan, Y. Liu, D. Rosenfeld, and Y. Ding (2011), The long-term impacts of aerosols on the vertical development of clouds and precipitation, *Nature Geosci.*, doi: 10.1038/NGEO1313.

Liu, C., D. J. Cecil, E. J. Zipser, K. Kronfeld, and R. Robertson (2012), Relationships between lightning flash rates and radar reflectivity vertical structures in thunderstorms over the Tropics and Subtropics, *J. Geophys. Res.*, 117, D06212, doi: 10.1029/2011JD017123.

Liu, C., E. J. Zipser, D. J. Cecil, S. W. Nesbitt, and S. Sherwood (2008), A cloud and

 precipitation feature database from nine years of TRMM observations. *J. Appl. Meteor.*

 Climatol., 47, 2712–2728, doi: http://dx.doi.org/10.1175/2008JAMC1890.1.

Lohmann, U. and J. Feichter (2005), Global indirect aerosol effects: A review. *Atmos. Chem.*

 Phys., 5, 715–737.

Lucas, C., E. J. Zipser, and M. A. LeMone (1994a), Convective available potential energy in the

 environment of oceanic and continental clouds: Corrections and comments. *J. Atmos.*

 Sci., 51, 3829–3830.

Lucas, C., E. J. Zipser, and M. A. LeMone (1994b), Vertical velocity in oceanic convection off

 tropical Australia. *J. Atmos. Sci.*, 51, 3183–3193.

Mackerras, D., M. Darveniza, R. E. Orville, E. R. Williams, and S. J. Goodman (1998), Global

 Lightning: Total, cloud, and ground flash estimates, *J. Geophys. Res.*, 103(D16), 19791–

 19809, doi: 10.1029/98JD01461.

Madden, R. A., and P. R. Julian (1971), Detection of a 40-50 day oscillation in the zonal wind in

 the tropical Pacific. *J. Atmos. Sci.*, 28, 702–708, doi: http://dx.doi.org/10.1175/1520-

 0469(1971)028<0702:DOADOI>2.0.CO;2.

Madden, R. A., and P. R. Julian (1972), Description of global-scale circulation cells in the

 Tropics with a 40-50 day period. *J. Atmos. Sci.*, 29, 1109–1123, doi:

 http://dx.doi.org/10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2.

Madden, R. A. (1986), Seasonal variations of the 40-50 day oscillation in the Tropics. *J. Atmos.*

 Sci., 43, 3138–3158.

Madden, R. A., and P. R., Julian (1994), Observations of the 40–50-Day Tropical Oscillation – A

Review. *Mon. Wea. Rev.*, 122, 814–837, doi: http://dx.doi.org/10.1175/1520-0493(1994)122<0814:OOTDTO>2.0.CO;2/.

Mansell, E. R. and C. L. Ziegler (2013), Aerosol effects on simulated storm electrification and precipitation in a two-moment bulk microphysics model. *J. Atmos. Sci.*, 70, 2032–2050, doi: 10.1175/JAS-D-12-0264.1.

May, P. T., G. Allen, G. Vaughan, and P. Connolly (2009), Aerosol and thermodynamic effects on tropical cloud systems during TWPICE and ACTIVE. *Atmos. Chem. Phys.*, 9, 15–24, doi:10.5194/acp-9-15-2009.

May, P. T., V. N. Bringi, and M. Thurai (2011), Do we observe aerosol impacts on DSDs in strongly forced tropical thunderstorms? *J. Atmos. Sci.*, 68, 1902–1910.

McGee, C. J. and S. C. van den Heever (2014), Latent heating and mixing due to entrainment in tropical deep convection. *J. Atmos. Sci.*, 71, 816–832, doi: 10.1175/JAS-D-13-0140.1.

Montgomery, D. C., E. A. Peck, and G. G. Vining (2012), *Introduction to Linear Regression Analysis*, 5th ed. Wiley, New York, 672 pp.

Moorthy, K. K. and A. Saha (2000), Aerosol study during INDOEX: observation of enhanced aerosol activity over the Mid Arabian Sea during the northern winter. *J. Atmos. Sol.-Terr. Phys.*, 62, 65-72, doi: 10.1016/S1364-6826(99)00081-4.

Morcrette, J.-J., et al. (2009), Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: Forward modeling, *J. Geophys. Res.*, 114, D06206, doi:10.1029/2008JD011235.

Morrison, H., and W. W. Grabowski (2013), Response of tropical deep convection to localized heating perturbations: Implications for aerosol-induced convective invigoration. *J. Atmos. Sci.*, 70, 3533–3555, doi:10.1175/JAS-D-13-027.1.

Nakazawa, T. (1988), Tropical super clusters within intraseasonal variations over the western Pacific. *J. Met. Soc. Japan*, 66(6), 823–839.

Nesbitt, S. W., E. J. Zipser, and D. J. Cecil (2000), A census of precipitation features in the Tropics using TRMM: Radar, ice scattering, and lightning observations. *J. Climate*, 13, 4087–4106.

Niu, F., and Z. Li, (2012), Systematic variations of cloud top temperature and precipitation rate with aerosols over the global Tropics, *Atmos. Chem. Phys.*, doi:10.5194/acp-12-8491-2012.

Orville, R. E., and R. W. Henderson (1986), Global distribution of midnight lightning: September 1977 to August 1978. *Mon. Wea. Rev.*, 114, 2640–2653.

Petersen, W. A., H. J. Christian, and S. A. Rutledge (2005), TRMM observations of the global relationship between ice water content and lightning. *Geophys. Res. Lett.*, 32, L14819, doi:10.1029/2005GL023236.

Petersen, W. A., R. Fu, M. Chen, and R. Blakeslee (2006), Intraseasonal forcing of convection and lightning activity in the southern amazon as a function of cross-equatorial flow. *J. Climate*, 19, 3180–3196.

Petersen, W. A. and S. A. Rutledge (2001), Regional variability in tropical convection: Observations from TRMM, *J. Climate*, 14, 3566–3586.

Petersen, W. A., S. A. Rutledge, and R. E. Orville (1996), Cloud-to-ground lightning observations from TOGA COARE: Selected results and lightning location algorithms. *Mon. Wea. Rev.*, 124, 602–620.

Pickering, K. E., A. M. Thompson, R. R. Dickerson, W. T. Luke, D. P. McNamara, J. P.

Greenburg, and P. R. Zimmerman (1990), Model calculations of tropospheric ozone production potential following observed convective events. *J. Geophys. Res.*, 95(D9), 14049–14062.

Pickering, K. E., Y. Wang, W.-K. Tao, C. Price, J.-F. Müller (1998), Vertical distributions of lightning NO_x for use in regional and global chemical transport models. *J. Geophys. Res.*, 103(D23), 31203–31216.

Pierce, J. R. and Adams, P. J. (2009), Uncertainty in global CCN concentrations from uncertain aerosol nucleation and primary emission rates. *Atmos. Chem. Phys.*, 9, 1339–1356, doi:10.5194/acp-9-1339-2009.

