ABSTRACT

Title of Document:	RELATIONSHIP BETWEEN COLUMN DENSITY AND SURFACE MIXING RATIO FOR O_3 AND NO_2: IMPLICATIONS FOR SATELLITE OBSERVATIONS AND THE IMPACTS OF VERTICAL MIXING
	Clare Marie Flynn, Doctor of Philosophy, 2016
Directed by:	Dr. Kenneth E. Pickering, Department of Atmospheric and Oceanic Science

Satellites have great potential for diagnosis of surface air quality conditions, though reduced sensitivity of satellite instrumentation to the lower troposphere currently impedes their applicability. One objective of the NASA DISCOVER-AQ project is to provide information relevant to improving our ability to relate satellite-observed columns to surface conditions for key trace gases and aerosols. In support of DISCOVER-AQ, this dissertation investigates the degree of correlation between O_3 and NO_2 column abundance and surface mixing ratio during the four DISCOVER-AQ deployments; characterize the variability of the aircraft *in situ* and model-simulated O_3 and NO_2 profiles; and use the WRF-Chem model to further investigate the role of boundary layer mixing in the column-surface connection for the Maryland 2011 deployment, and determine which of the available boundary layer schemes best captures the observations. Simple linear regression analyses suggest that O_3 partial column observations from future satellite instruments with sufficient sensitivity to the lower troposphere may be most meaningful for surface air quality under the conditions associated with the Maryland 2011 campaign, which included generally deep, convective boundary layers, the least wind shear of all four deployments, and few geographical influences on local meteorology, with exception of bay

breezes. Hierarchical clustering analysis of the *in situ* O_3 and NO_2 profiles indicate that the degree of vertical mixing (defined by temperature lapse rate) associated with each cluster exerted an important influence on the shapes of the median cluster profiles for O_3, as well as impacted the column vs. surface correlations for many clusters for both O_3 and NO_2. However, comparisons to the CMAQ model suggest that, among other errors, vertical mixing is overestimated, causing too great a column-surface connection within the model. Finally, the WRF-Chem model, a meteorology model with coupled chemistry, is used to further investigate the impact of vertical mixing on the O_3 and NO_2 column-surface connection, for an ozone pollution event that occurred on July 26-29, 2011. Five PBL schemes were tested, with no one scheme producing a clear, consistent "best" comparison with the observations for PBLH and pollutant profiles; however, despite improvements, the ACM2 scheme continues to overestimate vertical mixing.

RELATIONSHIP BETWEEN COLUMN DENSITY AND SURFACE MIXING RATIO FOR O_3 AND NO_2: IMPLICATIONS FOR SATELLITE OBSERVATIONS AND THE IMPACTS OF VERTICAL MIXING

by

Clare Marie Flynn

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Russell R. Dickerson, Chair
Adjunct Professor Kenneth E. Pickering, Co-Chair
Professor Ross J. Salawitch
Research Assistant Professor Timothy P. Canty
Professor Sheryl H. Ehrman, Dean's Representative

Acknowledgments

I first and foremost thank my advisor, Prof. Kenneth Pickering, for his guidance, flexibility, patience, and ability to provide useful feedback to point me in the right direction throughout this work. I would not have been successful during my time at the University of Maryland, and would not have had such amazing and enriching experiences as a graduate student, without such an excellent advisor. What a happy coincidence that Dr. Pickering took me on to be part of the NASA DISCOVER-AQ project! I am deeply and forever grateful to Dr. Pickering for this, as well as many of the other participants in the DISCOVER-AQ project, such as my colleagues Dr. Bryan Duncan, Dr. Melanie Follette-Cook, and Dr. Christopher Loughner of NASA Goddard; Dr. James Crawford and Dr. Gao Chen of NASA Langley; the members of NASA Langley LARGE and HSRL groups; and many, many others for including me, guiding me, and collaborating with me in such a rewarding project, which has encouraged my growth both as a scientist and as a person. I am happy to call several of the people I have met through DISCOVER-AQ friends. At first, I was unsure of participating in an air quality field campaign, but this has proven to be a seminal and wonderful period of my life; none of it is an experience I would change.

I also thank Prof. Russ Dickerson, Prof. Ross Salawitch, Prof. Tim Canty, Prof. Dale Allen, and the other members, both student and staff, of the atmospheric chemistry group here at Maryland for their guidance and helpful advice in my research. Science truly does not happen in a vacuum.

Last, but not least, I thank my family and friends for their unfailing support and encouragement during my years in graduate school. I could not have done this without their support and love. Life also does not happen in a vacuum! And is always enriched by the forging of a close connection to another living being. I have also made several deep and lasting friendships during my time at Maryland, for which I am very thankful. And let's not forget the cats! Always there to bug you when you need to work, but also to play and provide joy.

One essential lesson I have learned through my experiences while a student at Maryland:
"Pour yourself out like a fountain.
Flow into the knowledge that what you are seeking
finishes often at the start, and, with ending, begins."
-- Rainer Maria Rilke

Table of Contents

List of Acronyms and Abbreviations

ACAM	Airborne Compact Atmospheric Mapper
ACM2	Asymmetric Convective Model, version 2 PBL scheme
AGL	Above Ground Level
AMF	Air Mass Factor
AMSL	Above Mean Sea Level
BEIS	Biogenic Emissions Inventory System
BouLac	Bougeault-Lacarrére PBL scheme
CAMx	Comprehensive Air Quality Model with Extensions
CDPHE	Colorado Department of Public Health and Environment
CMAQ	Community Multiscale Air Quality model
CO	Carbon monoxide
COMMIT	NASA Chemical, Optical, and Microphysical Measurements of In-situ Troposphere trailer
CTM	Chemical transport model
DACOM	Differential Absorption Carbon Monoxide instrument
DISCOVER-AQ	Deriving Information on Surface conditions from Column and Vertically Resolved Observations Relevant to Air Quality EV-1 Mission
EPA	Environmental Protection Agency
FT	Free troposphere
GeoTASO	Geostationary Trace gas and Aerosol Sensor Optimization
GMI	NASA Global Modeling Initiative
HSRL	High Spectral Resolution Lidar
HYSPLIT	Hybrid Single Particle Lagrangian Integrated Trajectory Model
LARGE	NASA Langley Aerosol Research Group Experiment
MDE	Maryland Department of the Environment
MEGAN	Model of Emissions of Gases and Aerosols from Nature
MOVES	Motor Vehicle Emissions Simulator
MOZART	Model for Ozone And Related Tracers
MPL	MicroPulse Lidar
MU	The Millersville University of Pennsylvania
MYJ	Mellor-Yamada-Janjić PBL scheme
NASA	National Aeronautics and Space Administration
NARR	North American Regional Reanalysis
NATIVE	The Pennsylvania State University Nittany Atmospheric Trailer and Integrated Validation Experiment
NCAR	National Center for Atmospheric Research

NEI	National Emissions Inventory
NOAA	National Oceanic and Atmospheric Administration
NSERC	National Suborbital Education and Research Center
NO	Nitric oxide
NO_2	Nitrogen dioxide
NO_x	Nitrogen oxides ($= NO + NO_2$)
NO_y	Reactive nitrogen species
O_3	Ozone
OMI	Ozone Monitoring Instrument
PBL	Planetary boundary layer
PBLH	Planetary boundary layer height
PPB or PPBV	Parts per billion, by volume
PPM or PPMV	Parts per million, by volume
QNSE	Quasi-Normal Scale Elimination PBL scheme
SJV APCD	San Joaquin Valley Air Pollution Control District
TCEQ	Texas Commission on Environmental Quality
Theta	Potential temperature
UH	University of Houston
VOC	Volatile Organic Compound
WRF-ARW	Weather Research and Forecasting model, Advanced Research WRF core
WRF-Chem	WRF model with coupled Chemistry
YSU	Yonsei University PBL scheme

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

1.1: Overview of Tropospheric O_3 Formation

   Tropospheric ozone (O_3) is an Environmental Protection Agency (EPA) criteria pollutant with adverse human health effects, and a ubiquitous pollutant detrimental to human welfare (e.g., crop damage) (https://www3.epa.gov/ttn/naaqs/criteria.html). Ozone is formed from the oxidation of carbon monoxide (CO) or volatile organic compounds (VOCs) in the presence of nitrogen oxides (NO_x; $NO_x = NO + NO_2$). Sources of CO include biomass burning and fossil fuel combustion, while VOCs are emitted by trees, through combustion and evaporation of fossil fuels, and by various industries, such as petrochemical processing plants. Sources of nitric oxide (NO) also include biomass burning and fossil fuel combustion, such as from motor vehicles and electrical generating units; NO is directly emitted from these sources. Nitrogen dioxide (NO_2), in contrast, is usually not directly emitted but is formed photochemically from the reaction of NO and O_3 or the hydroperoxyl (HO_2) radical, which is itself produced photochemically by the oxidation of hydrocarbons and CO, as shown in reaction R5 below (Finlayson-Pitts and Pitts, 2000; Jacob, 1999).

Each O_3 formation mechanism is initiated by the production of the hydroxyl radical (OH):

$$O_3 + h\nu \rightarrow O_2 + O(^1D) \tag{R1}$$

$$O(^1D) + H_2O_{(g)} \rightarrow 2OH \tag{R2}$$

Ozone can then be produced from the oxidation of CO:

$$CO + OH \rightarrow CO_2 + H \tag{R3}$$

$$H + O_2 + M \rightarrow HO_2 + M \tag{R4}$$

$$HO_2 + NO \rightarrow OH + NO_2 \tag{R5}$$

$$NO_2 + h\nu \rightarrow NO + O \tag{R6}$$

$$O + O_2 + M \rightarrow O_3 \tag{R7}$$

$$CO + 2O_2 \rightarrow CO_2 + O_3 \quad\quad\quad\quad\quad\quad\quad\quad\text{(Net)}$$

Importantly, NO also reacts with O_3 to form NO_2, which leads to net O_3 depletion; thus, R5 is the rate-limiting step of the above O_3 formation mechanism, as conversion of NO to NO_2 through reaction with HO_2 favors O_3 production. Therefore, NO_2 is a critical precursor species to O_3, and is also an EPA criteria pollutant (https://www3.epa.gov/ttn/naaqs/criteria.html). Ozone can also be formed from the oxidation of VOCs:

$$RH + OH \rightarrow H_2O_{(g)} + R \quad\quad\quad\quad\quad\quad\quad\quad\text{(R8)}$$

$$R + O_2 + M \rightarrow RO_2 + M \quad\quad\quad\quad\quad\quad\quad\quad\text{(R9)}$$

$$RO_2 + NO \rightarrow NO_2 + RO \quad\quad\quad\quad\quad\quad\quad\quad\text{(R10)}$$

$$RO + O_2 \rightarrow HO_2 + R'CHO \quad\quad\quad\quad\quad\quad\quad\quad\text{(R11)}$$

$$HO_2 + NO \rightarrow OH + NO_2 \quad\quad\quad\quad\quad\quad\quad\quad\text{(R12)}$$

$$2 \text{ x } (NO_2 + h\nu \rightarrow NO + O) \quad\quad\quad\quad\quad\quad\quad\quad\text{(R13)}$$

$$2 \text{ x } (O + O_2 + M \rightarrow O_3 + M) \quad\quad\quad\quad\quad\quad\quad\quad\text{(R14)}$$

$$RH + 4O_2 \rightarrow R'CHO + H_2O + 2O_3 \quad\quad\quad\quad\quad\quad\quad\quad\text{(Net)}$$
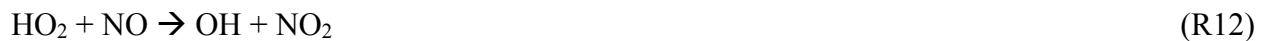
Here, RH represents the VOC species; R is the product of VOC oxidation, such as the CH_3 radical. Ozone production is limited in this case by both Reaction R12 and the sum of R10 over all the VOCs that participate in these reactions; NO_2 remains a critical precursor species within this mechanism. Both mechanisms are terminated by the reaction of the HO_2 radical with itself or by reaction of OH with NO_2 (Finlayson-Pitts and Pitts, 2000; Jacob, 1999):

$$HO_2 + HO_2 \rightarrow H_2O_2 + O_2 \quad\quad\quad\quad\quad\quad\quad\quad\text{(R15)}$$

$$OH + NO_2 + M \rightarrow HNO_3 + M \quad\quad\quad\quad\quad\quad\quad\quad\text{(R16)}$$

Unlike other pollutants, O_3 is often a regional rather than a local air quality problem. This is due to the long lifetime of O_3 (~10 hours within the boundary layer), which allows O_3 to be

vented out of the planetary boundary layer (PBL) and transported within the free troposphere (FT), where its lifetime is on the order of weeks. The current National Ambient Air Quality Standard (NAAQS) for O_3, set by the EPA under the authority of the Clean Air Act, is a maximum 8-hour average mixing ratio of 70 ppbv (as of October, 2015), attained if the 3-year average of the fourth-highest daily maximum 8-hour average O_3 mixing ratio at each monitor within a region does not exceed 70 ppbv. The current short-term standard for NO_2 is a one-hour average of 100 ppbv, and NO_2 has a lifetime on the order of 1-2 hours within the PBL (although NO_2 may be temporarily sequestered in a reservoir species such as nitric acid). Southern and central California, the mid-Atlantic region, eastern Texas, Chicago and industrial centers in the Ohio River valley, and urban centers in the Southeast violated the previous 75 ppbv O_3 NAAQS. These are also the regions of the US where most of the population resides (Jacob, 1999). However, O_3 air quality in the eastern US has improved since the EPA issued its NO_x State Implementation Plan (SIP) call in 1998, requiring 21 states to reduce their summertime NO_x emissions. By scrubbing NO_x emissions from power plants, transport of this precursor gas and resulting O_3 have decreased, and thus ozone air quality has improved in this region (Bloomer et al., 2009; Gégo et al., 2007; He et al., 2013). Reduction of mobile source emissions under Tier II requirements for vehicles has also led to NO_x and O_3 improvements.

1.2: Application of Satellite Observations to Tropospheric Air Quality

Satellite observations have been successfully applied to the study of air quality within the troposphere for nearly three decades, including quantifying the atmospheric abundances and distributions of many trace gas species, assessing temporal trends in these species, and top-down estimates of trace gas emissions (Fishman et al., 2008). For example, satellite data were used to determine that anthropogenic sources dominated over biogenic sources for tropospheric CO

during the 1980s and 1990s (Reichle et al., 1986; Reichle et al., 1999). Early research was able to determine the tropospheric O_3 column from the "residual" information after removal of the stratospheric component from the total observed column (Fishman et al., 1990). Retrievals able to take advantage of nadir measurements were later developed for O_3, NO_2, and other trace gases more suited to the troposphere, allowing direct observation of trace gas tropospheric columns (Bhartia et al., 1996; Liu et al., 2006; Richter and Burrows, 2002; Martin et al., 2002). Use of satellite data has demonstrated that most of the tropospheric NO_2 column resides in the lower troposphere near local emissions sources (Martin et al., 2006). Such retrievals led to the creation of global data sets of trace gas tropospheric column abundances, which have been applied to the study of many problems of atmospheric pollution. The global coverage, coupled with increasingly high spatial resolution, and fixed temporal resolution of such observations provide key advantages over other data sets. (Beirle et al., 2003; Boersma et al, 2008; Chatfield and Esswein, 2012; Fishman et al., 2008; Martin et al., 2008; Lamsal et al., 2011). The successful application of satellite observations to such challenges has thus led to the desire to relate column abundances directly to surface concentrations to address surface air pollution (Fishman et al., 2008). This may be especially applicable for regions that lack sufficient surface air quality monitors.

However, several factors currently complicate the applicability of the satellite-observed column abundances for surface air quality assessments. These include the biases inherent in satellite retrievals, the method for separation of the stratospheric and tropospheric burdens, and reduced sensitivity of satellite instruments to the lower troposphere, where the greatest concentrations of many pollutants are found. Furthermore, many current air quality satellite instruments that make observations in the UV and visible wavelengths are onboard low-Earth

orbit (LEO) satellites, limiting temporal coverage to one overpass per sunlit portion of the day at most sites. These observations thus miss the diurnal development of meteorology, emissions, and chemistry that are relevant to both the column abundance and surface mixing ratio. Because of these factors, uncertainties remain in the relationship between column abundances observed by satellites and surface mixing ratios, which are directly relevant to air quality management (Martin et al, 2008; Lee et al., 2011; Natraj et al., 2011). The upcoming National Aeronautics and Space Administration (NASA) satellite instrument, the Tropospheric Emissions: Monitoring of Pollution (TEMPO, Chance et al., 2013) which will be part of the Geostationary Coastal and Air Pollution Event (GEO-CAPE, Fishman et al., 2012) mission, will address some of these concerns. TEMPO will be on a geostationary satellite parked over the Equator for viewing North America, providing high spatial and temporal resolution observations of several key pollutants, and improved vertical resolution of O_3 profile retrievals. However, because a number of retrieval assumptions will still be necessary, the challenge of relating these satellite observed quantities to surface mixing ratios will remain.

1.2.1: Recent Works Addressing the Column-Surface Relationship

To address this challenge, recent work has focused on the use of models to relate columns and surface mixing ratios. Lamsal et al. (2008) developed a method to infer ground-level NO_2 mixing ratios from the Ozone Monitoring Instrument (OMI) tropospheric column abundance of NO_2 with the use of local scaling factors derived from the GEOS-Chem model. These scaling factors were defined as the ratio of the model-predicted surface mixing ratio to the model-predicted column at the model grid point nearest each OMI data point. After application of these scaling factors, Lamsal et al. (2008) obtained significant correlation between OMI-derived surface mixing ratio and *in situ* NO_2 measurements over much of the U.S. and Canada for the

year 2005. Ordóñez et al. (2006) employed seasonal NO_2 vertical profiles from the global MOZART-2 model to scale NO_2 *in situ* mixing ratios to column abundances over the Lombardy region of Italy for the years 1996-2002, and obtained good agreement between the column data derived from surface measurements and the coincident Global Ozone Monitoring Experiment (GOME) satellite column densities. Knepp et al. (2013) focused on tropospheric NO_2 column data from Pandora (Herman et al., 2009), a ground-based, sun-tracking spectrometer providing remotely sensed column densities, from Hampton, VA, from July 2010 to October 2011, and from Edgewood and Padonia, MD, during the DISCOVER-AQ deployment during July 2011. In this work, planetary boundary layer (PBL) heights derived from the EDAS-40 model and an assumption of a well-mixed lower tropospheric NO_2 profile were used to convert the Pandora data into mixing ratio values; significant correlation was obtained between the converted Pandora columns and *in situ* measurements.