Pierce, J. R., M. J. Evans, C. E. Scott, S. D. D'Andrea, D. K. Farmer, E. Sweitlicki, D. V. Spracklin (2013), Weak global sensitivity of cloud condensation nuclei and the aerosol indirect effect to Criegee+SO2 chemistry. *Atmos. Chem. Phys.*, 13, 3163–3176, doi:10.5194/acp-13-3163-2013.

Ramanathan, V. (2001), Aerosols, climate, and the hydrological cycle. *Science*, 294, 2119–2124, doi: 10.1126/science.1064034.

Ramanathan, V. et al. (2001), Indian Ocean Experiment: An integrated analysis of the climate forcing and effects of the great Indo-Asian haze. *J. Geophys. Res.*, 106 (D22), 28371–28398, doi: 10.1029/2001JD900133.

Rennó, N. O. et al. (2013), CHASER: An innovative satellite mission concept to measure the effects of aerosols on clouds and climate. *Bull. Amer. Met. Soc.*, 94, 685–694, doi: http://dx.doi.org/10.1175/BAMS-D-11-00239.1.

Reynolds, S . E., M. Brook, and M. F. Gourley (1957), Thunderstorm charge separation, *J . Meteorol.*, 14, 426-436.

Romps, D. M. and Z. Kuang (2010), Do undiluted convective plumes exist in the upper tropical troposphere?. *J. Atmos. Sci.*, 67, 468–484, doi: http://dx.doi.org/10.1175/2009JAS3184.1.

Rosenfeld, D. and I. Lensky (1998), Satellite-based insights into precipitation formation processes in continental and maritime convective clouds. *Bull. Amer. Meteor. Soc.*, 79(11), 2457-2476.

Rosenfeld, D. (1999), TRMM observed first direct evidence of smoke from forest fires inhibiting rainfall. *Geophys. Res. Lett.*, 26(20), 3105-3108, doi: 10.1029/1999GL006066.

Rosenfeld, D., R. Lahav, A. Khain, M. Pinsky (2002), The role of sea spray in cleansing air pollution over ocean via cloud processes. *Science*, 297, 1667–1670, doi: 10.1126/science.1073869.

Rosenfeld, D., U. Lohmann, G. B. Raga, C. D. O'Dowd, M. Kulmala, S. Fuzzi, A. Reissell, and M. O. Andreae (2008), Flood or drought: how do aerosols affect precipitation? *Science*, 321, 1309–1313.

Rosenfeld, D., et al. (2014), Global observations of aerosol-cloud-precipitation climate interactions, *Rev. Geophys.*, 52, 750–808, doi:10.1002/2013RG000441.

Rowe, A. K., and R. A. Houze Jr. (2015), Cloud organization and growth during the transition from suppressed to active MJO conditions, *J. Geophys. Res. Atmos.*, 120, doi:10.1002/2014JD022948.

Rutledge, S. A., E. R. Williams, and T. D. Keenan (1992), The Down Under Doppler and Electricity Experiment (DUNDEE): Overview and preliminary results. *Bull. Amer. Meteor. Soc.*, 73, 3–16.

Satheesh, S. K., K. K. Moorthy, and B. V. K. Murthy (1998), Spatial gradients in aerosol

characteristics over the Arabian Sea and Indian Ocean. *J. Geophys. Res.*, 103(D20), 26183–26192, doi:10.1029/98JD00803.

Satheesh, S. K., V. Ramanathan, X. Li-Jones, J. M. Lobert, I. A. Podgorny, J. M. Prospero, B. N. Holben, and N. G. Loeb (1999), A model for the natural and anthropogenic aerosols over the tropical Indian Ocean derived from Indian Ocean Experiment data. *J. Geophys. Res.*, 104(D22), 27421–27440, doi:10.1029/1999JD900478.

Saunders, C. P. R. (1993), A Review of Thunderstorm Electrification Processes. *J. Appl. Meteor.*, 32, 642–655, doi:http://dx.doi.org/10.1175/15200450(1993)032<0642:AROTEP>2.0.CO;2.

Schumann, U. and H. Huntrieser (2007), The global lightning-induced nitrogen oxides source. *Atmos. Chem. Phys.*, 7, 3823–3907.

Sheffield, A. M., S. M. Saleeby, and S. C. van den Heever (2015), Aerosol-induced mechanisms for cumulus congestus growth, *J. Geophys. Res. Atmos.*, 120, 8941–8952, doi:10.1002/2015JD023743.

Sherwood, S. C., V. T. J. Phillips, and J. S. Wettlaufer (2006), Small ice crystals and the climatology of lightning. *Geophys. Res. Lett.*, 33, L05804, doi:10.1029/2005GL025242.

Spracklen, D. V., K. S. Carslaw, U. Poschl, A. Rap, and P. M. Forster (2011), Global cloud condensation nuclei influenced by carbonaceous combustion aerosol. *Atmos. Chem. Phys.*, 11, 9067–9087, doi:10.5194/acp-11-9067-2011.

Squires, P. (1958), The microstructure and colloidal stability of warm clouds, *Tellus*, 10, 256– 271.

Stephens, G. L., P. J. Webster, R. H. Johnson, R. Engelen, and T. S. L'Ecuyer (2004),

Observational evidence for the mutual regulation of the tropical hydrological cycle and
tropical sea surface temperatures. *J. Climate*, 17, 2213–2224, doi:
http://dx.doi.org/10.1175/1520-0442(2004)017<2213:OEFTMR>2.0.CO;2.

Stevens, B., and G. Feingold (2009), Untangling aerosol effects on clouds and
precipitation in a buffered system. *Nature*, 461(1), doi:10.1038/nature08281.

Stolz, D. C., S. Businger, and A. Terpstra (2014), Refining the relationship between lightning
and convective rainfall over the ocean, *J. Geophys. Res. Atmos.*, 119, 964–981,
doi:10.1002/2012JD018819.

Stolz, D. C., S. A. Rutledge, and J. R. Pierce (2015), Simultaneous influences of
thermodynamics and aerosols on deep convection and lightning in the Tropics, *J.
Geophys. Res. Atmos.*, 120, 6207–6231, doi: 10.1002/2014JD023033.

Storer, R. L., S. C. van den Heever, and G. L. Stephens (2010), Modeling aerosol impacts on
convective storms in different environments. *J. Atmos. Sci.*, 67, 3904–3915. doi:
10.1175/2010JAS3363.1.

Storer, R. L. and S. C. van den Heever (2013), Microphysical processes evident in aerosol
forcing of tropical deep convective clouds. *J. Atmos. Sci.*, 70, 430–446. doi:
10.1175/JAS-D-12-076.1.

Storer, R. L. and S. C. van den Heever, and T. S., L'Ecuyer (2014) Observations of aerosol-
induced convective invigoration in the tropical east Atlantic. *J. Geophys. Res. Atmos.*,
119, 3963–3975, doi:10.1002/2013JD020272.

Straub, K. H. (2013), MJO initiation in the real-time multivariate MJO index. *J. Climate*, 26,
1130–1151, doi: 10.1175/JCLI-D-12-00074.1.

Szoke, E. J., E. J. Zipser, and D. P. Jorgensen (1986), A radar study of convective cells in

mesoscale systems in GATE. Part I: Vertical profile statistics and comparison with hurricanes. *J. Atmos. Sci.*, 43(2), 182–197.

Takahashi, T. (1978), Riming electrification as a charge generation mechanism in thunderstorms. *J. Atmos. Sci.*, 35, 1536–1548.

Tao, W.-K., J.-P Chen, Z. Li, C. Wang, and C. Zhang (2012), Impact of aerosols on convective clouds and precipitation, *Rev. Geophys.*, 50, RG2001, doi:10.1029/2011RG000369.

Tian, B., D. E. Waliser, R. A. Kahn, Q. Li, Y. L. Yung, T. Tyranowski, I. V. Geogdzhayev, M. I. Mishchenko, O. Torres, and A. Smirnov (2008), Does the Madden-Julian Oscillation influence aerosol variability? *J. Geophys. Res.*, 113, D12215, doi:10.1029/2007JD009372.

Tian, B., D. E. Waliser, R. A. Kahn, and S. Wong (2011), Modulation of Atlantic aerosols by the Madden-Julian Oscillation, *J. Geophys. Res.*, 116, D15108, doi:10.1029/2010JD015201.

Tompkins, A. M. (2001), Organization of tropical convection in low vertical wind shears: The role of cold pools. *J. Atmos. Sci.*, 58, 1650–1672.

van den Heever, S. C., G. G. Carrió, W. R. Cotton, P. J. DeMott, and A. J. Prenni (2006), Impacts of nucleating aerosol on Florida storms. Part I: Mesoscale simulations. *J. Atmos. Sci.*, 63, 1752–1775, doi: http://dx.doi.org/10.1175/JAS3713.1.

van den Heever, S. C. and W. R. Cotton (2007), Urban aerosol impacts on downwind convective storms. *J. App. Meteor. Clim.*, 46, 828–850.

van den Heever, S. C., G. L. Stephens, and N. B. Wood (2011), Aerosol indirect effects on tropical convection characteristics under conditions of radiative–convective equilibrium, *J. Atmos. Sci.*, 68, 699–718.

van der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, M. Mu, P. S. Kasibhatla, D. C.

Morton, R. S. DeFries, Y. Jin, and T. T. Van Leeuwen (2010), Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. and Phys.*, 10(23), 11707–11735, doi:10.5194/acp-10-11707-2010.

Várnai, T. and A. Marshak, (2009), MODIS observations of enhanced clear sky reflectance near clouds. *Geophys. Res. Let.*, 36, L06807, doi:10.1029/2008GL037089.

Várnai, T., A. Marshak, and W. Yang (2013), Multi-satellite aerosol observations in the vicinity of clouds. *Atmos. Chem. Phys.*, 13, 3899–3908, doi:10.5194/acp-13-3899-2013.

Venevsky, S. (2014), Importance of aerosols for annual lightning production at global scale. *Atmos. Chem. Phys. Discuss.*, 14, 4303–4325.

Vesperini, M. (2002), ECMWF analyses of humidity: Comparisons to POLDER estimates over land, *Remote Sens. Environ.*, 82, 469–480. doi:10.1016/S0034-4257(02)00070-6.

Wall, C., E. J. Zipser, and C. Liu (2014), An investigation of the aerosol indirect effect on convective intensity using satellite observations. *J. Atmos. Sci.*, 71, 430–447.

Wang, B. and H. Rui (1990), Synoptic climatology of transient tropical intraseasonal convection anomalies: 1975-1985. *Meteorol. Atmos. Phys.*, 44, 43-61.

Wang, C. (2005), A modeling study of the response of tropical deep convection to the increase of cloud condensation nuclei concentration: 1. Dynamics and microphysics, *J. Geophys. Res.*, 110, D21211, doi:10.1029/2004JD005720.

Webster, P. J. and R. Lukas (1992), TOGA-COARE: The coupled ocean-atmosphere response experiment. *Bull. Amer. Met. Soc.*, 73(9), 1377–1416, doi: http://dx.doi.org/10.1175/1520-0477(1992)073<1377:TCTCOR>2.0.CO;2.

Wheeler, M. and H. H. Hendon (2004), An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, 132,

1917–1932, doi:

http://dx.doi.org/10.1175/15200493(2004)132<1917:AARMMI>2.0.CO;2.

Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences*, 3rd ed. Academic Press,

Oxford, UK, 704 pp.

Williams, E. R. and G. Sátori (2004), Lightning, thermodynamic and hydrological comparison of

the two tropical continental chimneys, *J. Atmos. and Sol.-Ter. Phys.*, 66, 1213–1231, doi:

10.1016/j.jastp.2004.05.015.

Williams, E. R. and N. Renno (1993), An analysis of conditional instability of the tropical

atmosphere. *Mon. Wea. Rev.*, 121, 21–36.

Williams, E. R., R. Zhang, and J. Rydock (1991), Mixed-phase microphysics and cloud

electrification. *J. Atmos. Sci.*, 48(19), 2195–2203.

Williams, E. R., and S. Stanfill (2002), The physical origin of the land-ocean contrast in

lightning activity. *C. R. Phys.*, 3, 1277–1292.

Williams, E. R., et al. (1992), A radar and electrical study of tropical "hot towers", *J. Atmos. Sci.*,

49, 1386–1395.

Williams, E. R., et al. (2002), Contrasting convective regimes over the Amazon: Implications

for cloud electrification. *J. Geophys. Res.*, 107(D20), 8082, doi:10.1029/2001JD000380.

Williams, E. R., T. V. Mushtak, D. Rosenfeld, S. Goodman, and D. Boccippio (2005),

Thermodynamic conditions favorable to superlative thunderstorm updraft, mixed phase

microphysics and lightning flash rate. *Atmos. Res.*, 76, 288–306,

doi:10.1016/j.atmosres.2004.11.009.

Xu, W. and S. A. Rutledge (2014), Convective characteristics of the Madden–Julian Oscillation

over the central Indian Ocean observed by shipborne radar during DYNAMO. *J. Atmos. Sci.*, 71, 2859–2877, doi: 10.1175/JAS-D-13-0372.1.

Xu, W. and S. A. Rutledge (2015), Morphology, intensity, and rainfall production of MJO convection: Observations from DYNAMO shipborne radar and TRMM. *J. Atmos. Sci.*, 72, 623–640, doi: 10.1175/JAS-D-14-0130.1.