Lee et al. (2011), building upon the work of Lamsal et al. (2008), developed a scaling factor defined as the spatial average of the ratio of *in situ* NO_2 mixing ratio to the coincident OMI column over the area surrounding Windsor, Ontario, to transform OMI tropospheric column data into surface mixing ratios. Lee et al. (2011) obtained significant correlation between OMI-derived surface mixing ratio and *in situ* observations over seven 2-week sampling periods during 2007. Boersma et al. (2009) transformed surface *in situ* NO_2 mixing ratio data for several Israeli cities for the year 2006 into PBL column abundances with the assumption that NO_2 is well mixed within the PBL. Significant correlation was found between the derived PBL columns and OMI or SCIAMACHY column NO_2 data. Lastly, Chatfield and Esswein (2012) analyzed ozonesonde data over the U.S., and their results demonstrated that full tropospheric O_3 columns are rarely useful for surface air quality applications. However, they also demonstrated that lower

tropospheric partial O_3 columns (between approximately 0-3 km altitude) exhibited considerable correlation with near-surface O_3 (the average O_3 mixing ratio in the layer extending from the surface to approximately 500 m altitude).

Several of the papers described here relied upon models to relate surface and column data, and use of models will likely continue in future work. Vertical mixing within air quality models is one of several key parameters that have large impacts on the simulated air quality, thus impacting the model relationship between surface and column quantities. Lin et al. (2008) examined the sensitivity of the summertime U.S. O_3 diurnal cycle to PBL mixing, spatial resolution, and emissions of precursors within the Model for Ozone and Related Tracers, version 2, (MOZART-2), and found that vertical mixing exerted the most control over the diurnal cycle. Further, Lin et al. found that nonlocal mixing most realistically captured the observations of the ozone diurnal cycle over each U.S. region studied. This last finding is consistent with several other papers comparing PBL schemes within the WRF meteorological model, finding that non-local schemes better compared to observations of PBL height than local schemes (Hu et al., 2010; Shin and Hong 2011; Xie et al., 2012). Castellanos et al. (2011) examined CO column content and profiles within the regional Community Multiscale Air Quality (CMAQ) model over the northeastern U.S. to assess the vertical mixing. Simulated lower tropospheric CO columns agreed well with measured columns, while the CO profile from the model often failed to capture fine structure apparent in measured profiles. This suggested that model PBL vertical mixing may be too fast, while venting into the lower free tropospheric may be too slow.

1.3: Overview of the NASA DISCOVER-AQ Mission

The ultimate goal of the NASA DISCOVER-AQ (Deriving Information on Surface conditions from Column and Vertically Resolved Observations Relevant to Air Quality) project is

to provide information relevant to improving our ability to relate satellite-observed column densities to surface conditions for aerosols, O_3, NO_2, and formaldehyde. Additional goals include characterizations of differences in diurnal variability for surface and column observations and the horizontal scales of variability affecting satellites and model calculations. DISCOVER-AQ combines P-3B aircraft *in situ* profiling of trace gas species, aerosol properties, and key meteorological variables, UC-12 aircraft remote sensing of aerosols and trace gas columns, observations of surface conditions from the existing network of surface air quality monitors, remote sensing of trace gas columns and aerosols from a network of ground-based Pandora UV/vis spectrometers and a network of AERONET sun photometers collocated with the air quality monitors (and additional monitors at some sites), and model simulations for each of the four campaigns.

The first campaign was conducted in the Baltimore-Washington metropolitan region of Maryland during July 2011, which encountered deep, convective boundary layers, synoptic-scale stagnation under the influence of the Bermuda High, less wind shear than the other three campaigns, warm temperatures, the influence of the Chesapeake Bay breeze at the Edgewood and Essex spiral sties, and transport of polluted air from the Ohio River Valley. The second deployment was conducted in the San Joaquin Valley of California during January-February 2013, which encountered cold temperatures, air recirculation and valley drainage winds within the valley due to the influence of the nearby mountains, shallow boundary layers, and high concentrations of NO_x and aerosol species. The third campaign took place in the Houston, TX, metropolitan area during September 2013, which encountered deep, convective boundary layers, frequent cold frontal passages, wind shear, the Gulf breeze at the Galveston spiral site and the Galveston Bay breeze at the Smith Point spiral site, and a complex chemical environment due to

NO$_x$ emissions from automobiles and ships as well as VOC emissions from nearby petrochemical refineries. The fourth and final deployment was conducted in the Front Range region of Colorado during July-August 2014, which encountered upslope and downslope flow on the Front Range, the development of Denver Cyclone over the Denver-Boulder region (a solenoidal circulation), convective boundary layers, frequent afternoon pop-up thunderstorms, and a complex chemical environment with emissions of NO$_x$ as well as VOCs from fracking wells near some sites. Thus, the project covered a large range of meteorological and pollution conditions throughout these four campaigns, and allowed the collection of multiple, high quality column abundance, *in situ* profile, and surface mixing ratio data sets.

1.4: Thesis Objectives and Outline

The papers discussed above demonstrate the potential to relate surface mixing ratio and column abundance data for O$_3$ and NO$_2$. However, most of these studies have focused on the correlation between column and surface data that have been averaged over time and space. The focus has also been placed primarily on methods to transform one data type into the other, and not on processes controlling the column-surface relationship.

The work presented in this dissertation has been conducted in support of the goals of the DISCOVER-AQ project, and seeks to understand the degree of correlation between column and surface data, as well as the processes that influence the column-surface relationship, during the DISCOVER-AQ deployments. Additionally, this work will focus on the EPA criteria pollutants O$_3$ and NO$_2$, because of the importance of NO$_2$ as an O$_3$ precursor and the adverse impacts of O$_3$ on human health, crop yields, and the atmospheric radiation budget, as well as the established history of successful measurements by satellites. We aim to answer several important questions, including

1. Can any conclusions be drawn about the types of conditions under which surface air quality would be best (or most poorly) estimated from column observations from a geostationary satellite?

2. What is the diurnal variation of column abundances seen from integration of in-situ profile data, from ground-based spectrometers, from airborne remote sensing instruments, and from model simulations, and how do these column diurnal cycles compare to those of surface mixing ratio data?

3. Can a regional air quality model, such as CMAQ or WRF-Chem, potentially provide the shape factor profiles used in remote sensing retrievals, rather than a global chemical transport model?

4. Can the WRF-Chem model system be improved to bring the column-surface relationship within the model closer to that encountered in the observations?

This dissertation is divided into six chapters. The first (and present) chapter presents an introduction to tropospheric O_3 chemistry, the use of satellite data for air quality, and the current ability to relate satellite-observed columns to surface air quality conditions. Chapter 2 presents results of linear regression analyses of the P-3B lower tropospheric and Pandora full tropospheric O_3 or NO_2 column data versus surface mixing ratio data during the Maryland deployment, the first deployment of the DISCOVER-AQ mission. This work was the catalyst for all subsequent chapters, and represents the initial investigation into the degree of correlation between column and surface data and processes influencing these correlations. These analyses suggest that O3 partial column observations from future satellite instruments with sufficient sensitivity to the lower troposphere can be meaningful for surface air quality analysis, while planetary boundary layer height data add meaningful information to the NO_2 regressions. Furthermore, the degree of

correlation between the corresponding CMAQ lower tropospheric O_3 or NO_2 columns and surface mixing ratios was overestimated within the model, suggesting that vertical mixing is too strong within the WRF/CMAQ model system. The results of this study led to a paper that was published in *Atmospheric Environment* (Flynn et al., 2014).

Chapter 3 of this dissertation follows directly from the work of Chapter 2, and presents the results of an agglomerative hierarchical clustering analysis performed on the P-3B *in situ* lower tropospheric O_3 or NO_2 profiles for each of the four DISCOVER-AQ deployments; the results of correlation analyses between the column and surface data associated with each of the profile clusters obtained for all four deployments are also presented. This work characterizes the classes of profile shapes for each trace gas and each campaign, and explores the meteorological conditions that influenced the profile shape clusters and thus the column-surface correlations. These results suggest that satellites may be most relevant for surface air quality under the conditions associated with the Maryland deployment, which included deep, convective boundary layers and few interferences to the column-surface connection from complex meteorology, chemical environments, or orography. Further, vertical mixing and atmospheric stability exerted an important influence on the O_3 profile cluster shapes and correlations for each campaign, while O_3 photochemistry exerted the primary control on NO_2 profile variability. The CMAQ model captured the shape factors for O_3, and moderately well captured the NO_2 shape factors, for the conditions associated with the Maryland campaign, suggesting that a regional air quality may adequately specify *a priori* profile shapes for remote sensing retrievals. CMAQ shape factor profiles were not as representative of atmospheric observations for the other regions. Coarser vertical resolution in the NASA Global Modeling Initiative (GMI) global chemical transport model (CTM) model affected the ability of that model to reproduce the observations. These

results demonstrate the importance of resolution for accurate representation of pollutant profiles as *a priori* information within satellite retrievals, and for the ability to relate column abundances to surface concentrations. The results of this study led to a paper now under review in *Atmospheric Environment* (Flynn et al., 2016).

Chapter 4 builds upon the work of Chapter 2 and Chapter 3 through investigation of the campaign-average diurnal variation of O_3 and NO_2 column amounts within the observational lower tropospheric and full tropospheric datasets for each spiral site for each campaign. The average column diurnal cycles are compared to the campaign-average diurnal cycles of surface O_3 and NO_2 mixing ratios to determine if the column and surface cycles exhibit similar behavior, as well as to provide an indication of when satellite column observations may be most representative of surface concentrations for these two trace gases. These results indicate that neither full tropospheric nor lower tropospheric O_3 column abundances exhibited a clear diurnal cycle for any spiral site or campaign, indicating that O_3 column variability is largely independent of local synoptic meteorology and pollution conditions and is not connected to surface variability. Boundary layer dynamics play an important role in the regulation of the variability of these columns. However, the NO_2 full tropospheric and lower tropospheric column abundances, however, did display diurnal variability at most spiral sites during all campaigns, though the column diurnal variability was smaller in amplitude and offset in time relative to the surface diurnal variation. Column NO_2 variability is controlled by surface production of NO_2 and boundary layer mixing of NO_2 into the lower troposphere. Neither set of results for O_3 and NO_2 suggest a time of day when satellite column observations may be most representative of surface concentrations. Lastly, the observed column variability was compared to the column variability simulated by CMAQ and GMI for both gases. Both models replicated the shapes of the observed

column diurnal cycles for both gases and all campaigns. CMAQ was most able to capture the observed O_3 columns for the conditions associated with the Maryland and Texas campaigns, which included deep, convective boundary layers and adequate temperatures and sunlight for O_3 formation, while CMAQ best captured the observed NO_2 column magnitudes for all campaigns except California, which experienced greater NO_2 pollution and reduced O_3 photochemical production and NO_2 photolytic loss relative to the other three campaigns.

The ability of the coupled meteorology-chemistry WRF-Chem to effectively simulate the interplay between boundary layer mixing and O_3 and NO_2 vertical profiles, and the associated impacts on the column-surface correlations for these trace gases, is investigated in Chapter 5, for the surface O_3 pollution episode that occurred on July 26-29, 2011. Further, the relevance of the WRF-Chem model profiles for use in remote sensing retrievals is evaluated. Five PBL schemes are tested, including two nonlocal schemes (the ACM2 and YSU schemes) and three local schemes (the BouLac, MYJ, and QNSE schemes). Overall, no one PBL scheme was able to accurately simulate all observed meteorological or chemical species, as expected. However, the ACM2 scheme best captured the observed PBL heights and observed NO_2 column-surface correlation at each spiral site. All schemes compared well to the observed hourly O_3 median shape factors, and the BouLac scheme most accurately simulated the observed O_3 column-surface correlations. Additionally, WRF-Chem was able to replicate most of observed local minima and maxima in the O_3 shape factors at the correct altitudes, presenting a distinct advantage of a regional online meteorology-chemistry model over offline air quality models such as CMAQ. These results suggest that a regional, coupled meteorology-chemistry model may reasonably specify *a priori* profile shapes for remote sensing retrievals, particularly as PBL schemes continue to be improved. The results of this study led to a paper now in preparation, to be submitted this

summer to *Atmospheric Chemistry and Physics* (Flynn et al., in prep.). Finally, the major conclusions of this dissertation and directions for future research are presented in Chapter 6.

Chapter 2: Relationship Between Column-Density and Surface Mixing Ratio: Statistical Analysis of O_3 and NO_2 Data from the July 2011 Maryland DISCOVER-AQ Mission

2.1: Introduction

Satellite observations have made important contributions to the understanding of atmospheric chemistry and pollution over the past three decades, including quantifying the atmospheric abundances and distributions of many trace gas species, assessing temporal trends in these species, and top-down estimates of trace gas emissions (Fishman et al., 2008). Global coverage, coupled with increasingly high spatial resolution, and fixed temporal resolution of such observations provide key advantages over other data sets. Retrievals of tropospheric column abundances have also improved (Beirle et al., 2003; Boersma et al, 2008; Bucsela et al., 2013; Chatfield and Esswein, 2012; Fishman et al., 2008; Martin, 2008; Lamsal et al., 2011). Trace gas observations from satellites thus have great potential for diagnosis of near-surface conditions. This can be especially useful for monitoring the Environmental Protection Agency (EPA) criteria pollutants ozone (O_3) and nitrogen dioxide (NO_2), (http://www.epa.gov/air/criteria.html), pollutants known to have significant adverse impacts on human health, crop yields, and the atmospheric radiation budget. The greatest benefit may come in regions that lack sufficient surface air quality monitors**.**

However, several factors currently complicate the applicability of satellite-observed column abundances for surface air quality assessment. These include biases in satellite retrievals and reduced sensitivity of satellite instruments to the lower troposphere (Martin, 2008; Lee at al., 2011). Just as importantly, uncertainties remain in the relationship between column abundances observed by satellites and surface mixing ratios, which are directly relevant to air quality management. Recent work demonstrates progress in understanding this relationship. Chatfield and Esswein (2012) analyzed ozonesonde data over the U. S. and found substantial correlation

between partial-column O_3 (0-3km) and near-surface O_3 (in the lowest 500m). Consistent with Chatfield and Esswein (2012), Martins et al. (2015) found significant, positive correlations between upper troposphere (7-10 km) O_3 partial column and near-surface (1-3 km) O_3 partial columns from ozonesonde data measured at the Beltsville and Edgewood sites during the Maryland campaign. Martins et al. built a linear regression model to predict near-surface O_3 from upper troposphere abundances, and found agreement to within 11% between the ozonesonde observations and near-surface O_3 values from the regression model. This is encouraging that satellite instruments with sensitivity to the mid-upper troposphere may be related to O_3 abundances near the surface; however, difficulties remain in the ability to relate upper air O_3 partial columns derived from the ozonesondes to surface O_3 (0-100 m; Martins et al., 2015). Lamsal et al. (2008) developed a method to infer ground-level NO_2 mixing ratios from the Ozone Monitoring Instrument (OMI) tropospheric column abundances with the use of local scaling factors derived from the GEOS-Chem model. Significant correlation between OMI-derived and *in situ* surface NO_2 was observed (Lamsal et al., 2008, 2010). Other works have demonstrated significant correlation between satellite-observed NO_2 columns and surface NO_2 data scaled to obtain column amounts with the use of assumed NO_2 profiles (Ordóñez et al., 2006; Boersma et al., 2009). Knepp et al. (2013) used model-derived planetary boundary layer (PBL) heights to convert Pandora NO_2 tropospheric columns into average surface mixing ratios, also demonstrating high correlation between converted columns and surface data. Understanding the uncertainties in the relationship between column density and surface mixing ratio becomes more urgent with the up-coming NASA Tropospheric Emissions: Monitoring of Pollution satellite mission (TEMPO, Chance et al., 2013) which is likely to be one component of the Geostationary Coastal and Air Pollution Event (GEO-CAPE, Fishman et al., 2012) mission. TEMPO/GEO-

CAPE will be centered over ~100˚ W, allowing observations over North America from geostationary orbit with product horizontal resolution of 8 km × 4.5 km at the center of domain, much higher than current Low-Earth-Orbit (LEO) measurements. GEO-CAPE may combine multiple spectral regions to improve the vertical resolution of ozone profile retrievals, especially in the lowermost troposphere (Natraj et al., 2011). However, because a number of retrieval assumptions will still be necessary, the challenge of relating the satellite-observed quantities to surface mixing ratios will remain.

The objectives of and instrumentation platforms that comprised the DISCOVER-AQ project has been described previously in Chapter 1. The July 2011 campaign was conducted in the Baltimore-Washington metropolitan region and involved 6 surface air quality monitoring sites. These included Aldino, Beltsville, Edgewood, Essex, Fair Hill, and Padonia, MD, with locations mapped in Fig 2.1. The P-3B accomplished over 250 profiles on 14 flight days over six surface air quality monitoring sites and the Chesapeake Bay during the Maryland deployment. These flight days covered a range of conditions, including especially clean days on July 14th and 16th and pollution episodes during July 1-5 and July 18-23, as well as flights on weekdays and weekends. A complete description of DISCOVER-AQ measurements is publicly available at http://www-air.larc.nasa.gov/missions/discover-aq/discover-aq.html

In support of DISCOVER-AQ, results are presented of linear regression analyses between O_3 and NO_2 surface mixing ratio and column measurements, including column abundances integrated over *in situ* profile data from the P-3B aircraft, measured by the Pandora UV/Vis spectrometer, and observed by the Aura/OMI instrument. Through these analyses, the strength of the column-surface relationship and the ability to predict simultaneous surface mixing ratio from column abundance during the July 2011 deployment will be assessed. The column-

surface relationship in the CMAQ model is also evaluated and compared with the results obtained from the observations.
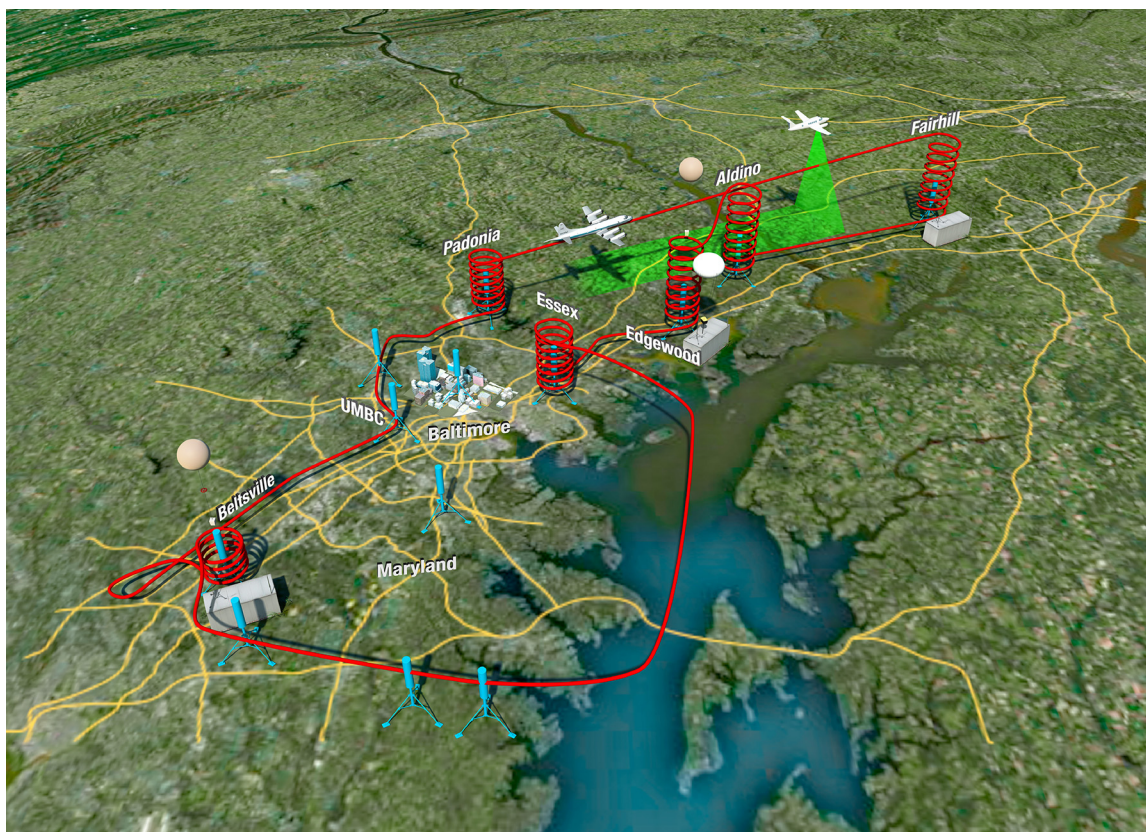
Fig. 2.1: Example P-3B flight track for the July 26th flight, displaying the locations of the 6 surface air quality monitoring sites.