Yamagata, T. and Y. Hayashi (1984), A simple diagnostic model for the 30-50 day oscillation in the Tropics. *J. Met. Soc. Jap.*, 62(5), 709–717.

Yoneyama, K., C. Zhang, and C.N. Long (2013), Tracking pulses of the Madden-Julian Oscillation, *Bull. Amer. Meteor. Soc.*, 94, 1871–1891, doi:10.1175/BAMS-D-12-00157.1.

Yuan, T., L. A. Remer, K. E. Pickering, and H. Yu (2011), Observational evidence of aerosol enhancement of lightning activity and convective invigoration, *Geophys. Res. Lett.*, 38, L04701, doi:10.1029/2010GL046052.

Yuan, T., L. A. Remer, H. Bian, J. R. Ziemke, R. Albrecht, K. E. Pickering, L. Oreopoulos, S. J. Goodman, H. Yu, and D. J. Allen (2012), Aerosol indirect effect on tropospheric ozone via lightning, *J. Geophys. Res.*, 117, D18213, doi:10.1029/2012JD017723.

Zhang, C. and M. Dong (2004), Seasonality in the Madden-Julian Oscillation. *J. Climate*, 17, 3169–3180.

Zhang, C. (2005), Madden-Julian oscillation, *Rev. Geophys.*, 43, RG2003, doi:10.1029/2004RG000158.

Zhang, C. (2013), Madden–Julian oscillation: Bridging weather and climate. *Bull. Amer. Meteor. Soc.*, 94, 1849–1870, doi: http://dx.doi.org/10.1175/BAMS-D-12-00026.1.

Zipser, E. J. and M. A. LeMone (1980), Cumulonimbus vertical velocity events in GATE. Part

II: Synthesis and model core structure. *J. Atmos. Sci.*, 37, 2458–2469, doi: http://dx.doi.org/10.1175/1520-0469(1980)037<2458:CVVEIG>2.0.CO;2.

Zipser, E. J. (1994), Deep cumulonimbus cloud systems in the Tropics with and without lightning. *Mon. Wea. Rev.*, 122, 1837–1851.

Zipser, E. J., and K. R. Lutz (1994), The vertical profile of radar reflectivity of convective cells: A strong indicator of storm intensity and lightning probability? *Mon. Wea. Rev.*, 122, 1751–1759.

Zipser, E. J. (2003), Some views on "Hot Towers" after 50 years of tropical field programs and two years of TRMM data. *Meteor. Monographs*, 29, 49–49, doi: http://dx.doi.org/10.1175/0065-9401(2003)029<0049:CSVOHT>2.0.CO;2.

Zipser, E. J., D. Cecil, C. Liu, S. Nesbitt, and D. Yorty (2006), Where are the most intense thunderstorms on Earth? *Bull. Am. Met. Soc.,* 87, 1057–1071.

APPENDIX

Appendix A.1 - Index of Terms

AVGHT30 – Average 30 dBZ echo top height

CAPE – Convective Available Potential Energy

CCN – Cloud Condensation Nucleus(ei)

CCF – Cold Cloud Feature

CF – Convective Feature

CIO – Central Indian Ocean

DMS – Dimethylsulfide

DYNAMO – Dynamics of the Madden-Julian Oscillation field campaign

ECMWF – European Center For Medium Range Weather Forecasts

EOF – Empirical Orthogonal Function

ERAi – European Centers For Medium Range Weather Forecasts Interim Reanalysis

GLD360 – Global Lightning Dataset (Vaisala)

GRL – Geophysical Research Letters

IR – Infrared

JGR-A – Journal of Geophysical Research-Atmospheres

LCL – Lifted-Condensation Level

LIS – Lightning Imaging Sensor onboard the TRMM satellite

LPCF – Lightning-Producing Convective Feature

MAXHT30 – Maximum 30 dBZ echo top height

MERRA – Modern Era Retrospective Analysis for Research and Applications

MJO – Madden-Julian Oscillation

N40 – Concentration of boundary-layer cloud condensation nuclei with diameters > 40 nm

NCAPE – Normalized Convective Available Potential Energy

PR – Precipitation Radar onboard the TRMM satellite

R08 – Rosenfeld et al., 2008

TLD – Total Lightning Density

TRMM – Tropical Rainfall Measuring Mission

VPRR – Vertical Profiles of Radar Reflectivity

WCD – Warm-Cloud Depth

Appendix A.2 - Statistical output – Supplement to Sec. 2.5

The overarching objective of the analysis in Sec. 2.5 was to discern the relative influence of the each predictor within independent predictor set on TLD and AVGHT30 using multiple linear regression analysis. A forward-regression analysis strategy was employed, whereby multiple iterations of the regression computations are conducted, each time adding a variable (or transformed variable) and assessing changes in regression performance. The optimal model for the data is found when the addition of new variables doesn't contribute improvements in regression strength beyond some pre-determined tolerance [*Montgomery et al.*, 2012]. In general, thirteen model forms consisting of linear combinations of NCAPE, CCN, WCD, and physically-explicable transformations of the independent variables were studied. The first twelve of these models were based on linear, quadratic, interaction, and logarithmic forms of the predictor set.

With the inclusion of higher-order terms in the regression model, it became apparent that increases in the multiple correlation between the dependent response variable (i.e., TLD and AVGHT30) and the independent predictor set were possible when including higher-order terms in the regression form. However, a consequence of including higher-order terms (e.g., NCAPE and $NCAPE^2$) in the same model was that multicollinearity within the predictor set led to decreased stability of individual parameter estimates. For completeness, the thirteenth model form was an experimental form that combined "random" combinations of linear, quadratic, or logarithmic forms of the independent variables (NCAPE, CCN, and WCD) in an effort to see if there was an optimal fit/significant increase in the multiple correlation using transformations of predictor variables.

In each statistical output table, several identifying characteristics are laid out in the upper left corner. The name of the geographical and temporal subset, the form of the predictand (0-TLD; 1-AVGHT30; log Y implies that the logarithmic transform of the predictand was used in regression), the minimum number of samples within a given subset of the three-dimensional parameter space before inclusion of that data point within the regression computation ("Bin minimum"), and the total number of samples (i.e., the number of subsets of the three-dimensional parameter space that were included in the regression computation) are shown. The forms of each regression for models 1-13 are shown in the upper center/right of the statistical output table. Note that regression form 13 varies according to which combination of transformed predictor variables explained the most amount of variance in the response.

Each statistical output table is subsequently organized by column according to the regression model form studied (the top row of the statistical output table is the regression number corresponding to entries in the upper right of each diagram). Note that for each regression model form (i.e., each column), each row for a given parameter corresponds to the quantity for NCAPE, CCN, and WCD, in that order. For the higher-order models including interaction, quadratic, and/or logarithmic transformations, the order of individual variables again follows the order of NCAPE, CCN, and WCD. The order for interaction terms is NCAPE·CCN, NCAPE·WCD, CCN·WCD.