2.2: Description of Observational Column Datasets

2.2.1: Aircraft Lower Tropospheric Columns

2.2.1.1: P-3B Column Air and Column Ground Lower Tropospheric Columns

Two different column amounts were available for P-3B *in situ* spirals, which differed in the method used to fill the gap between the lowest P-3B measurement altitude and the ground: column_air and column_ground. To compute column_air, the O_3 or NO_2 mixing ratio value at the lowest aircraft measurement level was held constant to the surface, while column_ground held the surface mixing ratio value constant up to the lowest aircraft measurement, when a

surface value was available. The profiles were then averaged into 100 m bins from the surface to the top of the P-3B spiral (typically ~3.2 km AMSL), giving an average mixing ratio for each vertical layer; average air density for each layer was also computed. These mixing ratios were then converted from mixing ratio to concentrations (molecules/cm^3), allowing computation of the partial column amounts for each 100 m bin. Finally, these partial columns were summed over the depth of the P-3B spiral to obtain the P-3B lower tropospheric *in situ* column_air and column_ground. The column amounts were then computed by integration of these lower tropospheric *in situ* profiles. The well mixed assumption inherent in the column_ground computation should be noted. These columns represent the aircraft lower tropospheric column amounts for each campaign. Furthermore, column_air and column_ground represent the lower and upper limits, respectively, on the true lower tropospheric column amount, as column_air does not assume the surface concentration is mixed into the PBL while column_ground assumes strong vertical mixing that efficiently communicated the surface value to the lower tropospheric column. Uncertainty in the column amounts is 5% for the O_3 column_air and column_ground, and 20% for NO_2 column_air and column_ground (Gao Chen and Andrew Weinheimer, personal communication).

2.2.1.2: UMD Cessna Column UMD Lower Tropospheric Columns

The Aldino P-3B profile shapes were compared to in situ profile shapes measured by the UMD Cessna aircraft for O3 and NO2 during the campaign. The Cessna always reached lower altitudes than the P-3B, so this comparison was used to identify which P-3B column better approximated the true column at Aldino. The Cessna profiles for O_3 (Fig. 2.2) typically remained well mixed to the lowest P-3B altitude (~300 m above ground level, or AGL). Differences in O_3 measured by the Cessna and P-3B may be due to the different interferences experienced by the

respective instruments after a spike in relative humidity (Arkinson et al., in preparation). The Cessna NO_2 profiles most often displayed a "boot shaped" appearance: very low mixing ratio magnitudes above the PBL, with magnitude increasing with decreasing altitude in the upper PBL, finally becoming well mixed within the lowest portion of the profile in the lower PBL; an increase in NO_2 mixing ratio was often encountered in the Cessna profiles near the lowest P-3B altitude (Fig. 2.2). This suggests that generally column_air was closer to the true O_3 and NO_2 columns; however, due to the "boot" in the NO_2 profile, column_air likely underestimates the true NO_2 partial column. However, the Cessna profiles also demonstrated that NO_2 mixing ratio could increase dramatically in magnitude below the lowest P-3B measurement altitude towards the surface mixing ratio value (Fig. 2.3); this is consistent with the 95th percentile NO_2 mixing ratio profile reported in Brent et al. (2015; see their Fig. 5). Additionally, these Cessna profiles were used to construct estimated profiles for the portion of the atmosphere below the lowest altitude of the Aldino P-3B spirals, and additional Aldino column amounts (column_UMD; uncertainty of 5%) were computed from these estimated profiles.
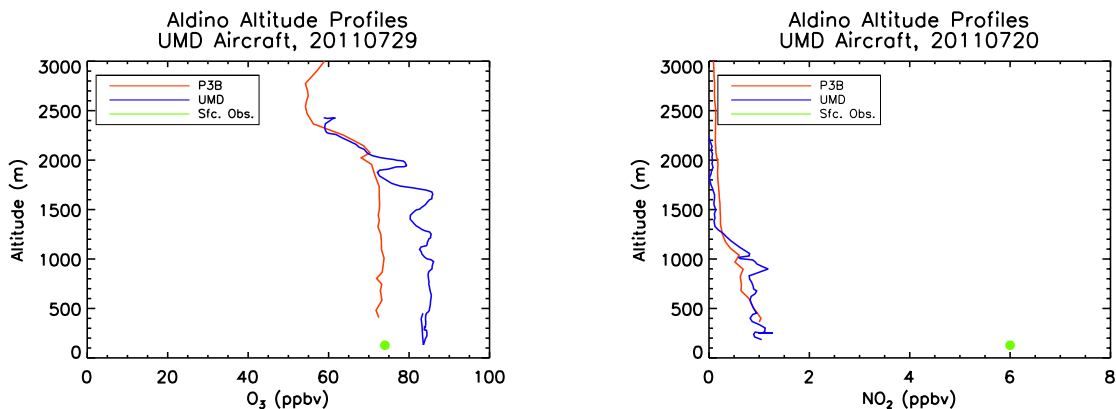
Figure 2.2: Example UMD Cessna altitude profiles for Aldino. O_3 profile plotted in the left profile as solid blue line; NO_2 profile plotted in the right profile. NO_2 profile displays the "boot shaped" appearance. Corresponding P-3B profiles also plotted for comparison (orange). Green circles represent surface O_3 and NO_y mixing ratio data, measured at the nearby Aldino ground monitoring site, averaged over the time of UMD profile and plotted at the elevation AMSL of the monitoring site.
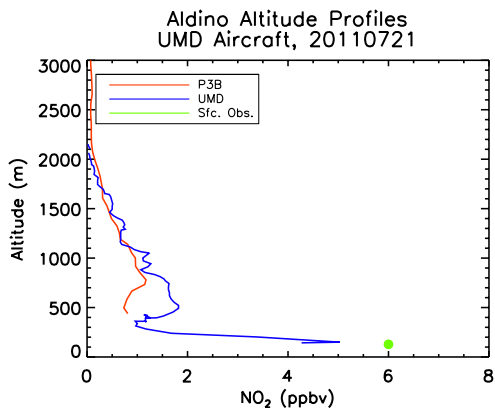
Figure 2.3: Example P-3B and UMD Cessna NO_2 altitude profiles for Aldino, displaying the increase in NO_2 mixing ratio below the lowest P-3B measurement altitude.

2.2.2: Pandora Full Tropospheric Columns

The ground-based Pandora UV/Vis spectrometers (Herman et al., 2009) were located at each spiral site for the Maryland campaign. The Pandora instruments observed O_3 and NO_2 total column amounts during daylight hours for all days during the campaign. Tropospheric columns were estimated by subtracting the stratospheric component derived by the OMI algorithms from

the Pandora total observed columns. Reed et al. (2015) found good agreement between OMI and Pandora for total column O_3 (residuals within ± 4.5%) and for total column NO_2 (residuals within ± 25%) during this campaign. Errors in Pandora tropospheric column amount were approximately 5% for O_3 and 7% for NO_2, and are due primarily to uncertainties in the OMI stratospheric column amount (approximately 2% for O_3 and 5% for NO_2), with some contribution from uncertainties in the total Pandora column. It should also be noted that the Pandora instrument continued to observe during cloudy conditions (though this impacted the retrieval quality), while P-3B flight days were chosen to minimize cloud cover and the aircraft actively avoided clouds during flight. It is also possible that the Pandora instrument observed different air masses from the P-3B spiral, if it did not point in the same direction as the aircraft spiral flight pattern.

2.2.3: OMI Full Tropospheric Columns

Tropospheric columns from the Ozone Monitoring Instrument (OMI) onboard the Aura satellite were retrieved with the Version 2.1 Goddard tropospheric NO_2 retrieval algorithm (Buscela et al., 2013; uncertainty of approximately 30%) and the ozone profile algorithm by Liu et al. (2011) with modifications as described in Kim et al. (2013) (uncertainty of approximately 5%). These data were screened for cloud fraction (effective cloud fraction less than 30%), the instrument row anomaly, and distance from the surface site (pixel center less than 100 km distance).

2.3: Description of Surface Volume Mixing Ratio Datasets

Surface volume mixing ratio data were provided by the Maryland Department of the Environment (MDE) for O_3 at all spiral sites from a molybdenum-converter chemiluminescence monitor, for NO_2 at Essex from a chemiluminescence monitor, and for NO_y from such monitors

at the Aldino and Beltsville sites. The EPA provided NO_2 measurements from chemiluminescence instruments with photolytic converters at the Edgewood and Padonia spiral sites. The NASA Chemical, Optical, and Microphysical Measurements of In-situ Troposphere (COMMIT; http://smartlabs.gsfc.nasa.gov) trailer provided O_3 and photolytic converter NO_2 measurements for Fair Hill. Uncertainties for the surface datasets were 5% for O_3, 10% for NO_2 (data archive) for the surface NO_2 data provided by MDE, and 10% for the surface NO_2 data provided by EPA.

2.4: Description of P-3B PBLH Dataset

Donald Lenschow (NCAR, retired) provided boundary layer height analyses based on the P-3B potential temperature profiles during the Maryland campaign. The PBL top was located where the potential temperature lapse rate exceeded approximately 3 K/km, with a relatively constant potential temperature lapse rate from the surface to the PBL top. The potential temperature profiles were also analyzed manually to ensure the algorithm accurately diagnosed the PBLH. Water vapor and ozone profiles were examined in addition to the potential temperature profiles.

2.5: Description of CMAQ Simulations and Column Amounts

2.5.1: Loughner et al. (2014) WRF/CMAQ Simulation

A simulation of the Maryland campaign was provided by Christopher P. Loughner of NASA GSFC (Loughner et al., 2014). The CMAQ model version 5.0 was used to simulate air quality for this deployment, and was driven offline by output from the Weather Research and Forecasting (WRF) meteorological model, specifically the Advanced Research WRF core (WRF-ARW; Skamarock et al., 2008). The North American Regional Reanalysis (NARR) was used for the initial and boundary conditions within WRF, and the WRF/CMAQ model system was run

with 34 vertical layers from the surface to 100 mb, with 16 layers within the lowest 2 km above ground level (AGL) to capture boundary layer processes. The WRF simulation also employed the Asymmetric Convective Model 2 (ACM2; J. E. Pleim, 2007a) scheme for vertical diffusion and convective mixing, the Pleim-Xiu surface layer scheme (J. Pleim, 2006), and the Pleim-Xiu land surface model (Xiu and Pleim, 2001).

Chemical initial and boundary conditions were provided by a simulation of the Model for Ozone and Related Chemical Tracers, version 4 (MOZART-4; Emmons et al., 2010). The CMAQ model used the Carbon Bond-05 (CB05; Yarwood et al., 2005) gas-phase chemical mechanism, the fifth generation aerosol module (aero5), and the ACM2 for vertical diffusion and convective mixing. The projected 2012 anthropogenic emissions based on the 2005 National Emissions Inventory (NEI), because 2011 emissions were not yet available at that time for the Maryland simulation. Anthropogenic mobile emissions were computed with the Motor Vehicle Emissions Simulator (MOVES, specifically MOVES2010; Kota et al., 2012), while biogenic emissions were computed with the Biogenic Emissions Inventory System within CMAQ, version 3.6 (BEIS; Pouliot and Pierce, 2009). CMAQ output was provided at 12 km horizontal resolution and in hourly averages. Lightning NO_x emissions were also calculated in-line within the model.

2.5.2: NOAA ARL CMAQ Forecasts

Forecasts of O_3 and NO_2 were provided by the National Oceanic and Atmospheric Administration (NOAA) Air Resources Laboratory (ARL) during the Maryland deployment, and this output was also analyzed. Like the Loughner et al. simulations, these forecasts used the CB05 mechanism. However, these forecasts employed an experimental version of CMAQ version 4.6, and were driven offline by the WRF-Nonhydrostatic Mesoscale Model (NMM) core, used the fourth generation aerosol module (aero4), the Mellor-Yamada-Janjic (MYJ; Janjić, 1994) scheme

for boundary layer mixing, the Noah land surface model, and the 2005 NEI for anthropogenic emissions. The available horizontal resolution was 12 km, and the available vertical resolution was 22 layers with 13 layers within the lowest 2 km.

2.5.3: CMAQ Partial Tropospheric Column Amounts

To be able to compare the CMAQ column-surface correlations to those from the P-3B, partial tropospheric columns were computed from the Loughner et al. or NOAA CMAQ output for O_3 and NO_2 over the depths of the P-3B spirals. CMAQ profiles coincident to each P-3B profile were sampled, and the CMAQ levels below or above the lowest or highest P-3B measurement altitudes were excluded from the column computation. The O_3 or NO_2 simulated partial column amounts were then computed from integration of the simulated lower tropospheric profile.

2.6: Linear Least Squares Regression Analyses

2.6.1: Simple Linear Least Squares Regression Analysis for the P-3B, Pandora, and OMI

A simple linear least squares regression analysis was performed between the P-3B column_air, P-3B column_ground, Pandora, and OMI O_3 and NO_2 columns and surface mixing ratio data for each surface-monitoring site. An additional analysis between the Aldino column_UMD and the surface data was also conducted for comparison to column_air and column_ground at this site. Surface data were averaged over the time of the aircraft spiral for use with the P-3B analyses. Hourly averages of the surface data for the hours between 7 AM-7 PM EDT were computed for use with the Pandora columns, while 15 minute averages centered on 2:45 PM EDT were computed for use with the OMI columns. Column abundance was used to predict the simultaneous surface volume mixing ratio (VMR), yielding a regression model of the form

surface VMR = β*(column) + intercept   Eqn (2.1)

where β is the regression coefficient. The NO_2 column and surface data followed an approximately lognormal distribution, and were therefore log-transformed (i.e., the natural logarithms of the NO_2 surface or column data were used instead of surface or column directly, to account for the non-Gaussian distribution of these data) before performing statistical analyses. Histograms of the Pandora column NO_2 observations and NO_2 surface mixing ratios at Essex are displayed in Fig. 2.4 to demonstrate how log-transformation produced data distributions that followed a Gaussian distribution more closely than the untransformed data. The Pandora O_3 column data were also approximately lognormal, and were also log-transformed. The degree of association between the column and surface data and the errors of the regression model relative to the observed data were assessed.
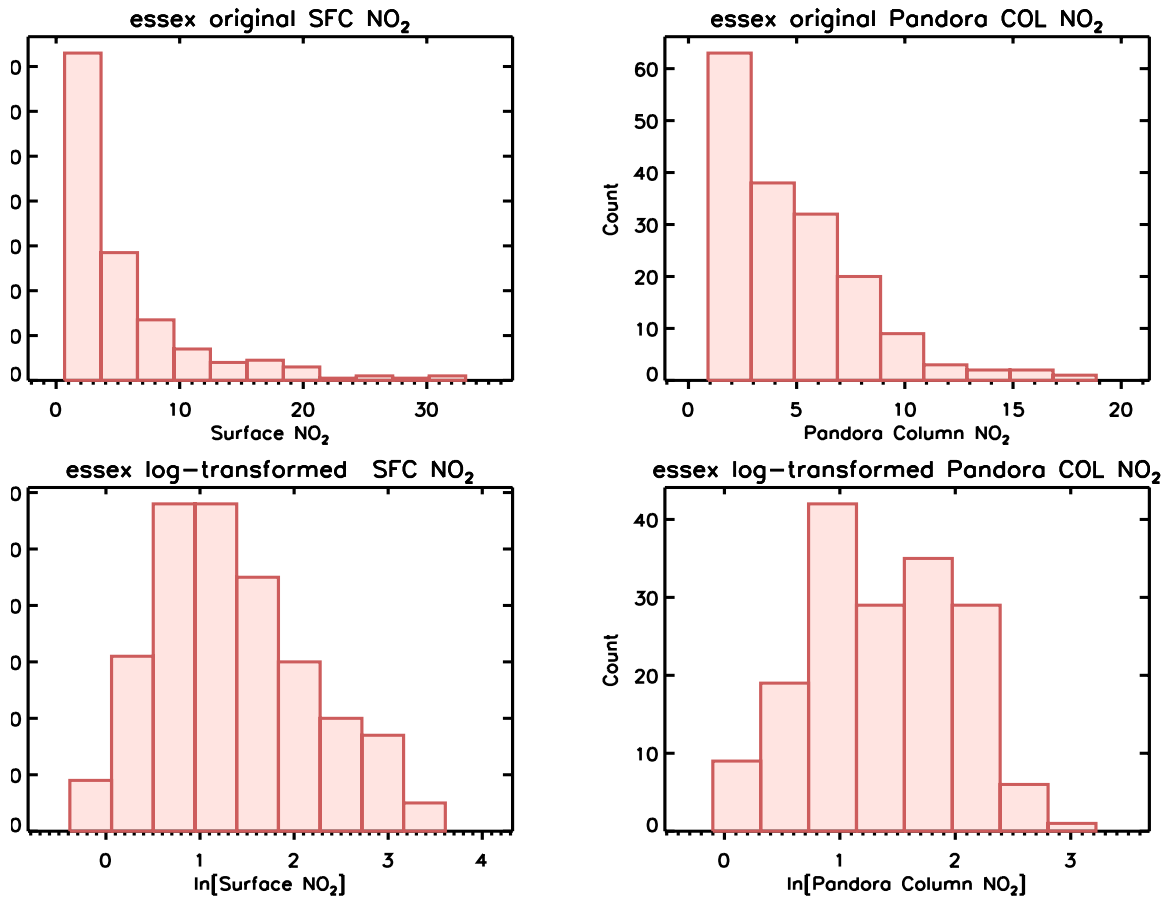
Figure 2.4: Example histograms for the surface NO$_2$ mixing ratio data and Pandora NO$_2$ tropospheric column data. Untransformed data distributions plotted in top row, and log-transformed data in bottom row. Surface data in left column and column data in right column.

2.6.2: Multivariate Linear Least Squares Regression Analysis for the P-3B and Pandora

A multivariate linear least squares regression analysis was performed for P-3B column_air, column_ground, and Pandora O$_3$ and NO$_2$. Column abundance and inverse PBL height (1/PBLH) were used as predictor variables. This yielded an equation of the form

$$\text{surface VMR} = \beta_1 * (\text{column}) + \beta_2 * (\text{PBL}^{-1}) + \text{intercept} \qquad \text{Eqn. (2.2)}$$

where β_1 is the regression coefficient associated with the column, and β_2 is the regression coefficient associated with the inverse PBL height. To prevent limitation of the available Pandora columns, the Pandora analyses used PBL height estimates derived from the

27

WRF/CMAQ model system. In the ACM2 PBL scheme, the PBL top is diagnosed as the height where the bulk Richardson number computed for the entrainment layer exceeds a critical value, typically set at 0.25.

2.6.3: Normalization by PBL Height for the P-3B and Pandora

The degree of correlation between NO_2 column and surface mixing ratio was re-evaluated after normalization of the P-3B or Pandora columns by the PBL height. Column abundances (molecules/cm^2) were divided by, or normalized by, the concurrent PBL height (cm), yielding an estimate of mean number concentration in the PBL. The PBL height estimates derived from the P-3B potential temperature profile were used with the P-3B analyses, while the Pandora analyses again used PBL height estimates derived from the WRF/CMAQ model system. This approach is similar to that of Knepp et al. (2013), and allowed a comparison between the results presented here and those obtained by Knepp et al. (2013).

2.6.4: Comparison of CMAQ Column versus Surface Relationships to Observations

A similar simple linear least squares regression analysis was applied to the Loughner et al. (2013) and NOAA CMAQ O_3 and NO_2 output to assess the correlation between column and surface within these model simulations. CMAQ model output was given in hourly increments. The NO_2 output was log-transformed before analysis. Additionally, correlation analyses for O_3 and NO_2 were also performed for several different conditions to further elucidate differences between the observations and the model. First, P-3B, Pandora, and CMAQ O_3 and NO_2 column and surface data were separated by the time of day at which they occurred, yielding a "Morning" group for data occurring before 12 PM EDT, and an "Afternoon" group for data occurring at or after 12 PM EDT. Second, column and surface data were separated by PBL height, yielding a "High PBL" group for data occurring when the PBL height was at or above 1000 m, and a "Low

PBL" group for data occurring when the PBL was below 1000 m. Estimates of PBLH based on the observed potential temperature profile were again used with the P-3B analysis, while estimates derived from the WRF/CMAQ system were used with the Pandora and CMAQ analyses. The correlation within the CMAQ model for both trace gases was compared to the correlation within the observations for each of these four data groups. Lastly, the CMAQ NO_2 columns were normalized by the concurrent PBL height estimate, and the results for the correlation were re-evaluated.

2.6.5: Significance Tests for Correlation Analyses

2.6.5.1: F-test to Test Significance of the Correlation Analyses

An F-test of overall significance was used to test the significance of the regression analyses presented in this chapter and in all subsequent correlation analyses. The null hypothesis tested states that all regression coefficients β (such as column abundance or inverse PBLH) are equal to zero:

$$\beta_1 = \beta_1 = \ldots = 0 \tag{2.1}$$

The alternative hypothesis states that at least one regression coefficient is not equal to zero:

$$B_j \neq 0 \tag{2.2}$$

The F-statistic is computed by the Interactive Data Language (IDL) algorithm REGRESS.pro, by setting the FTEST flag when calling the program. The F-statistic represents the ratio of explained variance to the unexplained variance. The degrees of freedom (DF) are computed as follows:

$$DF = \text{(number of data pairs in regression)} - \text{(number of predictors)} -1 \tag{2.3}$$

A significance level α of 0.05 (hence confidence level of 95%) was chosen for all significance tests. The sample F-statistic is then compared to the F-distribution, as computed by the IDL with the algorithm F_PDF.pro, for the appropriate degrees of freedom and significance level. This algorithm returns the p-value associated with the sample F-statistic, and is compared to the significance level to determine the significance of the regression: if the p-value is greater than 0.05, the regression is not significant, and if less than 0.05, the regression is significant. Because only one regression coefficient is used in the majority of correlation analyses (column amount used to predict surface value) in this chapter and all subsequent chapters, an insignificant regression indicates an insignificant correlation. In the case of the multivariate regression analyses with two predictors, this test determines only if at least one of the predictors is statistically different from zero, but does not indicate which predictors are different from zero.

2.6.5.2: Z-test to Compare Two Correlation Coefficients

To compare two correlation coefficients to determine if they are statistically significantly different from each other, a Fisher R-to-Z transformation is employed. Each correlation coefficient R_i is first transformed to a Z-value through

$$Z_i = \frac{1}{2} \ln \left(\frac{(1+R_i)}{(1-R_i)} \right) \tag{2.4}$$

The sample Z-statistic is then computed from the Z-values for each correlation coefficient and the number of data points used in each correlation N_i:

$$Z = \left. (Z_1 - Z_2) \middle/ \sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-1}} \right. \tag{2.5}$$

The Z-statistic assumes a Gaussian distribution. The p-value associated with the sample Z-statistic is computed from the Gaussian distribution with the IDL algorithm GAUSS_PDF.pro. A significance level of 0.05 is again employed: if the p-value is less than or equal to 0.05, the two correlations are significantly different from each other, while if it is larger than 0.05, the two correlations are statistically the same.

2.6.5.3: Durbin-Watson Test Statistic to Test for Autocorrelation of Residuals

The Durbin-Watson test statistic is used to test for the presence of autocorrelation of the residuals, or prediction errors, of the regression model. For a perfect regression model, the residuals would approximate the random errors associated with the data used to build the regression model. Thus, correlation among the residuals would violate one of the assumptions inherent in linear regression and indicate that the regression model did not fit the data well. The test statistic d is computed as

$$d = \left. \sum_2^n (e_n - e_{n-1})^2 \middle/ \sum_1^n (e_n)^2 \right. \tag{2.6}$$

where e is the regression residual and n is the number of observations. The test statistic is compared to the lower and upper critical values at the chosen significance level α ($d_{L,\alpha}$ and $d_{U,\alpha}$, respectively):

$d < d_{L,\alpha}$: statistical evidence that the residuals are positively autocorrelated.

$d > d_{U,\alpha}$: no statistical evidence that the residuals are positively autocorrelated.

$d_{L,\alpha} < d < d_{U,\alpha}$: the test is inconclusive.

To test for negative autocorrelation, the quantity (4-d) is used instead of d, and compared to the lower and upper critical values, with the same interpretations.

2.7: Results of Simple Linear Regression Analysis for P-3B, Pandora, and OMI

2.7.1: Evaluation of the Correlation Between Column and Surface for the Full Data Set

The degree of correlation between surface mixing ratio and column abundance found from the simple linear regression analyses for the P-3B, Pandora, and OMI data sets is summarized in Table 2.1. To assign a degree of correlation to an analysis, the correlations for at least four of the six surface sites must have fallen within one of the categories of correlation degree. Values of R^2 are given in Table 2.2 (O_3 analyses) and Table 2.3 (NO_2 analyses), and representative scatter plots of the correlation are displayed in Figure 2.5 (O_3) and Figure 2.6 (NO_2). Most P-3B O_3 (Figure A1 in Appendix A of the Dissertation Supplementary Material**),** P-3B NO_2 (Figure A4), Pandora O_3 (Figure A2), and Pandora NO_2 (Figure A5) regressions were statistically significant at a confidence level of 95% (Tables 2.2-2.3). The simple linear regression analyses performed with the Pandora total column O_3 and NO_2 data were not significantly different from those for the tropospheric column data. The poor correlation between OMI column O_3 or NO_2 and surface data may be partly due to the large OMI footprint size; the pixel size at nadir is 13 x 24 km^2, and increases towards the ends of the OMI swath. The OMI O_3 retrieval also loses sensitivity to the lower troposphere (Liu et al., 2010). The Pandora correlations for both trace gases were also poorer than those for the P-3B likely because of the uncertainties in the Pandora column estimates due to subtraction of the OMI stratospheric column, as well as the possibility that the Pandora instruments sometimes observed different air masses from the P-3B spiral: the Pandora is a sun-synchronous instrument, and so may not have always pointed in the same direction as the P-3B spiral. P-3B NO_2 column_ground demonstrated larger correlation than did column_air, reflecting the influence of the surface data in the column computation; P-3B O_3 column_ground and column_air demonstrated similar values. This indicates that O_3 is vertically and horizontally better mixed than NO_2 at each site. The Aldino

column_UMD correlations were not significantly different from those for column_air, but were significantly smaller than those for column_ground for both gases; this suggests that the column_air analyses were more representative of the true correlation between lower tropospheric column and the surface (Tables 2.2 and 2.3). Figure 2.7 displays scatter plots of NO_2 mixing ratio measured at the lowest P-3B aircraft measurement altitude and surface NO_2 mixing ratio measurements coincident to each spiral at each spiral site (when both measurements were available). These plots demonstrate that the surface values were typically larger than the coincident P-3B lowest altitude values, and that they typically did not correlate well. As the values of χ^2_{red} are typically close to or less than a value of 1.0 for Edgewood, Essex, Fair Hill, and Padonia, with the data points lying in a cluster with no correlation and larger error bars for either the surface or aircraft data (overestimated uncertainty), this suggests that the aircraft and surface measurements were scattered about the same mean; thus, the bias between aircraft and surface NO_2 values may be due to the NO_2 mixing ratio vertical gradient in the lowest 300 m below the P-3B aircraft (Section 2.2.1.2 briefly discusses NO_2 profile below the P-3B). This likely contributed to the poor correlation between column and surface data. In contrast, Aldino and Beltsville displayed a low degree of correlation between surface and aircraft data, but with values of χ^2_{red} less than 1.0; in this case, the surface and aircraft data may be scattered about different means and thus representative of different chemical conditions, as the surface data at these two sites were actually NO_2 with interferences from NO_y species, unlike the other four sites. This likely impacted the bias between aircraft and surface at these two sites and the column-surface correlations. Extended analysis of the simple linear regressions is also presented in Appendix A.
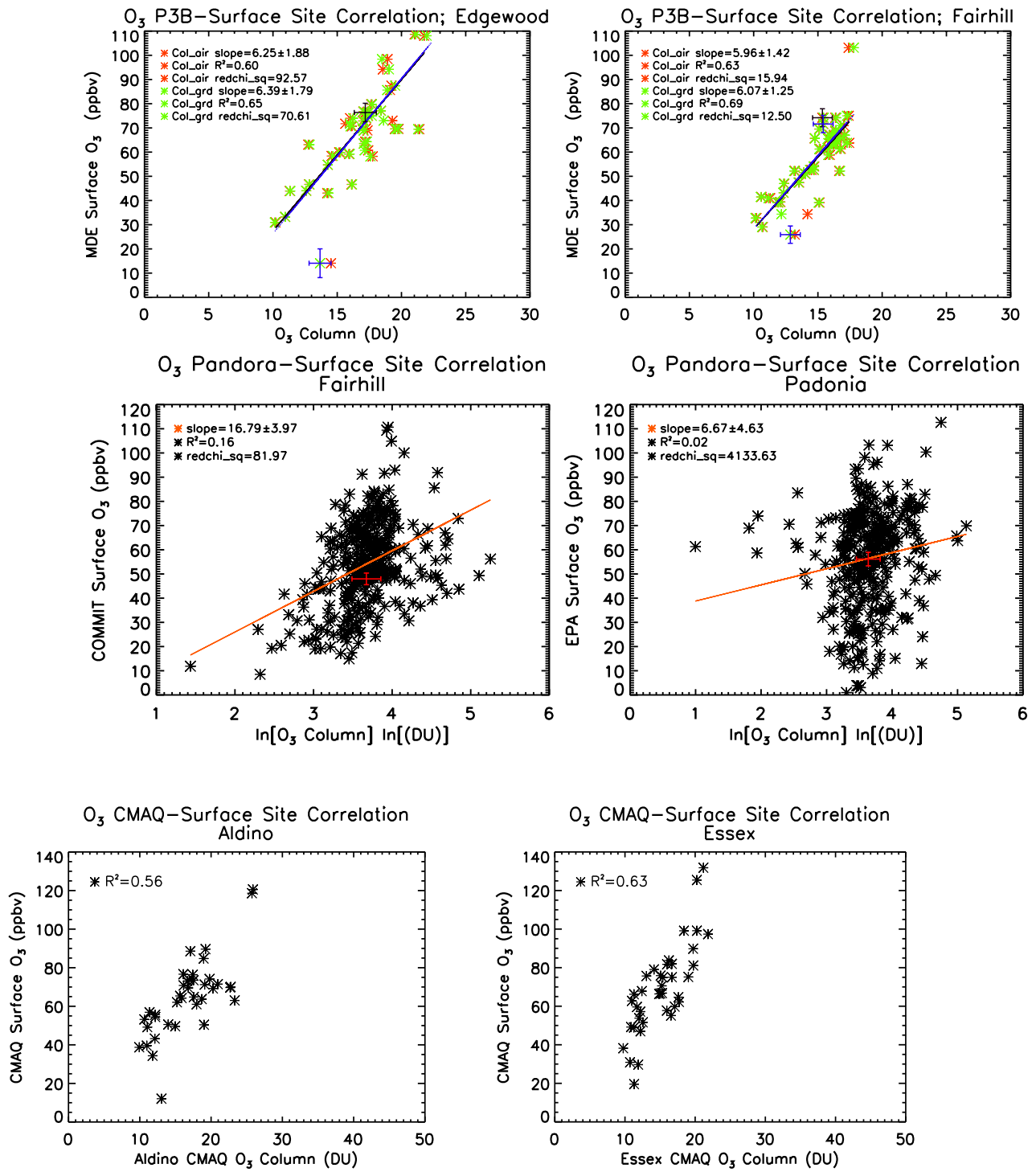
Figure 2.5: Example scatter plots of O3 column vs. surface O3 mixing ratio for P-3B, (top) Pandora (middle), and CMAQ (Loughner et al, 2013, bottom) correlation analyses. Plots chosen represent the most typical behavior of the column-surface relationship for that

data set. Correlation shown between all available column and surface data for each data set from the simple linear regression analysis. Slopes of the regression, χ^2_{red} values, and R^2 values displayed in the upper left corner of each plot for the P-3B and Pandora; R^2 values displayed in the upper left corner of each plot for CMAQ. Example uncertainty bars displayed data points for the P-3B (in black for column_air and blue for column_ground) and Pandora (in red) plots.

Figure 2.6: Example scatter plots of NO₂ column vs. surface NO₂ mixing ratio for P-3B, (top) Pandora (middle), and CMAQ (Loughner et al, 2013, bottom) correlation analyses. Plots chosen represent the most typical behavior of the column-surface relationship for that data set. Correlation shown between all available column and surface data for each data set from the simple linear regression analysis. Slopes of the regression, χ^2_{red} values, and R^2 values displayed in the upper left corner of each plot for the P-3B and Pandora; R^2 values displayed in the upper left corner of each plot for CMAQ. Example uncertainty bars displayed data points for the P-3B (in black for column_air and blue for column_ground) and Pandora (in red) plots.

36

Figure 2.7: Scatter plots of NO₂ at lowest P-3B measurement altitude vs. surface NO₂ mixing ratio for each spiral site correlation analysis. Correlation shown between all available data for each site. χ^2_{red} values and R² values displayed in the upper left corner of each plot; line displayed is the 1:1 line. Example uncertainty bars for surface and aircraft NO₂ data also displayed on one data point.

	NO$_2$	O$_3$
P-3B Col_air	Low	High
P-3B Col_ground	Moderate	High
Pandora	Moderate	Low
OMI	Not Significant	Not Significant
CMAQ (Loughner et al.)	High	Moderate
CMAQ (NOAA)	High	High

Table 2.1: Summary of degree of correlation found from the simple linear regression analyses between column amounts and surface mixing ratios. Low correlation: R^2=0-0.16; Moderate: R^2=0.16-0.64; High: R^2=0.64-1.0.

	P-3B Col_Air R^2	P-3B Col_Ground R^2	P-3B Col_Air F-ratio	P-3B Col_Ground F-ratio	P-3B Column_UMD R^2	P-3B Column_UMD F-ratio	Pandora R^2	Pandora F-ratio
Aldino	0.76	0.79	112.57 (<0.001)	131.81 (<0.001)	0.73	98.9 (<0.001)	0.06	21.72 (<0.001)
Beltsville	0.83	0.88	192.10 (<0.001)	267.95 (<0.001)	--	--	0.04	12.8 (<0.001)
Edgewood	0.61	0.65	62.94 (<0.001)	77.17 (<0.001)	--	--	0.01	3.53 (0.057)
Essex	0.58	0.63	52.25 (<0.001)	61.74 (<0.001)	--	--	0.03	9.39 (0.002)
Fair Hill	0.64	0.70	72.58 (<0.001)	95.47 (<0.001)	--	--	0.16	68.98 (<0.001)
Padonia	0.65	0.72	60.25 (<0.001)	85.58 (<0.001)	--	--	0.02	8.02 (0.004)

Table 2.2: Summary of the R^2 statistic and F-ratio (p-value) for the P-3B and Pandora O$_3$ simple linear regressions.