Multicollinearity statistics depict the degree of linear dependence in the predictor set, i.e., the likelihood that a given predictor can be represented by a linear combination of the remaining predictors in the set. The two rows that follow beneath the multicollinearity statistics consist of the multiple correlation and adjusted multiple correlation (i.e., R^2 and R_{adj}^2, respectively) for the total regression to illustrate the "goodness of fit". The adjusted multiple correlation is meant to

express the "goodness of fit" while accounting for the degrees of freedom in the regression and residuals. Individual coefficients of determination (i.e., Individual R^2) for regression of the predictand onto each predictor in succession have been provided to illustrate potential differences in the "goodness of fit" between individual variables and the predictand vs. between the full predictor set and the predictand. Parameter estimates, i.e., the coefficients of the regression assigned to individual predictors, and the regression constant (i.e., β_0 term in Eq. 2.1) follow next. Mean-square error (MSE) is meant to quantify the accuracy of each model and in theory, this quantity should decrease with improving "goodness of fit". The degrees of freedom for the regression (ν_{reg}) and for the residuals (ν_{error}) are provided along with the F-statistic; these quantities are used in evaluating the null hypothesis for the regression (i.e., $\beta_p = 0$) according to the theoretical F-statistics for significance levels, p=0.10, 0.05, and 0.01. Lastly, the two-tailed Student's t-test results are shown to evaluate whether individual parameter estimates in the regression are statistically significant; these values represent the probability that each t-statistic for a given parameter estimate is not random and values approaching 1.0 are more favorable. Note that these values are negatively correlated with the multicollinearity statistics near the top of the statistical output table; as linear dependence in the predictor set increases, the statistical significance of individual parameter estimates decreases.

For each predictand (e.g., TLD and AVGHT30), statistics for global, continental, oceanic, hemispheric, and regional data subsets on the annual time scale are followed by hemispheric data subsets for individual seasons. Statistical output tables for TLD precede the statistical output tables for AVGHT30 and each table can be identified using the information in the upper left corner of each diagram. These tables represent the statistics *for logarithmic forms*

of the predictand and the results for linear forms of the predictand have been withheld for reasons described in Sec. 2.5. See below for statistical output tables.

Location,Season: Globe, ANNUAL

Predictand: 0 log Y

Bin minimum: 20

Samples: 885

Regression Forms:

1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, log$_{10}$CCN, WCD (optimal)

Regression Number: 1 – 13

Multicollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

v_{est}; v_{gust}; Computed F-Statistic

Theoretical F-Statistic (p=0.10,0.05,0.01)

Two-sided t-test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t-value)

Location,Season: Continents, ANNUAL

Predictand: 0 log Y

Bin minimum: 20

Samples: 716

Regression Forms:

1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, log$_{10}$CCN, WCD (optimal)

Regression Number: 1 – 13

Multicollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

v_{est}; v_{gust}; Computed F-Statistic

Theoretical F-Statistic (p=0.10,0.05,0.01)

Two-sided t-test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t-value)

Location,Season: Oceans, ANNUAL

Predictand: 0 log Y

Bin minimum: 20

Samples: 615

Regression Forms:

1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, log$_{10}$CCN, WCD (optimal)

Regression Number: 1 – 13

Multicollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

v_{est}; v_{gust}; Computed F-Statistic

Theoretical F-Statistic (p=0.10,0.05,0.01)

Two-sided t-test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t-value)

Location,Season: Northern Hemi. (Continents), ANNUAL

Predictand: 0 log Y

Bin minimum: 20

Samples: 631

Regression Forms:

1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, log$_{10}$CCN, WCD (optimal)

Regression Number: 1 – 13

Multicollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

v_{est}; v_{gust}; Computed F-Statistic

Theoretical F-Statistic (p=0.10,0.05,0.01)

Two-sided t-test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t-value)

Top-left panel

Location,Season: Southern Hemi. (Continents), ANNUAL

Predictand: 0 log Y

Bin minimum: 20

Samples: 535

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE', log₁₀CCN, WCD' (optimal)

Regression Number:

	1	2	3	4	5	6	7	8	9	10	11	12	13

Multicollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_i

Regression Constant, b_0

Mean Square Error

$v_{est}; v_{error};$ Computed F–Statistic

Theoretical F–Statistic (p=0.10,0.05,0.01)

Two–sided t–test: Significance of b_i (probability that the absolute value of a random variate will be less than parameter t–value)

Top-right panel

Location,Season: Northern Hemi. (Oceans), ANNUAL

Predictand: 0 log Y

Bin minimum: 20

Samples: 522

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE', log₁₀CCN, WCD' (optimal)

Regression Number:

	1	2	3	4	5	6	7	8	9	10	11	12	13

Multicollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_i

Regression Constant, b_0

Mean Square Error

$v_{est}; v_{error};$ Computed F–Statistic

Theoretical F–Statistic (p=0.10,0.05,0.01)

Two–sided t–test: Significance of b_i (probability that the absolute value of a random variate will be less than parameter t–value)

Bottom-left panel

Location,Season: Southern Hemi. (Oceans), ANNUAL

Predictand: 0 log Y

Bin minimum: 20

Samples: 444

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE', log₁₀CCN, WCD' (optimal)

Regression Number:

	1	2	3	4	5	6	7	8	9	10	11	12	13

Multicollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_i

Regression Constant, b_0

Mean Square Error

$v_{est}; v_{error};$ Computed F–Statistic

Theoretical F–Statistic (p=0.10,0.05,0.01)

Two–sided t–test: Significance of b_i (probability that the absolute value of a random variate will be less than parameter t–value)

Bottom-right panel

Location,Season: Amazon, ANNUAL

Predictand: 0 log Y

Bin minimum: 10

Samples: 202

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE', log₁₀CCN, WCD' (optimal)

Regression Number:

	1	2	3	4	5	6	7	8	9	10	11	12	13

Multicollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_i

Regression Constant, b_0

Mean Square Error

$v_{est}; v_{error};$ Computed F–Statistic

Theoretical F–Statistic (p=0.10,0.05,0.01)

Two–sided t–test: Significance of b_i (probability that the absolute value of a random variate will be less than parameter t–value)

174

Location,Season: Central America Offshore, ANNUAL
Predictand: 0 log Y
Bin minimum: 10
Samples: 58

Regression Forms:
1: NCAPR
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPR, $\log_{10}$CCN, $\log_{10}$WCD (optimal)

Location,Season: Central Indian Ocean, ANNUAL
Predictand: 0 log Y
Bin minimum: 10
Samples: 51

Regression Forms:
1: NCAPR
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPR, CCN, $\log_{10}$WCD (optimal)

Location,Season: Congo, ANNUAL
Predictand: 0 log Y
Bin minimum: 10
Samples: 137

Regression Forms:
1: NCAPR
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPR, $\log_{10}$CCN, WCD (optimal)