	P-3B Col_Air R^2	P-3B Col_Grd R^2	P-3B Col_Air F-ratio	P-3B Col_Grd F-ratio	P-3B Col_UMD R^2	P-3B Col_UMD F-ratio	Pandora R^2	Pandora F-ratio
Aldino	0.13	0.78	4.93 (0.03)	119.58 (<0.001)	0.16	6.48 (<0.016)	0.01	4.75 (0.030)
Beltsville	0.13	0.55	6.16 (0.02)	50.44 (<0.001)	--	--	0.20	80.94 (<0.001)
Edgewood	0.02	0.56	0.62 (0.429)	43.11 (<0.001)	--	--	0.21	61.03 (<0.001)
Essex	0.05	0.37	0.812 (0.380)	9.59 (0.009)	--	--	0.29	68.57 (<0.001)
Fair Hill	0.18	0.80	4.69 (0.040)	86.40 (<0.001)	--	--	0.09	7.43 (0.009)
Padonia	0.07	0.49	2.22 (0.148)	30.70 (<0.001)	--	--	0.27	94.69 (<0.001)

Table 2.3: Summary of the R^2 statistic and F-ratio (p-value) statistic for the P-3B and Pandora NO_2 simple linear regressions.

2.7.2.1: Evaluation of the Errors of the Simple Linear Regression Model

An overview of the average error of the regressions relative to the observations for O_3 and NO_2 is presented for the P-3B data sets. The column_air- and column_ground-measured surface values were first combined into one data set, as were the regression estimated surface values for the column_air and column_ground regression analyses, before computation of the average percentage error of the regression relative to the observations. The average error for P-3B O_3 was typically less than 10% at each site, with the exception of Padonia; this was due to the presence of a very low surface observation that was not a statistical outlier. Additionally, approximately 50 to 75% of regression estimations fell within a ±10% error of the observed value (Table 2.4). These results support the conclusions presented for P-3B O_3 in the previous

section. The average error for the Pandora O_3 regressions, however, was much more variable among sites, and could be much larger than seen for the P-3B results. The percentage of estimations falling within ±10% error was typically less than 25% (Table 2.5). The Durbin-Watson test statistic was used to test for the presence of autocorrelation of the residuals, which would violate the assumption of independent regression errors. All Pandora O_3 regressions demonstrated positive autocorrelation and large average errors, indicating errors in the computation of the Pandora tropospheric column O_3. This may be due to subtraction of the OMI stratospheric column, which may not be representative of the true column at each surface site due to the large OMI footprint size.

	O_3 Mean Error	% of Cases w/in ±10% Error	NO_2 Mean Error	% of Cases w/in ±10% Error	% of Cases w/in ±50% Error
Aldino	3.1 (± 24.6) %	61.3%	9.7 (± 37.2) %	28.6%	88.5%
Beltsville	6.0 (± 37.4) %	74.1%	3.2 (± 18.7) %	47.7%	98.8%
Edgewood	7.1 (± 44.9) %	51.2%	-58.2 (± 607.3) %	0.00%	26.3%
Essex	6.1 (± 38.2) %	59.7%	36.5 (± 83.4) %	11.1%	52.8%
Fair Hill	2.5 (± 18.8) %	63.9%	4.2 (± 416.0) %	10.4%	39.6%
Padonia	40.2 (± 236.8) %	61.4%	4.0 (± 93.9) %	21.1%	63.6%

Table 2.4: Summary of percentage errors (standard deviation) simple linear regression for all sites relative to observed surface values. Column_air and column_ground are analyzed together for each site.

	O_3 Mean Error	% of Cases w/in ±10% Error	NO_2 Mean Error	% of Cases w/in ±10% Error	% of Cases w/in ±50% Error
Aldino	16.9 (± 66.6)%	24.8%	27.3 (± 88.5)%	24.7%	81.0%
Beltsville	61.9 (± 270.9)%	21.0%	20.4 (± 59.6)%	22.1%	74.6%
Edgewood	29.9 (± 112.9)%	23.6%	-125.5 (± 2381.7)%	4.3%	28.1%
Essex	30.6 (± 116.6)%	23.9%	58.1 (± 374.1)%	11.2%	62.9%
Fair Hill	12.1 (± 43.7)%	23.9%	-382.6 (± 2554.9)%	6.8%	44.6%
Padonia	47.6 (± 318.4)%	23.1%	119.5 (± 941.2)%	15.4%	62.3%

Table 2.5: Summary of percentage errors of Pandora simple linear regression for all sites relative to observed surface values.

The P-3B NO_2 regressions resulted in an average percentage error relative to the observations similar to the O_3 regressions at most sites, with the exceptions of Edgewood and Essex. Less than 30% of regression estimations fell within a ±10% error of the observed value except at Beltsville; a typically larger but more variable percentage fell within ±50% error of the observed value (Table 2.4). However, the Pandora NO_2 regressions displayed larger average errors than the Pandora O_3 regressions except at Beltsville, and larger average errors than the P-3B NO_2 regressions. Approximately 50% or more of regression estimations fell within ±50% error of the observed value at most sites (Table 2.5). Plots of the regression residuals revealed other problems with this simple linear regression analysis for P-3B and Pandora NO_2. The Durbin-Watson test statistic again indicated positive autocorrelation of the residuals in the Pandora NO_2 regressions. Lag-1 residual correlation plots are another tool used to indicate

graphically the presence of autocorrelation of the regression residuals, Each residual is plotted against the residual immediately preceding it in time (for a time step of 1); if the errors of the regression are random, then structure should be present in these plots. However, structure is evident in the lag-1 plots of the Pandora residuals at all sites, indicating the presence of autocorrelation of the residuals and a violation of the assumptions of linear regression (Fig. 2.8). Histograms of column_air residuals and Pandora NO_2 residuals demonstrated some deviation from normality at some sites. These problems with the simple linear regression are more severe for Pandora, but further suggest that a simple linear regression is not as appropriate for NO_2 as for O_3.
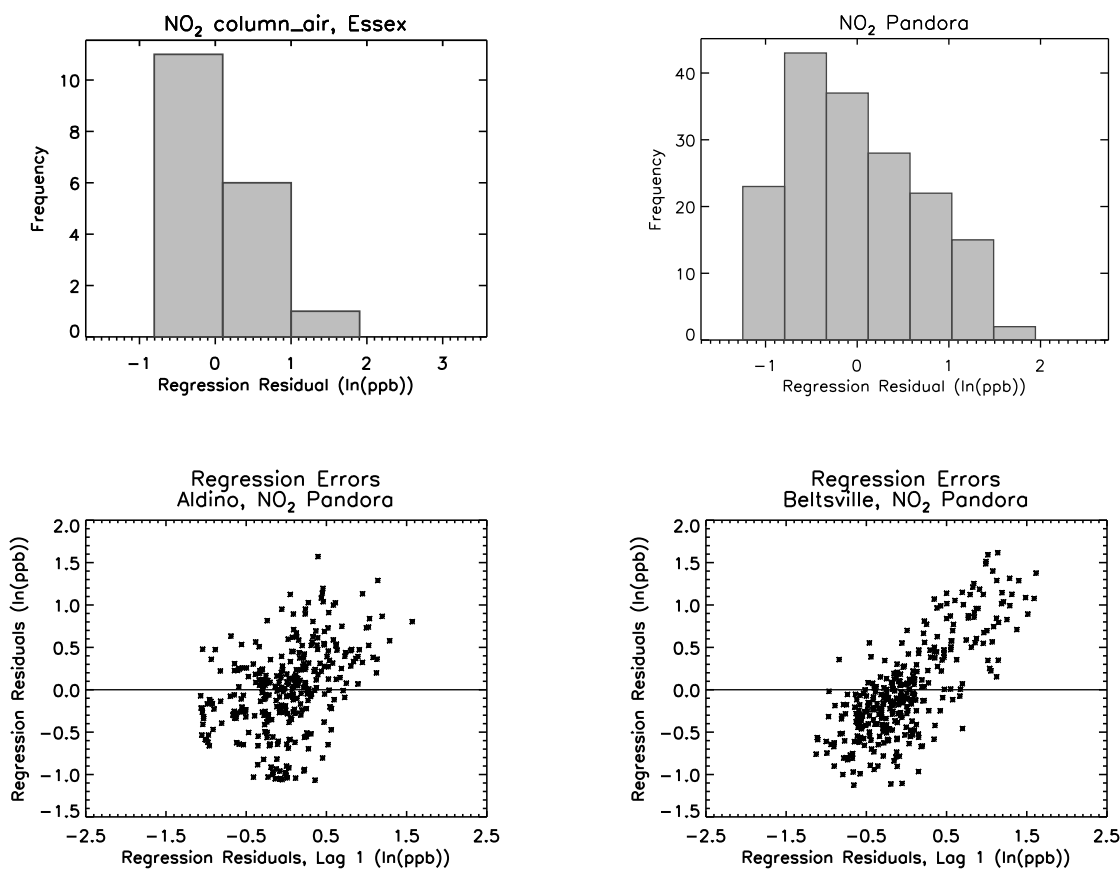
Figure 2.8: Example scatter plots for NO₂ simple linear regression residuals. (top) Histograms of NO₂ residuals for Essex P-3B column_air and Pandora. (bottom) Pandora NO₂ residuals plotted against the lagged-1 residuals at Aldino and Beltsville.

Simple linear regression analyses were also performed for P-3B and Pandora O_3 and NO_2 for the aggregate of the data over all sites; the values of R^2 for the correlations between column and surface data for the aggregate analyses are displayed in Table 2.6. The values of R^2 for the aggregate P-3B NO_2 column_air, NO_2 column_ground, Pandora O_3, and Pandora NO_2 regressions were similar to those for the individual sites for these datasets; the aggregate R^2 values were much smaller than those for the individual sites for P-3B O_3 column_air and column_ground. All P-3B and Pandora O_3 and NO_2 aggregate analyses suffered from positive autocorrelation of the residuals. The average percentage error for the aggregate P-3B O_3 analysis

was larger than for the individual sites, and the percentage of estimations within a ±10% error of the observed value was smaller; the average errors for the aggregate P-3B NO_2, Pandora O_3, and Pandora NO_2 regressions were similar to the individual sites. However, the standard deviations of the percentage errors for the aggregate P-3B O_3 and NO_2 regressions were larger compared to the individual sites (Table 2.7). Aggregation of the data thus worsened the regressions for the P-3B datasets, relative to the individual sites, and did not result in an improved regression for the Pandora datasets.

	O_3 R^2	O_3 F-Ratio	NO_2 R^2	NO_2 F-Ratio
P_3B Col_Air	0.26	83.41 (<0.001)	0.01	2.35 (0.126)
P_3B Col_Air	0.28	89.66 (<0.001)	0.43	140.17 (<0.001)
Pandora	0.04	81.86 (<0.001)	0.14	4.75 (0.030)

Table 2.6: Summary of the R^2 statistic and F-ratio (p-value) statistic for the P-3B and Pandora simple linear regressions for the aggregate over all data.

	O_3 Mean Error	% of Cases w/in ±10% Error	NO_2 Mean Error	% of Cases w/in ±10% Error	% of Cases w/in ±50% Error
P-3B	17.6 (± 117.5)%	33.5%	-4.6 (± 1155.9)%	7.6%	56.3%
Pandora	33.9 (± 190.5)%	80.9%	-10.8 (± 1807.3)%	50.4%	85.5%

Table 2.7: Summary of percentage errors (standard deviation) simple linear regression for the aggregate over all data relative to observed surface values. Column_air and column_ground are analyzed together for each site for the P-3B analyses.

2.7.2.2: Evaluation of the Slopes for the Simple Linear Regression Model

Tables 2.8-2.10 list the slopes and values of the reduced chi-squared (χ^2_{red}) goodness-of-fit test obtained from the simple linear regression analyses at the individual spiral sites and for the aggregate analyses for P-3B and Pandora O_3 and NO_2. The reduced chi-squared statistic

evaluates how well the linear regression model captured the observed data, to within the uncertainty of the data:

$$\chi^2_{red} = \left(\frac{1}{N} - n\right) * \sum_{n=i}^{N}(\frac{y_i - fit(x_i)}{\sigma_i})^2 \qquad (2.7)$$

where N is the total number of data points and n is the index of each data point. A reduced chi-squared value near 1.0 indicates that the regression well captured the relationship to within the data uncertainty, while a value much less than 1.0 indicates that the uncertainty may have been overestimated, that the regression model overfit the data, or both, and a value much greater than 1.0 indicates that the uncertainty was underestimated, the regression underfit the data, or both. Column_air and column_ground O_3 presented very similar values for the slopes and the uncertainties of the slopes at each of the individual sites and for the aggregate analyses, which is consistent with the similar values of R^2 obtained for these two datasets. Additionally, the confidence intervals for both the column_air and column_ground slopes overlapped for almost all spiral sites, indicating that the slopes were not statistically different among the sites; Beltsville was the exception to this and presented a statistically distinct slope for these two P-3B data sets, as its confidence interval did not overlap with the intervals of any other site. Beltsville also presented the steepest P-3B slopes, which may be partially due to greater O_3 photochemical production at this site relative to the other sites. The Pandora O_3 regressions also did not produce statistically distinct slopes at most spiral sites. Further, the confidence interval of the aggregate Pandora O_3 slope overlapped with the intervals for all individual sites, indicating that the aggregate Pandora regression produced the same result as those for the individual sites, while the aggregate column_air and column_ground O_3 slopes overlapped with only the Edgewood and Essex slopes. It is also notable that the P-3B O_3 regressions displayed smaller uncertainties in the slopes, and thus smaller confidence intervals, than did the Pandora regressions; the same is

true of the comparison of the values of χ^2_{red} between the P-3B and Pandora. This is again likely due to the uncertainties in the Pandora tropospheric column that also impacted the Pandora column-surface correlations: uncertainty due to subtraction of the OMI stratospheric column from the Pandora total column, and the distinct possibility that the Pandora instruments sometimes observed different air masses from the P-3B spiral: Pandora is a sun-synchronous instrument, and if the Pandora instrument was pointed in a different direction following the sun than the aircraft spiral overhead of the instrument, the Pandora may have picked up a different air mass from that sampled by the P-3B. However, the χ^2_{red} values indicate some advantage of the regression at Fair Hill for the P-3B and Pandora analyses; however, it should be noted that these values were smallest at the Fair Hil for both the P-3B and Pandora analyses relative to the other sites, but had large magnitudes, indicating inadequacy of the regression model to fit the data. Thus, full and lower tropospheric column amounts may be best related to surface mixing ratios at a relatively rural, clean site such as Fair Hill. However, the values of the Pandora slopes were smallest at Edgewood and Essex; these two sites were also fairly polluted sites, and experienced bay breezes; bay breezes tend to accumulate pollution over nearby coastal sites, as they interrupt horizontal transport, and so this "keeping pollution in place" over Edgewood and Essex may have contributed to the small slopes and greater ability to relate surface and column abundances from the perspective of a remote sensing instrument retrieving the full tropospheric column..

The confidence intervals for the slopes for the NO_2 column_air regressions overlapped for all individual sites, and for most individual sites for the column_ground regressions (Table 2.9). Additionally, the confidence intervals of the aggregate NO_2 column_air and column_ground regressions overlapped with the intervals for almost all individual sites, indicating no statistical

difference between a regression built for an individual site and for the entire study region. However, the confidence intervals for the Edgewood, Essex, Padonia, and aggregate column_air analyses included 0.0, indicating that these slopes are not statistically significant and no predictive relationship exists between the column and surface data for these analyses; the confidence intervals for column_ground did not include 0.0, but it should again be noted that the NO_2 column_ground estimates are likely unrealistic. In contrast, the slopes were statistically distinct from those of all other sites and the aggregate analysis at Aldino and Edgewood for the Pandora analyses (Table 2.10). It is notable the magnitude of the slope for Aldino was the smallest of all sites for the Pandora regressions, while the slope was largest at Edgewood; the confidence intervals for the Pandora and column_air Aldino slopes overlapped, indicating that these slopes were statistically the same despite the use of full tropospheric column data in one regression and lower tropospheric column data in the other. This should perhaps not be surprising, as most of the NO_2 column burden resides in the lower atmosphere. The values of χ^2_{red}, however, did not suggest a clear advantage of the regression model at any individual site for either the Pandora or the P-3B, though they do suggest that an aggregate regression over a large region, such as the Baltimore-Washington metropolitan region, is inappropriate for NO_2, likely due to greater spatial heterogeneity than for O_3.

	Column_Air O_3 Slope (uncertainty)	Column_Ground O_3 Slope (uncertainty)	Column_Air O_3 χ^2_{red}	Column_Ground O_3 χ^2_{red}
Aldino	7.69 (1.46)	7.65 (1.36)	25.86	22.43
Beltsville	10.96 (1.59)	10.49 (1.30)	73.77	39.15
Edgewood	4.96 (1.88)	5.12 (1.79)	92.57	70.61
Essex	4.56 (1.29)	4.77 (1.23)	65.50	52.54
Fair Hill	5.97 (1.42)	6.07 (1.25)	15.94	12.50
Padonia	7.12 (1.86)	7.15 (1.57)	2706.42	1845.86
Aggregate	2.69 (0.60)	2.76 (0.60)	596.66	539.12

Table 2.8: Summary of slopes and χ^2_{red} values for each of the individual spiral sites and the aggregate analysis for P-3B column_air and column_ground O_3. Uncertainties of the slope given in parentheses next to the slope values.

	Column_Air NO_2 Slope	Column_Ground NO_2 Slope	Column_Air NO_2 χ^2_{red}	Column_Ground NO_2 χ^2_{red}
Aldino	0.41 (0.38)	1.07 (0.20)	675.40	157.42
Beltsville	0.26 (0.21)	0.69 (0.20)	16.34	6.99
Edgewood	0.25 (0.62)	1.32 (0.41)	4.55	1.12
Essex	-0.28 (0.66)	0.89 (0.61)	719.30	566.02
Fair Hill	1.05 (1.00)	1.67 (0.37)	1775.98	188.22
Padonia	0.32 (0.43)	1.19 (0.44)	240.12	87.55
Aggregate	0.17 (0.22)	1.08 (0.18)	456.13	241.56

Table 2.9: Summary of slopes and χ^2_{red} values for each of the individual spiral sites and the aggregate analysis for P-3B column_air and column_ground NO_2. Uncertainties of the slope given in parentheses next to the slope values.

	Pandora O_3 Slope	O_3 χ^2_{red}	Pandora NO_2 Slope	NO_2 χ^2_{red}
Aldino	13.09 (5.52)	188.55	0.11 (0.10)	9.67
Beltsville	12.77 (7.01)	3079.45	0.65 (0.14)	1305.69
Edgewood	4.31 (4.51)	543.82	0.94 (0.24)	80.74
Essex	8.07 (5.18)	579.81	0.70 (0.17)	19.57
Fair Hill	16.79 (3.97)	81.97	0.38 (0.28)	21.72
Padonia	6.67 (4.63)	4133.63	0.45 (0.10)	8.44
Aggregate	9.42 (2.04)	1496.06	0.43 (0.05)	327.37

Table 2.10: Summary of slopes and χ^2_{red} values for each of the individual spiral sites and the aggregate analysis for Pandora O_3 and NO_2. Uncertainties of the slope given in parentheses next to the slope values.

2.7.3: Multivariate Linear Least Squares Regression Analysis for the P-3B and Pandora

All P-3B O_3 column_air and column_ground multivariate regressions were significant at a confidence level of 95%, and the associated R^2 values demonstrated modest improvement over those for the simple linear regressions (Table 2.11). The average percentage errors and standard deviations were consistently smaller than for the simple linear regressions, indicating that the range of the residuals had decreased. Likewise, the percentage of estimations falling within a ±10% error of the observed value was somewhat larger than or similar to the percentage for the simple linear regression at each site (Table 2.12). All Pandora O_3 regressions were significant, and demonstrated larger improvement relative to the simple linear regressions than did the P-3B regressions. The average percentage errors and standard deviations were much smaller, and the percentage of cases falling within a ±10% error of the observed value much larger (Tables 2.13, 2.14). However, the Durbin-Watson results for Essex column_air and column_ground O_3 and all Pandora O_3 indicated positive autocorrelation of residuals. Some structure and fanning behavior, in which the range of residuals either increases or decreases as the abscissa increases, was also

present in plots of the residuals against inverse PBL height for Pandora, indicating limitations of

this regression for the Pandora O_3 data (Figure 2.8).
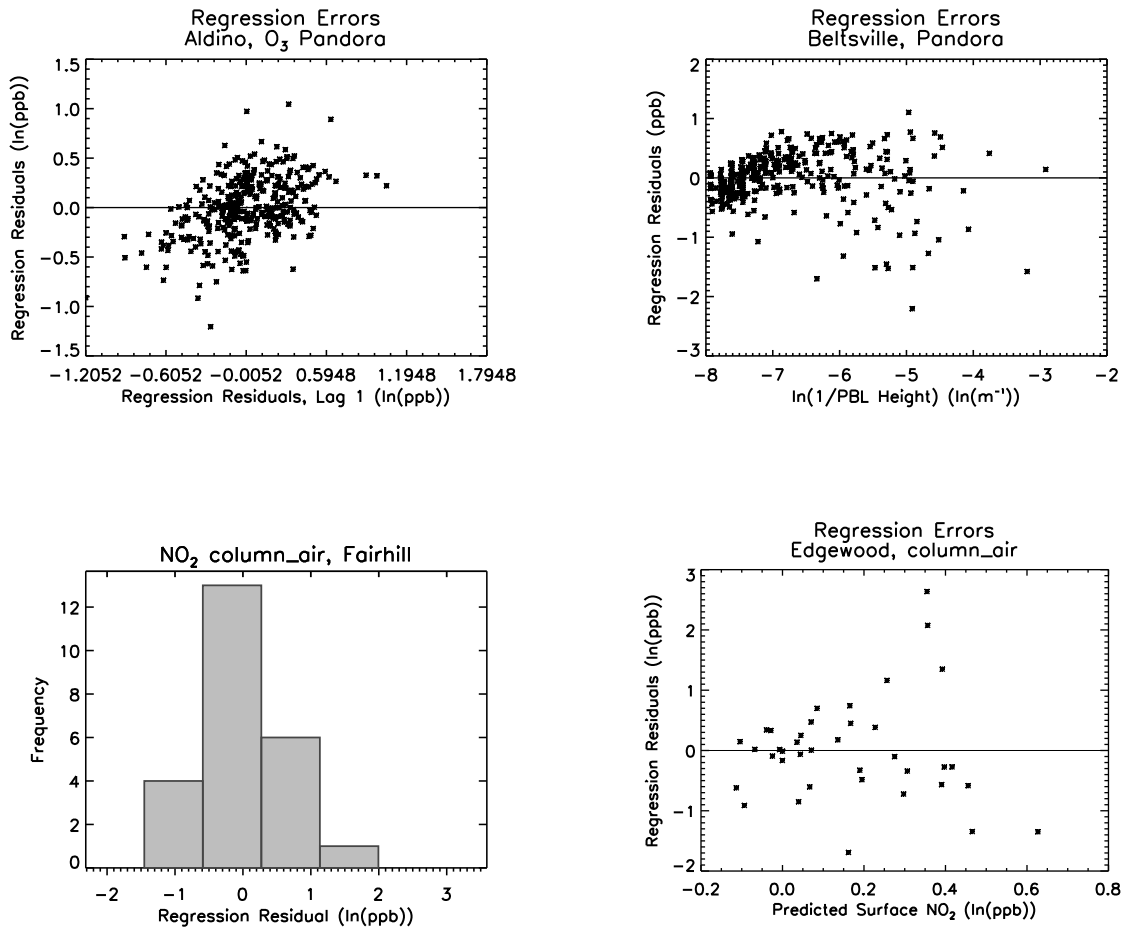
Figure 2.9: Example scatter plots for O_3 and NO_2 multivariate residuals. (top) Pandora O_3 residuals plotted against lagged-1 residuals at Aldino and against the logarithm of the inverse PBL height at Beltsville. (bottom) Histogram of residuals for Fair Hill NO_2 column_air regression and plot of residuals vs. predicted surface NO_2 for Edgewood column_air regression for the multivariate regression analysis.

	P-3B O$_3$ Col_Air R^2	P-3B O$_3$ Col_Ground R^2	P-3B O$_3$ Col_Air F-ratio	P-3B O$_3$ Col_Ground F-ratio	P-3B NO$_2$ Col_Air R^2	P-3B NO$_2$ Col_Ground R^2	P-3B NO$_2$ Col_Air F-ratio	P-3B NO$_2$ Col_Ground F-ratio
Aldino	0.82	0.83	75.62 (<0.001)	81.87 (<0.001)	0.30	0.78	6.69 (0.003)	55.32 (<0.001)
Beltsville	0.90	0.93	169.79 (<0.001)	227.91 (<0.001)	0.44	0.66	14.88 (<0.001)	37.58 (<0.001)
Edgewood	0.67	0.70	38.36 (<0.001)	44.48 (<0.001)	0.05	0.63	0.85 (0.437)	26.47 (<0.001)
Essex	0.72	0.74	42.99 (<0.001)	47.24 (<0.001)	0.21	0.53	1.91 (0.182)	8.17 (0.009)
Fair Hill	0.66	0.69	37.64 (<0.001)	43.09 (<0.001)	0.19	0.80	2.52 (0.104)	41.53 (<0.001)
Padonia	0.74	0.78	46.66 (<0.001)	57.47 (<0.001)	0.43	0.63	11.23 (<0.001)	25.58 (<0.001)

Table 2.11: Summary of the R^2 statistic and F-ratio (p-value) for the P-3B O$_3$ and NO$_2$ multivariate regressions.

	O$_3$ Mean Error	% of Cases w/in ±10% Error	NO$_2$ Mean Error	% of Cases w/in ±10% Error	% of Cases w/in ±50% Error
Aldino	2.0 (± 18.2) %	54.8%	8.3 (± 33.3) %	22.1%	89.7%
Beltsville	3.5 (± 23.6) %	77.2%	2.2 (± 15.6) %	52.4%	98.8%
Edgewood	5.8 (± 34.9) %	62.2%	-89.2 (± 439.3) %	5.6%	20.8%
Essex	3.9 (± 24.4) %	60.3%	30.6 (± 81.8) %	11.8%	61.8%
Fair Hill	2.3 (± 18.6) %	67.9%	11.6 (± 436.4) %	8.3%	43.8%
Padonia	26.5 (± 154.1) %	58.6%	0.8 (± 71.7) %	21.2%	72.8%

Table 2.12: Summary of percentage errors of P-3B multivariate regression for all sites relative to observed surface values. Col_air and col_ground are analyzed together for each site.

	O_3 R^2	O_3 F-ratio	NO_2 R^2	NO_2 F-ratio
Aldino	0.42	130.0 (<0.001)	0.34	84.2 (<0.001)
Beltsville	0.54	201.7 (<0.001)	0.68	354.2 (<0.001)
Edgewood	0.36	89.2 (<0.001)	0.52	123.7 (<0.001)
Essex	0.25	47.7 (<0.001)	0.57	112.7 (<0.001)
Fair Hill	0.47	155.7 (<0.001)	0.27	13.5 (<0.001)
Padonia	0.51	183.3 (<0.001)	0.59	85.3 (<0.001)

Table 2.13: Summary of the R^2 statistic and F-ratio (p-value) for the Pandora O_3 and NO_2 multivariate regressions.

	O_3 Mean Error	% of Cases w/in ±10% Error	NO_2 Mean Error	% of Cases w/in ±10% Error	% of Cases w/in ±50% Error
Aldino	0.68 (± 8.6) %	82.6%	19.9 (± 71.4) %	26.2%	84.6%
Beltsville	3.2 (± 27.6) %	72.7%	9.8 (± 39.3) %	31.0%	85.9%
Edgewood	1.4 (± 13.9) %	78.9%	-131.7 (± 1027.3) %	6.9%	42.4%
Essex	1.8 (± 16.6) %	72.9%	38.2 (± 267.3) %	17.6%	67.1%
Fair Hill	0.56 (± 7.6) %	85.3%	- (± 1699.0) %	5.4%	43.2%
Padonia	1.4 (± 17.7) %	79.8%	62.6 (± 465.0) %	17.7%	68.8%

Table 2.14: Summary of percentage errors of Pandora multivariate regression for all sites relative to observed surface values.

The regressions for P-3B NO_2 column_ground at all sites and for NO_2 column_air at half of the sites were significant at a confidence level of 95%; the R^2 values also improved (Table

2.11). Like the P-3B O_3 results, the average percentage error and associated standard deviation decreased relative to the simple linear regression at most sites. Most sites also saw an increased percentage of regression estimations falling within a ±10% error and ±50% error of the observed value (Table 2.12). The Pandora NO_2 regressions also demonstrated marked improvement in the average percentage errors and the standard deviations at most sites. However, the percentage of regression estimations falling within a ±10% error and ±50% error of the observed value demonstrated marginal improvement (Tables 2.13, 2.14). These results indicate that both the column and inverse PBL height contain useful information for NO_2. Fewer sites displayed histograms of residuals for column_air and Pandora NO_2 that departed from normality, and fewer sites indicated autocorrelation of the residuals for Pandora. However, plots of the residuals for column_air and column_ground NO_2 against predicted surface NO_2 at Edgewood displayed some fanning structure (Figure 2.8). Though some improvement to the regression model is needed, these results indicate that the inverse PBL height adds useful information for the O_3 and NO_2 regressions, and thus mixing within the PBL has an important impact on the column-surface relationship for these gases. Because future geostationary air quality satellites will capture the diurnal cycle of their observations, these results further imply that care should be taken for the impact of PBL development on column quantities.

2.7.4: Normalization by PBL Height for the P-3B and Pandora

The normalization of Pandora column NO_2 abundances by estimates of PBL height derived from the WRF/CMAQ model system resulted in a consistently moderate degree of correlation (Table 2.15; Figure 2.10). Normalization by PBL height also resulted in statistically significant increases in the value of R^2 relative to the Pandora full data set correlation analyses at the 95% confidence level at most sites. Overall, a moderate degree was obtained for most sites

for after normalization of the Pandora columns, rather than the low degree obtained with the simple linear regression analysis. The P-3B column NO_2 normalization analyses presented more mixed results. Normalization of P-3B column NO_2 by estimates of PBL height derived from the observed potential temperature profile resulted in moderate correlation, when significant, for column_ground, and column_air (Table 2.15; Figure 2.9). However, normalization did not produce significantly different results relative to the P-3B full data set correlations. The lack of improvement for the P-3B normalization analyses may be due to the "well mixed PBL" assumption inherent in the gap-filling methods used for the column computations. Because one value is held constant within the lowermost PBL, the P-3B columns likely rely on a better mixed NO_2 profile than the Pandora columns, such that normalization by PBL depth does not add as much useful information for the P-3B as it did for Pandora NO_2.
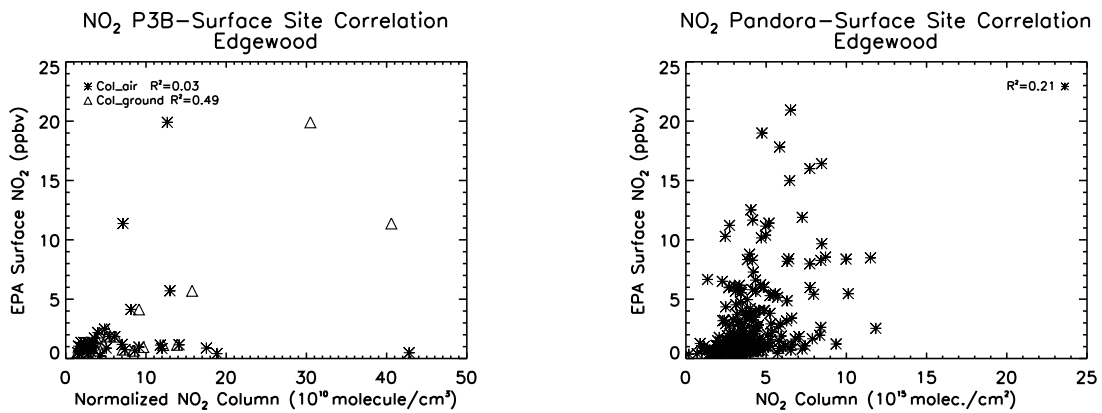
Figure 2.10: Example scatter plots of NO_2 column vs. surface NO_2 mixing ratio at Edgewood for the P-3B (left) and Pandora (right). Normalization by PBL height correlation analysis (surface vs. [column/PBLH] correlation). R^2 values displayed at the top of each plot.

	P-3B NO$_2$ Col_Air R^2	P-3B NO$_2$ Col_Ground R^2	Pandora NO$_2$ R^2
Aldino	0.32	0.62	0.31
Beltsville	0.33	0.51	0.67
Edgewood	NS	0.49	0.50
Essex	NS	0.46	0.57
Fair Hill	NS	0.27	0.28
Padonia	0.37	0.61	0.58

Table 2.15: Summary of the R^2 statistic for the P-3B and Pandora NO$_2$ PBL-normalization analysis. NS denotes non-significant correlation.

The correlations between Pandora NO$_2$ surface mixing ratio (ppb) and column abundance (cm^{-2}) at Edgewood and Padonia presented in Section 2.7.4 compared well to the results obtained by Knepp et al. (2013) for their comparison of hourly-averaged Pandora NO$_2$ surface (ppb) and column data (cm^{-2}). This agreement held after Knepp et al. (2013) excluded surface NO$_2$ mixing ratios less than 1 ppb, and for their comparison of raw surface and column NO$_2$ data (see Knepp et al., 2013, Table 2). The correlations between P-3B surface mixing ratio (ppb) and column_ground (cm^{-2}) at Edgewood and Padonia also compared modestly well to Knepp et al. (2013), though the P-3B correlations were larger than either Pandora analysis. The correlations between the surface mixing ratios and Pandora or P-3B column_ground NO$_2$ columns normalized by PBL height presented here also compared well to the correlations obtained by Knepp et al. (2013) after application of their PBL-correction factor at Edgewood and Padonia (see Knepp et al., 2013, Eqn. 1 and Table 3). Knepp et al. (2013) employed PBL heights from the EDAS40 model, OMI stratospheric NO$_2$ column data, and air density to transform Pandora NO$_2$ total columns into PBL-average mixing ratio values, and then examined the correlation between

between *in situ* and Pandora-estimated mixing ratios; improved correlation was also obtained after application of this correction factor for Pandora. Example scatter plots are presented in Figure 7. Differences in the correlations presented here and in Knepp et al. (2013) may be due to the exclusion of data occurring at a solar zenith angle greater than 75° by Knepp et al. (2013), and differences in the PBL height derived from the WRF/CMAQ model system, the model used by Knepp et al. (2013), and from the P-3B potential temperature profile. This agreement between the results presented here and in Knepp et al. (2013) for Pandora further demonstrates the influence of mixing within the PBL on the NO_2 column-surface relationship, as application of a PBL correction (either the normalization analysis presented here or the correction factor of Knepp et al.) resulted in improved column-surface correlations. The results presented in this section also bolster the conclusion found by Knepp et al. (2013) that, to a first order, NO_2 column abundances can be relevant to surface air quality.

2.8: Comparison of CMAQ Analyses to Observational Analyses

The degree of correlation between surface mixing ratio and column abundance found from the simple linear regression analyses for the Loughner et al. (2013) CMAQ model output is summarized in Table 2.1. Values of R^2 are given in Table 2.16, and representative scatter plots of the correlation are displayed in Figures 2.3-2.4. All regressions were statistically significant at a confidence level of 95%. Unlike the P-3B correlations, the CMAQ O_3 correlations (Figure A3) were not generally larger than the CMAQ NO_2 correlations (Figure A6). Significant differences in correlation between the CMAQ and P-3B analyses occurred between CMAQ and column_air NO_2 at most sites; CMAQ generally presented larger correlation than column_air. As discussed previously, the Aldino column_UMD analysis suggested that the P-3B column_ground correlations were likely too high. The CMAQ O_3 or NO_2 correlations were also statistically

significantly greater than those for Pandora O_3 or NO_2 at most sites. This indicates that O_3 and

NO_2 may be too well mixed vertically and horizontally within the model.

	Loughner et al. O_3 R^2	Loughner et al. NO_2 R^2	Loughner et al. O_3 F-ratio	Loughner et al. NO_2 F-ratio	NOAA O_3 R^2	NOAA NO_2 R^2	NOAA O_3 F-ratio	NOAA NO_2 F-ratio
Aldino	0.56	0.76	46.09 (<0.001)	46.09 (<0.001)	0.86	0.67	225.81 (<0.001)	74.01 (<0.001)
Beltsville	0.75	0.39	126.24 (<0.001)	26.58 (<0.001)	0.84	0.74	221.54 (<0.001)	117.05 (<0.001)
Edgewood	0.53	0.49	49.32 (<0.001)	42.81 (<0.001)	0.82	0.65	190.83 (<0.001)	79.50 (<0.001)
Essex	0.63	0.63	62.90 (<0.001)	63.43 (<0.001)	0.71	0.88	91.06 (<0.001)	270.38 (<0.001)
Fair Hill	0.54	0.93	48.68 (<0.001)	544.03 (<0.001)	0.83	0.88	205.91 (<0.001)	305.17 (<0.001)
Padonia	0.81	0.68	160.75 (<0.001)	78.30 (<0.001)	0.78	0.68	134.42 (<0.001)	78.13 (<0.001)

Table 2.16: Summary of the R^2 statistic and F-ratio (p-value) for the CMAQ O_3 and NO_2 simple linear regression analysis.