Location,Season: Gulf of Guinea, ANNUAL
Predictand: 0 log Y
Bin minimum: 10
Samples: 58

Regression Forms:
1: NCAPR
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPR, $\log_{10}$CCN, $\log_{10}$WCD (optimal)

Location,Season: Gulf Stream, ANNUAL

Predictand: 0 log Y
Bin minimum: 10
Samples: 87

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, $\log_{10}$CCN, WCD (optimal)

Regression Number: 1–13

Sections: Multicollinearity; R^2, R_{adj}^2; Individual R^2; Parameter estimates, b_i; Regression Constant, b_0; Mean Square Error; v_{reg}; v_{err}; Computed F-Statistic; Theoretical F-Statistic ($p=0.10,0.05,0.01$); Two-sided t-test: Significance of b_i (probability that the absolute value of a random variate will be less than parameter t-value)

Location,Season: India, ANNUAL

Predictand: 0 log Y
Bin minimum: 10
Samples: 25

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, $\log_{10}$CCN, WCD (optimal)

Location,Season: South America Offshore, ANNUAL

Predictand: 0 log Y
Bin minimum: 10
Samples: 29

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: $\log_{10}$NCAPE, CCN, $\log_{10}$WCD (optimal)

Location,Season: United States, ANNUAL

Predictand: 0 log Y
Bin minimum: 10
Samples: 128

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, $\log_{10}$CCN, WCD (optimal)

176

Location,Season: Warm Pool, ANNUAL

Predictand: 0 log Y

Bin minimum: 10

Samples: 54

Regression Forms:
- 1: NCAPR
- 2: CCN
- 3: WCD
- 4: Linear
- 5: Interaction
- 6: Quadratic
- 7: Logarithmic
- 8: Linear,Interaction
- 9: Linear,Quadratic
- 10: Interaction,Quadratic
- 11: Linear,Interaction,Quadratic
- 12: Linear,Interaction,Quadratic,Logarithmic
- 13: $\log_{10}$NCAPR, $\log_{10}$CCN, WCD* (optimal)

Location,Season: South Africa Offshore, ANNUAL

Predictand: 0 log Y

Bin minimum: 10

Samples: 22

Regression Forms:
- 1: NCAPR
- 2: CCN
- 3: WCD
- 4: Linear
- 5: Interaction
- 6: Quadratic
- 7: Logarithmic
- 8: Linear,Interaction
- 9: Linear,Quadratic
- 10: Interaction,Quadratic
- 11: Linear,Interaction,Quadratic
- 12: Linear,Interaction,Quadratic,Logarithmic
- 13: NCAPR*, $\log_{10}$CCN, WCD* (optimal)

Location,Season: Northern Hemi. (Continents), JJA

Predictand: 0 log Y

Bin minimum: 20

Samples: 439

Regression Forms:
- 1: NCAPR
- 2: CCN
- 3: WCD
- 4: Linear
- 5: Interaction
- 6: Quadratic
- 7: Logarithmic
- 8: Linear,Interaction
- 9: Linear,Quadratic
- 10: Interaction,Quadratic
- 11: Linear,Interaction,Quadratic
- 12: Linear,Interaction,Quadratic,Logarithmic
- 13: NCAPR*, $\log_{10}$CCN, WCD* (optimal)

Location,Season: Southern Hemi. (Continents), DJF

Predictand: 0 log Y

Bin minimum: 20

Samples: 413

Regression Forms:
- 1: NCAPR
- 2: CCN
- 3: WCD
- 4: Linear
- 5: Interaction
- 6: Quadratic
- 7: Logarithmic
- 8: Linear,Interaction
- 9: Linear,Quadratic
- 10: Interaction,Quadratic
- 11: Linear,Interaction,Quadratic
- 12: Linear,Interaction,Quadratic,Logarithmic
- 13: NCAPR*, $\log_{10}$CCN, WCD (optimal)

Location,Season: Northern Hemi. (Oceans), JJA

Predictand: 0 log Y

Bin minimum: 20

Samples: 287

1: NCAPR
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPR, $\log_{10}$CCN, $\log_{10}$WCD (optimal)

Regression Number:

Multicollinearity:

R^2,R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

v_{reg}; v_{error}; Computed F–Statistic

Theoretical F–Statistic (p=0.10,0.05,0.01)

Two–sided t–test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t–value)

Location,Season: Southern Hemi. (Oceans), DJF

Predictand: 0 log Y

Bin minimum: 20

Samples: 204

1: NCAPR
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPR, $\log_{10}$CCN, WCD* (optimal)

Regression Number:

Multicollinearity:

R^2,R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

v_{reg}; v_{error}; Computed F–Statistic

Theoretical F–Statistic (p=0.10,0.05,0.01)

Two–sided t–test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t–value)

Location,Season: Globe, ANNUAL

Predictand: 1 log Y

Bin minimum: 30

Samples: 1052

1: NCAPR
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPR, $\log_{10}$CCN, WCD (optimal)

Regression Number:

Multicollinearity:

R^2,R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

v_{reg}; v_{error}; Computed F–Statistic

Theoretical F–Statistic (p=0.10,0.05,0.01)

Two–sided t–test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t–value)

Location,Season: Continents, ANNUAL

Predictand: 1 log Y

Bin minimum: 30

Samples: 844

1: NCAPR
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPR, $\log_{10}$CCN, WCD (optimal)

Regression Number:

Multicollinearity:

R^2,R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

v_{reg}; v_{error}; Computed F–Statistic

Theoretical F–Statistic (p=0.10,0.05,0.01)

Two–sided t–test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t–value)

Location,Season: Northern Hemi. (Continents), ANNUAL

Predictand: 1 log Y

Bin minimum: 30

Samples: 736

Regression Number:

1: NCAPR
2: CCN
3: WCD
4: Linear

Regression Forms:

5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPR, log₁₀CCN, WCD (optimal)

Multicollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

v_{reg}, v_{error}; Computed F-Statistic

Theoretical F-Statistic (p=0.10,0.05,0.01)

Two-sided t-test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t-value)

Location,Season: Southern Hemi. (Continents), ANNUAL

Predictand: 1 log Y

Bin minimum: 30

Samples: 634

Regression Number:

1: NCAPR
2: CCN
3: WCD
4: Linear

Regression Forms:

5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPR, log₁₀CCN, WCD (optimal)

Location,Season: Northern Hemi. (Oceans), ANNUAL

Predictand: 1 log Y

Bin minimum: 30

Samples: 711

Regression Number:

1: NCAPR
2: CCN
3: WCD
4: Linear

Regression Forms:

5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPR, log₁₀CCN, WCD (optimal)

Location,Season: Southern Hemi. (Oceans), ANNUAL

Predictand: 1 log Y

Bin minimum: 30

Samples: 580

Regression Number:

1: NCAPR
2: CCN
3: WCD
4: Linear

Regression Forms:

5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPR, log₁₀CCN, WCD' (optimal)