Values of R^2 for the correlation separation analyses are given in Appendix B. Comparing

the Loughner et al. (2013) CMAQ simulation to the P-3B for the correlation analyses separated

by time of day, the CMAQ NO_2 correlations were significantly larger than those for P-3B NO_2

column_air for the Afternoon group at four of the six MDE sites. However, CMAQ produced

significantly larger correlations relative to Pandora for the O_3 and NO_2 Afternoon analyses at all

sites (Tables B1-B2, B6). The larger CMAQ NO_2 correlations relative to P-3B column_air but

not column_ground for the Afternoon analysis suggests that these large CMAQ correlations

during afternoon may be related to the growth of the boundary layer during the day and that too

much horizontal and vertical mixing within the boundary layer is occurring in the model. For the

separation by PBL height analyses, CMAQ produced correlations significantly larger than those

for P-3B NO_2 column_ground for the High PBL group at three of six MDE sites, and larger than those for P-3B NO_2 column_air for both PBL data groups at all sites. CMAQ produced significantly larger correlations relative to the Pandora O_3 analyses for the Low and High PBL groups at all sites, and relative to Pandora NO_2 for both data groups at four of six sites (Tables B3-B4, B6). Because most significant differences occurred with either the Afternoon group or High PBL group, this suggests that mixing influences the column-surface relationship within CMAQ and that horizontal and vertical mixing may be too strong within the model. Furthermore, the correlation between CMAQ PBL height-normalized column NO_2 and surface NO_2 was significantly larger than the full data set correlations at only two sites; normalization by PBL height does not add as much information to the CMAQ correlations because NO_2 is too well mixed within the model (Table 2.17).

	CMAQ (Loughner et al.) NO_2 R^2	CMAQ (NOAA) NO_2 R^2
Aldino	0.88	0.79
Beltsville	0.85	0.67
Edgewood	0.63	0.50
Essex	0.77	0.88
Fair Hill	0.95	0.91
Padonia	0.90	0.82

Table 2.17: Summary of the R^2 statistic for the CMAQ NO_2 PBL-normalization analysis. NS denotes non-significant correlation.

A high degree of correlation was found between both O_3 and NO_2 surface and column output within the NOAA CMAQ forecast (Table 2.1); the correlations within this forecast were significantly larger than the P-3B O_3 and NO_2 and Loughner et al. (2013) simulation correlations

at several sites, and was significantly larger than all Pandora O_3 and NO_2 correlations (Table 2.17). The results for the correlation separation analyses for the NOAA CMAQ forecast are consistent with the results for the Loughner et al. (2013) simulation, though the impacts within the NOAA forecast were greater. For example, in addition to presenting significantly larger correlations relative to the P-3B NO_2 column_air analyses for both Low and High PBL groups, NOAA CMAQ NO_2 also produced larger correlations relative to the NO_2 column_ground analyses for the High PBL at four of six MDE sites (Tables B7-B8). Additionally, no correlations between PBL-normalized NO_2 column and surface mixing ratios were significantly different than the full data set correlations for the NOAA simulation (Table 2.17). These results again indicate that vertical and horizontal mixing within the model may be too strong, and that inaccuracies within model mixing schemes can have an important impact on the column-surface relationship for O_3 and NO_2 within CMAQ.

2.9: Conclusions

A wide range of degrees of correlation resulted from the simple linear regression analyses between the O_3 and NO_2 column and surface data. The OMI tropospheric O_3 and NO_2 data resulted in non-significant correlations, the P-3B column_air NO_2 and Pandora O_3 demonstrated a low degree of correlation, P-3B column_ground NO_2, CMAQ O_3 and NO_2, and Pandora NO_2 demonstrated a moderate degree of correlation, and P-3B column_air and column_ground O_3 demonstrated a high degree of correlation with surface air quality observations. These results indicate that O_3 is generally well mixed in the vertical, while NO_2 is not. Further, a simple linear regression model was found to fit the P-3B O_3 column and surface data well, while it struggled to capture the column versus surface relationships for the P-3B NO_2, Pandora O_3, and Pandora NO_2 data. The multivariate regression analyses and the PBL normalization correlation analyses

indicate that PBL height (an indicator of mixing) add meaningful information to the column-surface relationship.

The O_3 correlations within the Loughner et al. (2013) simulation and NOAA CMAQ forecast were similar to the P-3B O_3 correlations, but were more similar to column_ground than column_air for NO_2. Both sets of CMAQ output demonstrated greater correlation between the O_3 and NO_2 column and surface during the afternoon and for conditions associated with a maturely developed PBL than did the observations. These results suggest that vertical and horizontal mixing within the model is stronger than in the observational data sets. In future work, we will investigate how the vertical mixing in CMAQ can be improved.

The large OMI footprint likely contributes to the non-significant correlations obtained between OMI tropospheric O_3 or NO_2 column and surface observations; the insufficient sensitivity of the OMI instrument to the lower troposphere also contributes for the OMI O_3 analyses. The DISCOVER-AQ measurements suggest that O_3 observations from future satellite instruments can be meaningful for surface air quality analysis if they have sufficient sensitivity to the lowest 2-3 km of the troposphere.

Chapter 3: Variability of O_3 and NO_2 Profile Shapes during the DISCOVER-AQ Project: Implications for Satellite Observations and Comparisons to Modeled Profiles

3.1: Introduction

Satellite observations have great potential for diagnosis of near-surface air quality conditions because of their global coverage, increasingly high spatial resolution, fixed temporal resolution, and improved retrievals of tropospheric column abundances (Beirle et al., 2003; Boersma et al., 2008; Chatfield and Esswein, 2012; Fishman et al., 2008; Flynn et al., 2014; Lamsal et al., 2011; Martin, 2008). Satellite column observations can be especially useful for diagnosis of near-surface abundances of the Environmental Protection Agency (EPA) criteria pollutants (http://www.epa.gov/air/criteria.html) ozone (O_3) and nitrogen dioxide (NO_2). Recent work has demonstrated that such an application should be possible for NO_2. Lamsal et al. (2008) and Lamsal et al. (2010) observed significant correlation between *in situ* surface NO_2 mixing ratios and ground-level NO_2 observations inferred from Ozone Monitoring Instrument (OMI) column amounts, after application of local scaling factors derived from the GEOS-Chem model. Other works have instead scaled surface NO_2 values to obtain column amounts with the use of assumed NO_2 profiles, and have found good agreement between OMI tropospheric columns and these scaled columns (Boersma et al., 2009; Knepp et al., 2013; Ordóñez et al., 2006). Chatfield and Esswein (2012) examined ozonesonde data over the U.S. and observed a significant correlation between partial column (0-3 km) and near-surface O_3 (500 m) observations. Flynn et al. (2014) examined partial or full tropospheric column amounts derived from aircraft or Pandora UV/Vis spectrometer data sets for O_3 and NO_2 during July 2011 in the Baltimore-Washington metropolitan region, and found a wide range of degrees of correlation between column and surface data, with O_3 generally demonstrating a greater correlation than NO_2. These results suggest that satellite-observed ozone observations could be used to estimate surface ozone,

provided there is sufficient sensitivity to the lowermost troposphere. Natraj et al. (2011) demonstrated that lower tropospheric O_3 retrievals could be greatly improved by flying both UV and thermal IR sensors together on future satellites.

These studies are encouraging that column amounts and surface concentrations can be related. However, such work also highlighted the considerable difficulties inherent in understanding this relationship. Biases remain in satellite retrievals, while current satellite instruments have reduced sensitivity to the lower troposphere (Liu et al., 2010; Martin, 2008). The wide range of correlations obtained and the use of model- or data-derived scaling factors demonstrate a need for better understanding of the processes connecting the column and surface.

Understanding of the variability of *in situ* profile shapes is useful for understanding the degree of correlation between column and surface data. Satellite data may be more useful for air quality applications in some parts of the day than others and under certain meteorological conditions. How well do the assumed profile shapes used in satellite retrievals capture observed conditions and ultimately what is the resulting impact on the ability of satellite-observed columns to represent surface air quality? Additionally, the assumed profile shapes used in retrievals are given not as profiles of volume mixing ratios, but as shape factors that are provided by simulations of global chemical transport models such as GMI. The shape factor is defined as the ratio of the partial column within a vertical layer z (Ω_z) to the full tropospheric column (Ω_{trop}), and thus depends indirectly on the mixing ratio profile:

$$S(z) = \Omega_z/\Omega_{trop} \tag{3.1}$$

For the NASA standard NO_2 product (Bucsela et al., 2013), shape factors are used in the radiative transfer model to calculate the air mass factor (AMF), used to convert the slant column abundance to the vertical column abundance according to

$$AMF = \Omega_s/\Omega_v \tag{3.2}$$

where Ω_s is the slant column and Ω_v is the vertical column (Chance, 2002; Lamsal et al., 2014; Palmer et al., 2001). The AMF is also used with the differential optical absorption spectroscopy (DOAS) technique, such as used to retrieve the OMI total vertical O_3 columns (Bhartia, 2002). Liu et al. (2010) employ a similar method in their retrieval of OMI O_3 vertical columns, in which the AMF is not used, but rather *a priori* partial O_3 column amounts at each model vertical layer in the retrieval computations. Accurate representation of the O_3 or NO_2 profile shapes is critical to accurate representation of the shape factors used in satellite retrievals, and thus retrieval accuracy. A high-resolution NO_2 retrieval from OMI has been performed by Russell et al. (2011) using NO_2 profiles from the WRF-Chem model at 4-km horizontal resolution. This work evaluates the ability of a regional air quality model to produce accurate NO_2 profiles for use in satellite retrievals.

Previous studies have examined the observed trace gas or aerosol profile variability over several regions. Diab et al. (2003) investigated O_3 aircraft profiles over Johannesburg, South Africa, through application of a clustering technique. Six unique O_3 clusters were found, which were further related to air mass origin through back trajectory modeling. Diab et al. (2004) clustered ozonesonde profiles over Irene, South Africa, for the periods 1990-1994 and 1998-2002; these clusters were also related to air mass origin and meteorological conditions. Using a self-organizing maps technique, Stauffer et al. (2016) obtained nine distinct clusters of tropospheric O_3 profiles from four long-term U.S. ozonesonde locations, which corresponded to distinct meteorological and pollution conditions. Additionally, this work determined that O_3 profile climatologies greatly underrepresented O_3 profiles at these sites. Taubman et al. (2006) and Hains et al. (2008) examined summertime trace gas and aerosol aircraft vertical profiles over

the mid-Atlantic U.S. during 1997-2003. Through a cluster analysis of back trajectories or pollutant profiles, distinct pollution regimes and their associated meteorological conditions and emissions were identified for the summertime Mid-Atlantic region. However, while these studies examined long time record data sets, they were limited in spatial extent and occasionally also by time of year examined. These studies also did not investigate NO_2 profile variability, nor evaluate the ability of air quality models to replicate these profile shape clusters.

In support of DISCOVER-AQ (described previously in Chapters 1 and 2), results are presented of an agglomerative hierarchical cluster analysis of P-3B *in situ* profiles for O_3 and NO_2 for each of the four campaigns. Through these analyses, the variability of the *in situ* P-3B O_3 or NO_2 profiles will be characterized for each campaign. Classes of profiles are identified for each trace gas and each campaign, and meteorological conditions influencing these classes will be investigated as well as the associated column vs. surface correlations for each cluster. Shape factors are computed from the O_3 and NO_2 observations. The observed cluster shape factor results are also compared to shape factors from the Community Multiscale Air Quality (CMAQ) model and NASA Global Modeling Initiative (GMI) model to assess model performance and the relevance of the model profiles for use in satellite retrievals will be evaluated.

3.2: Description of P-3B *In Situ* Profile Measurements

A complete description of the DISCOVER-AQ measurements is publicly available at http://www-air.larc.nasa.gov/missions/discover-aq/discover-aq.html. *In situ* trace gas volume mixing ratio data were collected by the P-3B aircraft over a network of six surface air quality monitoring sites during the Maryland, California, and Colorado campaigns, and over a network of eight monitoring sites during the Texas campaign. Typically, three spirals were accomplished over each spiral site during each flight day (typically morning, midday, and afternoon), with 10-

15 flight days per campaign. Spiral sites for each campaign are mapped in Fig. 3.1a)-d). The National Center for Atmospheric Research (NCAR) $NO_{xy}O_3$ instrument, a 4-channel chemiluminescence instrument for the measurement of NO, NO_2, NO_y, and O_3, provided the P-3B O_3 (uncertainty of 5%) and NO_2 (uncertainty of 10%) *in situ* observations used here. P-3B carbon monoxide (CO) *in situ* observations were provided by the Differential Absorption Carbon Monoxide instrument (DACOM; uncertainty of 2%). The National Suborbital Education and Research Center (NSERC) P-3B data acquisition and distribution system provided the *in situ* observations of altitude and meteorological observations used in these analyses.
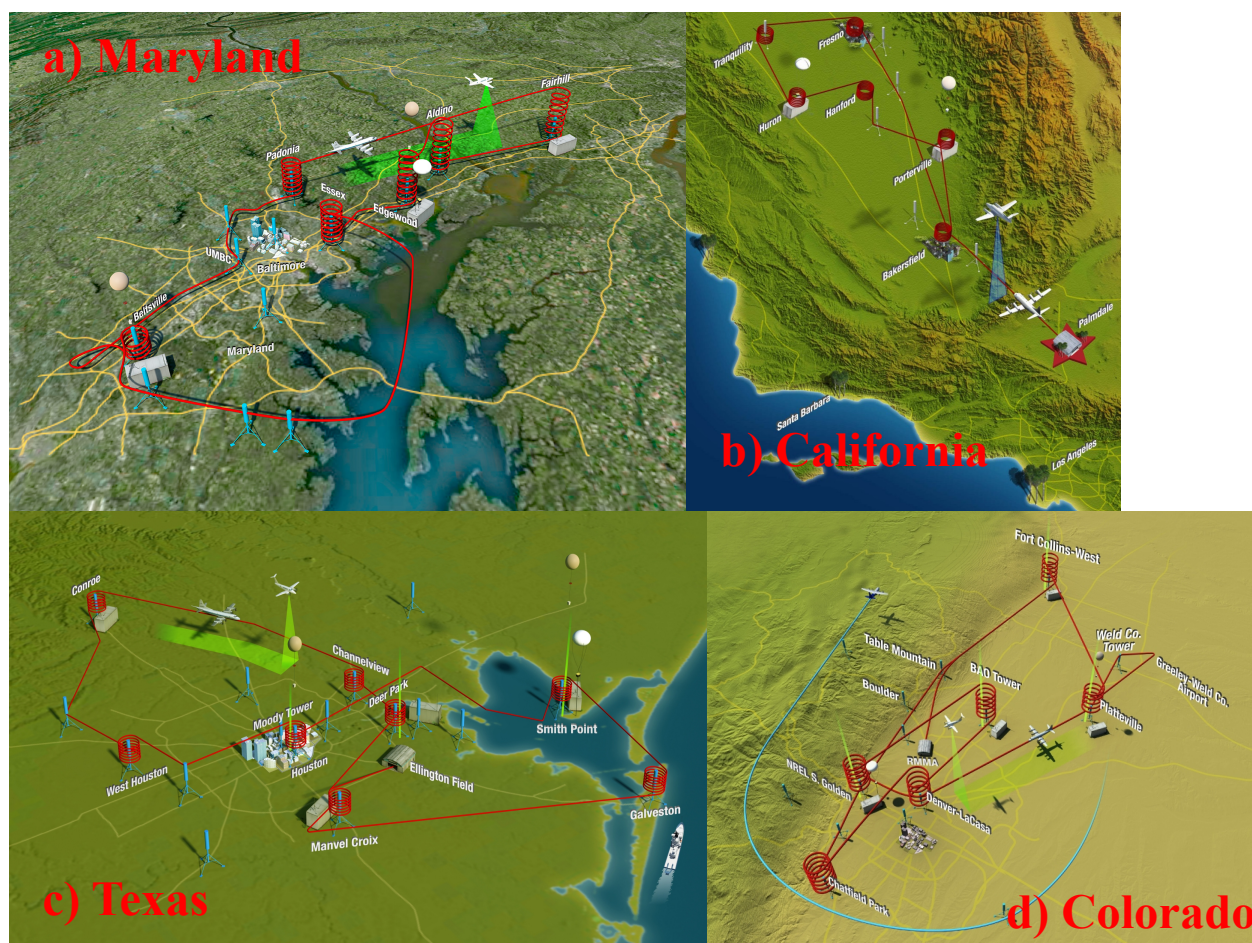
Fig. 3.1: a) The six spiral sites for the Maryland campaign; b) the six spiral sites for the California campaign; c) the eight spiral sites for the Texas campaign; and d) the six spiral sites for the Colorado campaign. Spiral sites named in white font.

3.3: Description of P-3B Column_Air and Column_Ground Lower Tropospheric Columns

The P-3B column_air and column_ground O_3 and NO_2 values were computed as described previously in Chapter 2, Section 2.2.1, for all four DISCOVER-AQ campaigns. The well mixed assumption inherent in the column_ground computation is worth noting again. For the California, Texas, and Colorado campaigns, the aircraft profiles also included missed approach data from local airports located near some spiral sites, which reached as low as ~80 m AMSL, rather than the typical ~300 m AMSL.

3.4: Description of Surface Volume Mixing Ratio Datasets

The surface volume mixing ratio data sets available for the Maryland campaign have been described in Chapter 2, Section 2.3. Uncertainties for the following surface volume mixing ratio datasets are the same as those described for the Maryland surface data in Chapter 2.

Several different surface mixing ratio data sets were available for the California campaign. Surface O_3 data were provided by the San Joaquin Valley Air Pollution Control District (SJV) at all California spiral sites except Huron. The EPA provided surface photolytic converter NO_2 measurements at Bakersfield and Porterville, The Pennsylvania State University Nittany Atmospheric Trailer and Integrated Validation Experiment (NATIVE; http://ozone.met.psu.edu/Native/) provided O_3 and NO_y data at the Porterville site, and The Millersville University of Pennsylvania (MU) provided O_3 and molybdenum-converter NO_x data at the Huron site. SJV surface NO_y data were also available at the Hanford site. The California Air Resources Board (CARB) provided surface NO_2 at the Fresno site. Photolytic surface NO_2 data were used instead of molybdenum converter NO_2 (which has interferences from NO_y species) data for the correlation analyses that follow, at those sites for which photolytic data were available (Bakersfield and Porterville).

The Texas Commission on Environmental Quality (TCEQ) provided surface O_3 data from a chemiluminescence monitor at the Conroe, Channelview, Deer Park, Galveston, and Manvel Croix spiral sites. TCEQ also provided surface molybdenum-converter NO_2 data at the Conroe, Channelview, and Deer Park sites, and NO_y data at the Galveston and Manvel Croix sites. The University of Houston (UH) provided O_3 and NO_y surface data at the Moody Tower site; the surface at this site was approximately 70 m AGL, as these monitors sat on the rooftop of the Moody Tower on the UH campus. NATIVE provided O_3 and NO_y at Smith Point. Almost no surface data were available for the West Houston site. Lastly, NOAA provided photolytic NO_2 data from a chemiluminescence monitor equipped with a photolytic converter at Galveston, and provided NO_2 surface measurements at the Manvel Croix from a cavity ring down instrument. Photolytic or cavity ring down surface NO_2 data were used in place of NO_y measurements, where available. As with California, photolytic surface NO_2 data were used in place of NO_y measurements, where available.