Amazon, ANNUAL

Location,Season: Amazon, ANNUAL
Predictand: 1 log Y
Bin minimum: 20
Samples: 278

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, log$_{10}$CCN, log$_{10}$WCD (optimal)

Regression Number:

	1	2	3	4	5	6	7	8	9	10	11	12	13
Multicollinearity:													
NaN	NaN	NaN	NaN	0.0971	0.8633	0.0739	0.0978	0.9950	0.9314	0.8834	0.9959	0.9976	0.1273
	−777.0	−777.0	−777.0	0.1506	0.4519	0.0769	0.2105	0.9939	0.8813	0.9416	0.9967	0.9987	0.2250
	−777.0	−777.0	−777.0	0.1818	0.3348	0.1192	0.2268	0.8002	0.9975	0.9624	0.9984	0.9996	0.2487
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.8870	0.9309	0.9386	0.9127	0.9149	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9951	0.8731	0.9727	0.9963	0.9963	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9940	0.9975	0.2944	0.9965	0.9967	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9396	0.9868	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9070	0.9496	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9983	0.9999	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9753	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9084	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9999	−777.0
R^2,R_{adj}^2													
	0.1600	0.5946	0.1440	0.6501	0.5261	0.5236	0.7095	0.6573	0.7191	0.7144	0.7241	0.7373	0.7163
	0.1570	0.5931	0.1409	0.6463	0.5209	0.5184	0.7063	0.6498	0.7129	0.7081	0.7149	0.7254	0.7132
Individual R^2													
	0.1600	0.5946	0.1440	0.1700	0.4603	0.1479	0.1700	0.1700	0.4603	0.1700	0.1700	0.1700	0.1700
	−777.0	−777.0	−777.0	0.4672	0.1222	0.2700	0.5946	0.4672	0.1222	0.4672	0.4672	0.4672	0.5946
	−777.0	−777.0	−777.0	0.1464	0.4504	0.1481	0.1440	0.1464	0.4504	0.1464	0.1464	0.1464	0.1440
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.4603	0.1479	0.4603	0.4603	0.4603	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.1222	0.2700	0.1222	0.1222	0.1222	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.4504	0.1481	0.4504	0.4504	0.4504	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.1479	0.1479	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.2700	0.2700	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.1481	0.1481	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.1600	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.0946	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.1440	−777.0
Parameter estimates, b_p													
	0.4001	0.7711	−0.379	0.4143	0.0354	0.4420	0.3473	0.5428	0.4190	0.1648	0.1337	−0.376	0.3640
	−777.0	−777.0	−777.0	0.5320	0.2647	0.3880	0.6460	−0.447	1.2372	0.4038	1.2285	−1.9985	0.6296
	−777.0	−777.0	777.0	0.302	0.6101	0.394	0.190	−0.383	0.196	1.0331	0.268	−3.852	0.212
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−0.129	−0.022	−0.060	0.2238	0.2857	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−0.091	−0.733	−0.692	0.2547	0.3215	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	1.0610	−0.039	−0.392	−0.086	−0.712	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−0.065	0.1429	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−0.811	−0.478	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.0050	1.7923	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.2120	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.3673	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	1.8915	−777.0
Regression Constant, b_0													
	−7.8e−15	−9.4e−15	−1.2e−15	−9.0e−15	−8.7e−15	−8.4e−15	−4.7e−15	−8.4e−15	−9.1e−15	−8.6e−15	−9.7e−15	−5.4e−15	−4.9e−15
Mean Square Error													
	0.8429	0.4068	0.8590	0.3536	0.4790	0.4815	0.2936	0.3501	0.2870	0.2918	0.2850	0.2745	0.2792
$v_{reg}; v_{error}$; Computed F−Statistic													
	1.0000	1.0000	1.0000	3.0000	3.0000	3.0000	3.0000	6.0000	6.0000	6.0000	9.0000	12.000	3.0000
	276.00	276.00	276.00	274.00	274.00	274.00	274.00	271.00	271.00	271.00	268.00	265.00	274.00
	52.609	404.86	46.464	169.76	101.39	100.41	223.10	86.668	115.64	112.99	78.179	61.994	230.61
Theoretical F−Statistic (p=0.10,0.05,0.01)													
	2.7238	2.7238	2.7238	2.1038	2.1038	2.1038	2.1038	1.7859	1.7859	1.7859	1.6550	1.5707	2.1038
	3.8753	3.8753	3.8753	2.6375	2.6375	2.6375	2.6375	2.1321	2.1321	2.1321	1.9149	1.7888	2.6375
	6.7275	6.7275	6.7275	3.8538	3.8538	3.8538	3.8538	2.8690	2.8690	2.8690	2.4740	2.2523	3.8538
Two−sided t−test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t−value)													
	1.0000	1.0000	1.0000	1.0000	0.1725	1.0000	0.9999	0.4731	0.9311	0.6501	0.1124	0.2372	1.0000
	−777.0	−777.0	777.0	1.0000	0.9987	1.0000	1.0000	0.4377	1.0000	0.8943	0.7566	0.7428	1.0000
	−777.0	−777.0	−777.0	0.9999	0.9999	1.0000	0.9943	0.9953	0.1333	1.0000	0.1415	0.9987	0.9975
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.5271	0.0792	0.1975	0.7280	0.8334	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.0846	0.9999	0.9999	0.9999	0.2522	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.8250	0.0267	0.9999	0.0771	0.4986	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.2117	0.2147	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9999	0.9246	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.0000	0.6894	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.4199	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.9344	−777.0
	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	−777.0	0.0986	−777.0

Central America Offshore, ANNUAL

Location,Season: Central America Offshore, ANNUAL
Predictand: 1 log Y
Bin minimum: 20
Samples: 159

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, log$_{10}$CCN, log$_{10}$WCD (optimal)

Regression Number:

	1	2	3	4	5	6	7	8	9	10	11	12	13
Multicollinearity:													
NaN	NaN	NaN	NaN	0.1257	0.8977	0.0489	0.1778	0.9986	0.9354	0.9290	0.9987	0.9989	0.1886
R^2,R_{adj}^2													
	0.4642	0.3777	0.0063	0.6490	0.6236	0.6498	0.5976	0.5617	0.7130	0.7043	0.7331	0.7718	0.7260
	0.4608	0.3737	−2.101	0.6422	0.6163	0.6430	0.5898	0.6483	0.7016	0.6926	0.7169	0.7530	0.7207
Mean Square Error													
	0.5391	0.6262	1.0000	0.3577	0.3836	0.3569	0.4101	0.3516	0.2983	0.3073	0.2830	0.2469	0.2792

Central Indian Ocean, ANNUAL

Location,Season: Central Indian Ocean, ANNUAL
Predictand: 1 log Y
Bin minimum: 20
Samples: 150

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, log$_{10}$CCN, log$_{10}$WCD (optimal)

Regression Number:

	1	2	3	4	5	6	7	8	9	10	11	12	13
Multicollinearity:													
NaN	NaN	NaN	NaN	0.0269	0.8554	0.0145	0.0411	0.9979	0.9373	0.8580	0.9981	0.9989	0.0497
R^2,R_{adj}^2													
	0.3460	0.3052	0.0031	0.5711	0.5135	0.5255	0.5255	0.5895	0.5969	0.5795	0.6087	0.6710	0.6165
	0.3416	0.3005	−0.003	0.5623	0.5443	0.5158	0.5833	0.5722	0.5696	0.5619	0.5835	0.6422	0.6086
Mean Square Error													
	0.6583	0.6994	1.0035	0.4376	0.4556	0.4841	0.4166	0.4277	0.4303	0.4380	0.4164	0.3577	0.3913

Congo, ANNUAL

Location,Season: Congo, ANNUAL
Predictand: 1 log Y
Bin minimum: 20
Samples: 167

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, log$_{10}$CCN, WCD² (optimal)

Regression Number:

	1	2	3	4	5	6	7	8	9	10	11	12	13
Multicollinearity:													
NaN	NaN	NaN	NaN	0.1193	0.8351	0.1145	0.1172	0.9968	0.9237	0.8445	0.9980	0.9984	0.1145
R^2,R_{adj}^2													
	0.0402	0.3668	0.0100	0.4158	0.0320	0.3726	0.4435	0.4333	0.4363	0.4685	0.4513	0.4815	0.4900
	0.0344	0.3630	0.0439	0.4051	0.2891	0.3611	0.4333	0.4333	0.4090	0.4486	0.4326	0.4518	0.4806
Mean Square Error													
	0.9655	0.6369	0.9560	0.5948	0.7108	0.6388	0.5666	0.5909	0.5513	0.5673	0.5481	0.5014	0.5193

Location,Season: Gulf of Guinea, ANNUAL

Predictand: 1 log Y

Bin minimum: 20

Samples: 113

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, $\log_{10}$CCN, $\log_{10}$WCD (optimal)

Regression Number: 1–13

Multicollinearity:
R^2,R_{adj}^2:
Individual R^2:
Parameter estimates, b_p:
Regression Constant, b_0:
Mean Square Error
v_{est}; v_{oerr}; Computed F–Statistic
Theoretical F–Statistic (p=0.10,0.05,0.01)
Two–sided t–test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t–value)

Location,Season: Gulf Stream, ANNUAL

Predictand: 1 log Y

Bin minimum: 20

Samples: 197

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, CCN, $\log_{10}$WCD (optimal)

Regression Number: 1–13

Location,Season: India, ANNUAL

Predictand: 1 log Y

Bin minimum: 20

Samples: 75

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE², CCN, WCD² (optimal)

Regression Number: 1–13

Location,Season: South America Offshore, ANNUAL

Predictand: 1 log Y

Bin minimum: 20

Samples: 126

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE², $\log_{10}$CCN, WCD² (optimal)

Regression Number: 1–13

Location,Season: United States, ANNUAL

Predictand: 1 log Y

Bin minimum: 20

Samples: 180

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPE', log₁₀CCN, WCD' (optimal)

Regression Number:

1	2	3	4	5	6	7	8	9	10	11	12	13

[Dense numerical data table — Multicollinearity, R²,R²adj, Individual R², Parameter estimates b₀, Regression Constant b₀, Mean Square Error, vᵣₑₗ;vₑᵣᵣₒᵣ; Computed F-Statistic, Theoretical F-Statistic (p=0.10,0.05,0.01), Two-sided t-test]

Location,Season: Warm Pool, ANNUAL

Predictand: 1 log Y

Bin minimum: 20

Samples: 161

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPE, log₁₀CCN, WCD (optimal)

Location,Season: South Africa Offshore, ANNUAL

Predictand: 1 log Y

Bin minimum: 20

Samples: 93

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: log₁₀NCAPE, log₁₀CCN, WCD' (optimal)

Location,Season: Northern Hemi. (Continents), JJA

Predictand: 1 log Y

Bin minimum: 30

Samples: 544

1: NCAPE
2: CCN
3: WCD
4: Linear

Regression Forms:
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction

9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic

13: NCAPE, log₁₀CCN, WCD (optimal)

Table 1 (top left)

Location,Season: Southern Hemi. (Continents), DJF

Predictand: 1 log Y

Bin minimum: 30

Samples: 505

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, CCN, WCD² (optimal)

Regression Number:

	1	2	3	4	5	6	7	8	9	10	11	12	13

Multcollinearity:

R^2, R_{adj}^2

Individual R^2

Parameter estimates, b_p

Regression Constant, b_0

Mean Square Error

$v_{reg}; v_{error};$ Computed F–Statistic

Theoretical F–Statistic (p=0.10,0.05,0.01)

Two–sided t–test: Significance of b_p (probability that the absolute value of a random variate will be less than parameter t–value)

Table 2 (top right)

Location,Season: Northern Hemi. (Oceans), JJA

Predictand: 1 log Y

Bin minimum: 30

Samples: 444

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE, $\log_{10}$CCN, $\log_{10}$WCD (optimal)

Table 3 (bottom left)

Location,Season: Southern Hemi. (Oceans), DJF

Predictand: 1 log Y

Bin minimum: 30

Samples: 350

Regression Forms:
1: NCAPE
2: CCN
3: WCD
4: Linear
5: Interaction
6: Quadratic
7: Logarithmic
8: Linear,Interaction
9: Linear,Quadratic
10: Interaction,Quadratic
11: Linear,Interaction,Quadratic
12: Linear,Interaction,Quadratic,Logarithmic
13: NCAPE², $\log_{10}$CCN, $\log_{10}$WCD (optimal)

*Appendix A.3 **-** Data availability*

The convective feature (CF) database is accessible online and may be downloaded via public http or anonymous ftp from the University of Utah server (ftp://trmm.chpc.utah.edu/pub/trmm/level_3/). Daily, six-hourly data from the ERA-Interim Reanalysis Modern Era Retrospective Analysis for Research and Applications (MERRA) were downloaded from the European Centre for Medium Range Weather Forecasts online server (http://apps.ecmwf.int/datasets/data/interim_full_daily/) and Global Modeling and Assimilation Office server (http://gmao.gsfc.nasa.gov/merra/), respectively. Level 2 orbital data from the Tropical Rainfall Measuring Mission satellite are available National Aeronautics and Space Administration (NASA) servers via the Goddard Earth Sciences Data and Information Services Center (http://disc.sci.gsfc.nasa.gov/TRMM). Information about the GEOS-Chem chemical transport model (www.geos-chem.org) with the online TOMAS aerosol microphysics module is available online (http://wiki.seas.harvard.edu/geos-chem/index.php/TOMAS_aerosol_microphysics). Data from the GEOS-Chem simulations analyzed during this research are available by special request (dstolz@atmos.colostate.edu).