The Colorado Department of Health and Environment (CDPHE) provided O_3 surface mixing ratio data from chemiluminescence monitors at the Golden, Chatfield Park, Fort Collins, and La Casa spiral sites and molybdenum converter chemiluminescence NO_2 from the La Casa site. NATIVE provided surface O_3 and NO_y data at the Platteville site, while the NASA Langley Research Center Langley Aerosol Research Group Experiment (LARGE) provided O_3 and NO_2 data at the Boulder Atmospheric Observatory (BAO) site from their LARGE mobile van suite of instruments. EPA provided photolytic NO_2 measurements at the Golden and Fort Collins sites. Neither NO_2 nor NO_y data were available for the Chatfield Park site.

3.5: Description of Model Simulations

3.5.1 HYSPLIT Simulations

Back trajectories were computed for each P-3B spiral for each of the four campaigns with the National Oceanic and Atmospheric Administration (NOAA) Air Resources Laboratory (ARL) Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model version 4 (Draxler and Rolph, 2003), to analyze air mass source regions and transport. Meteorological inputs were taken from the North American Mesoscale (NAM) model 40 km Eta Data Assimilation System (EDAS) 3-hour data archive. Back trajectories were computed for each spiral's center latitude and longitude back to three days prior to each spiral, and were computed for each 500 m increment in altitude between 500 m and 3500 m AGL. To compensate for these errors in this analysis, back trajectories associated with a spiral site were clustered using the clustering algorithm within the HYSPLIT model. The HYSPLIT clustering algorithm is based on the k-means clustering approach (a different clustering technique from agglomerative hierarchical clustering). This cluster analysis was performed over the entirety of the campaign to which that spiral site belonged, to determine the source regions and transport patterns that most contributed to air mass transport at the site during the campaign period.

3.5.2: Loughner et al. CMAQ Simulations

The CMAQ model version 5.0 was used to simulate air quality for the Maryland campaign, as described in Chapter 2, Section 2.5.1.

CMAQ version 5.0.2 was used to simulate air quality for the Texas campaign, again driven off-line by WRF meteorology. In this case, WRF used the North American Model (NAM) 12-km analyses for initial and boundary conditions. The same WRF and CMAQ options used for the Maryland campaign were used in this simulation, with the exception of 45 vertical layers instead of 34 layers, with 18 layers within the lowermost 2 km. However, 2012 Texas Commission on Environmental Quality (TCEQ) anthropogenic emissions were used instead of

projected emissions based on the NEI.; as with the Maryland simulation, BEIS and MOVES were used to compute biogenic emissions and anthropogenic mobile emissions, respectively. The Texas simulations were also run iteratively, such that only the second, improved air quality simulation was used in the analyses presented below. This technique required WRF to be run twice: the first WRF run performed analysis nudging on all domains based on the 12 km North American Model (NAM) output, and the second WRF run performed analysis nudging on all domains based on the NAM with the exceptions of 2 m temperature and humidity for the 4 km horizontal resolution domain; the 2 m temperature and humidity from the first WRF simulation at 12 km was used to nudge these two parameters for the second WRF simulation at the 4 km resolution. The second iterative WRF run was used to drive CMAQ. CMAQ output was provided in 20-minute averages for the Texas campaign, while it was given in hourly averages for the Maryland campaign. The 4 km output was used for the following Texas analyses to capture bay and sea breeze events (Christopher P. Loughner, personal communication).

3.5.3: NOAA ARL CMAQ Forecasts

NOAA ARL provided forecasts of O_3 and NO_2 from an experimental version of CMAQ Version 4.6 during each deployment; these simulations were examined for campaigns for which Loughner et al. simulations were unavailable. The CB05 chemical mechanism was also used in the NOAA simulation. However, the NOAA model runs were driven offline by WRF (Nonhydrostatic Mesoscale Model, or NMM, core) meteorology, and used the fourth generation aerosol module (aero4), the Mellor-Yamada- Janjić (MYJ; Janjić, 1994) scheme for boundary layer mixing, the Noah land surface model, and the 2005 NEI for anthropogenic emissions; lightning NO_x emissions were not included. The available horizontal resolution was 12 km for the California campaign and 4 km for the Colorado campaign, and the available vertical resolution

was 22 layers for California and 27 layers for Colorado, with 13 layers and 17 layers, respectively, within the lowest 2 km.

3.5.4: NASA GMI Simulations

Profiles of O_3 and NO_2 were obtained from the NASA GMI coupled troposphere-stratosphere chemical transport model. GMI has a horizontal resolution of 2° latitude by 2.5° longitude, with 72 vertical levels (Duncan et al., 2007). Specifically, the GMI HindcastFF simulations (Strode et al., 2015) were used in the following analyses. These simulations are driven by meteorology from the Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al., 2011). Emissions inputs include year-specific fossil fuel emissions based on the Emission Database for Global Atmospheric Research (EDGAR) 2000 emissions inventory, regional anthropogenic emissions inventories for other years, year-specific Global Fire Emissions Database v3 (GFEDv3; van der Werf et al., 2010) biomass burning emissions, Asian fossil fuel emissions from the 2006 Intercontinental Chemical Transport Experiment (INTEX-B; Zhang et al., 2009) experiment scaled to other years, and biofuels from the EPA/NEI99 over the U.S. Lightning NO_x emissions are also included (Allen et al., 2010).

3.6: Agglomerative Hierarchical Cluster Analysis

An agglomerative hierarchical cluster analysis was applied to the P-3B O_3 and NO_2 profiles for each campaign, following the approach of Hains et al., 2008. Any type of statistical cluster analysis seeks to group together objects with the smallest differences, such that objects (in this case, the difference D_{ij} between profiles, as defined below) within one cluster are more similar to each other than to objects within different clusters. Agglomerative hierarchical cluster analysis initially treats each object as its own cluster, and would continue to cluster until all

objects are grouped into a single cluster. A combination of a statistical criterion and manual inspection is required to determine a meaningful number of clusters.

As in Hains et al. (2008), the O_3 or NO_2 mixing ratio data for an individual profile were first averaged into altitude layers of 100 m, then grouped into altitude bins covering larger depths. Only the profiles that covered the full altitude range for each campaign were included in this analysis; it should be noted that all altitude data are above mean sea level (AMSL). Hains et al. (2008) employed the following equations to quantify the difference D_{ij} between profiles i and j within each profile pair, which are also employed in this work:

$$D_{ij}(1) = \left[\sum_{a=1}^{a=n} abs(C_i(a) - C_j(a))\right]^2 \tag{3.3}$$
$$D_{ij}(2) = 1 + [1 - r] \tag{3.4}$$
$$D_{ij}(3) = 1 - \exp\left[-(s-1)^2\right] \tag{3.5}$$
$$D_{ij} = \sum_{k=1}^{k=b} [D_{ij}(1) * (D_{ij}(2) + D_{ij}(3))] \tag{3.6}$$

where k is the index for the b altitude bins, a is the index for the n altitude layers within a bin, C is the mixing ratio for the *ith* and *jth* profiles, r is the correlation coefficient for each pair of profiles within each of the k bins, and s is the slope between each profile pair within a bin. The correlation coefficient and slope were obtained from a regression analysis between each profile pair within each bin. These differences D_{ij} represent the total difference between magnitude of the mixing ratios throughout the altitude range, as well as how much the slope and correlation coefficients deviate from unity, thus accounting for how different the profile shapes are at different altitude levels (Hains et al., 2008). The objects of the cluster analysis presented here are these D_{ij} values, and were clustered with a hierarchical clustering algorithm in the Interactive Data Language (IDL). An average linkage method was used to determine the similarity between clusters, in which the distance between two clusters was defined as the average difference between the data points in the first cluster and the data points in the second cluster.

The optimal number of meaningful clusters was determined with a combination of manual inspection and a technique based on total root mean square deviation (TRMSD), as described by Taubman et al. (2006). The dendrogram (tree diagram displaying the arrangement of cl7usters) produced by the clustering algorithm was initially inspected to determine a reasonable maximum number of clusters to consider, before application of the TRMSD technique. In this technique, an average profile was calculated for each cluster, and then the root mean square deviation (RMSD) of each profile within the cluster from the average cluster profile was computed. These RMSD values were then summed over all clusters under consideration to give the TRMSD. This TRMSD value was computed first over all clusters initially under consideration based on inspection of the dendrogram, and then computed again each time the number of clusters was reduced. The percentage change in TRMSD as clusters were condensed was then computed, and the optimal number of clusters taken as the number of clusters immediately before a large increase (~10%) in TRMSD, with the percentage change remaining relatively high upon further agglomeration (Taubman et al., 2006). The member profiles of each cluster obtained with the TRMSD technique were also manually inspected to assess the meaningfulness of the algorithm results.

The above techniques determine the optimum number of meaningful clusters, but do not necessarily ensure that each median cluster profile obtained after application of these techniques is significantly different from the other median profiles. In other words, each median cluster profile shape and relative magnitude at each vertical level may differ from those of the other median clusters, but the distribution of magnitudes at each vertical level may overlap, for which the clustering algorithm and TRMSD computation do not test. Thus, to determine that a cluster is significantly different from the other clusters for a campaign, the cluster median profile must have

exhibited at least five consecutive altitude layers in which the 25th and 75th percentile values (error bars) did not overlap with those of any other median profile.

3.7: Shape Factor Computation

Model simulated profiles were evaluated relative to the P-3B *in situ* profiles in terms of the O_3 or NO_2 shape factor, or column relative vertical distribution, for each P-3B cluster. Simulated profiles most coincident in time and space to the P-3B profiles included within a cluster were sampled, and a median simulated shape factor profile was computed for each cluster from both the model and the observations. The shape factors were computed according to Eqn. 3.1. For both observations and model simulations, the partial columns were computed over the depth of each CMAQ or GMI vertical layer; thus, the vertical distribution was computed over the model vertical grid, analogously to the shape factors computed from global models used in conjunction with satellite retrieval algorithms (Lamsal et al., 2014; Palmer, et al., 2001). It should be noted that the CMAQ or GMI tropospheric columns used in this analysis were also computed over the P-3B spiral depths only, and so were lower tropospheric column abundances rather than full tropospheric columns, to allow for a direct comparison to the P-3B shape factors.

3.8: Ozone P-3B *In Situ* Profile Clusters and Comparison to Models

3.8.1: P-3B Ozone Profile Clusters

Figure 3.2a)-d) shows the median profiles of O_3 for each of the clusters obtained for each of the four DISCOVER-AQ deployments. Clusters containing fewer than four members had too few members for a meaningful analysis and are excluded here. To determine that a cluster is significantly different from the other clusters for a campaign, the cluster median profile must have exhibited at least five consecutive altitude layers in which the error bars did not overlap with those of any other median profile. With few exceptions, the median profiles were

significantly different only within the planetary boundary layer (PBL), and lost distinction from each other within the free troposphere. Based on this significance criterion, the Texas campaign (Fig. 3.2c) demonstrated the greatest number of distinct clusters (five), as well as the greatest range of mixing ratio values, of all four campaigns. Few clusters were obtained for the California and Colorado campaigns (Fig. 3.2b) and Fig. 3.2d), respectively), though these clusters were distinct. The Maryland campaign (Fig. 3.2a), however, demonstrated only one significantly independent cluster (Cluster 4), unlike the other campaigns. This is contrary to initial expectations, and is surprising, given that the largest range in O_3 mixing ratio magnitudes within the PBL was encountered during the Maryland campaign. However, the clustering algorithm considered both mixing ratio magnitude and profile shape, producing the results described in this section for Maryland.
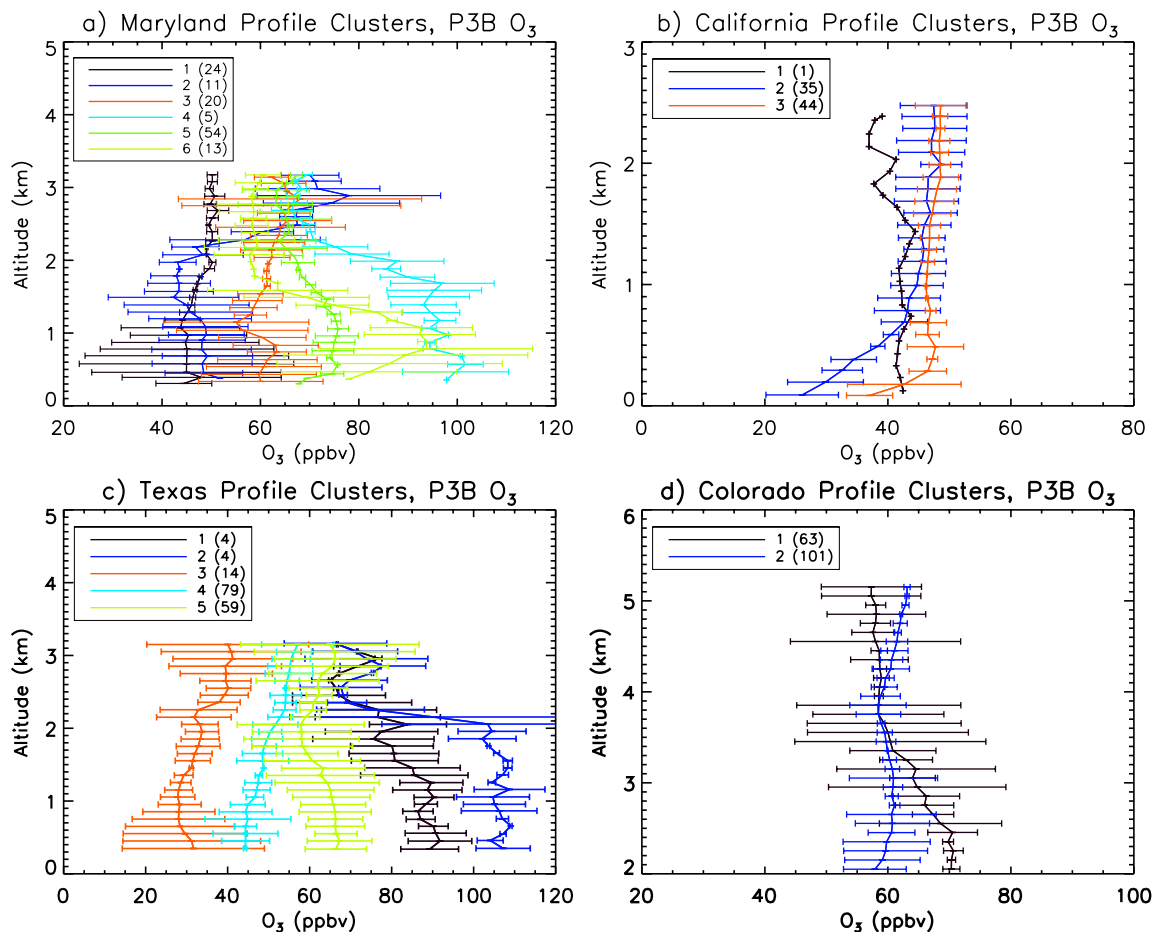
Fig. 3.2: The median profiles for each *in situ* O$_3$ profile cluster a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. Cluster numbers displayed in legend, with number of profiles in each cluster given in parentheses. Error bars represent 25th and 75th percentile values.

These differences in profile variability, as indicated by the number of significant clusters, may be due to synoptic conditions favoring or inhibiting O$_3$ formation. For example, the California campaign took place during winter, and thus experienced less sunlight and the coldest temperatures relative to the other campaigns, which inhibited O$_3$ formation. Though the Colorado campaign took place during summer, cooler temperatures were also experienced here relative to the other warm season campaigns due to its elevation in addition to more convection than initially expected, which also led to some inhibition of O$_3$ formation. Just as importantly, the effects of complex terrain may have led to greater horizontal mixing within the study regions

during the California and Colorado campaigns, further limiting the variability of the observed O_3 profiles. Plots of the wind fields from the NARR displayed complex interactions of the synoptic scale flow with local topography within the San Joaquin Valley during the California campaign and within the foothills and plains at the base of the Front Range during the Colorado campaign at several pressure levels. For example, along-valley flow or valley circulations were often seen due to upslope or downslope flow on opposing sides of the San Joaquin Valley, leading to horizontal mixing within the valley (Fig. 3.3). Interactions of downslope flow from the Front Range often interacted with northerly or southerly winds over the plains, leading to horizontal circulation patterns; these wind flow patterns also changed throughout the course of a day (Fig. 3.4). On some days, a well-defined upslope flow formed on the east side of the Front Range during the daytime, which also influenced pollution profiles. In contrast, the Maryland campaign study region was embedded within the synoptic scale flow, which tended to be westerly or northwesterly, at most levels, and, due to this, polluted air masses from the Ohio River Valley or Great Lakes Region were often transported into the study region (Fig. 3.5). Ozone was thus likely the most well mixed vertically and horizontally during the Maryland campaign, dampening its profile variability. Wind patterns had less of an impact during the Texas campaign than for the other three, suggesting that local chemistry and emissions were much more important to O_3 profile variability than local meteorology (except for the last two flights). This is perhaps due to the more complex chemical environment of the Houston metropolitan area, with large, nearby sources of volatile organic compounds (VOCs).

NARR 750mb Geopotential (dm) 20130131, 10am PST
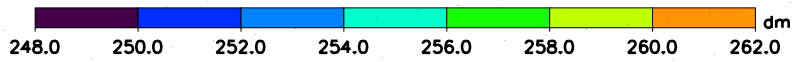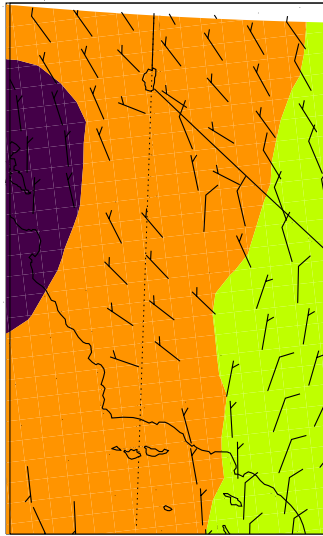
NARR 850mb Geopotential (dm) 20130131, 10am PST
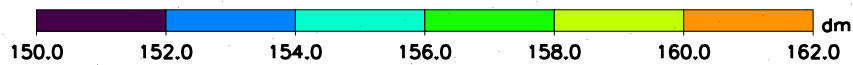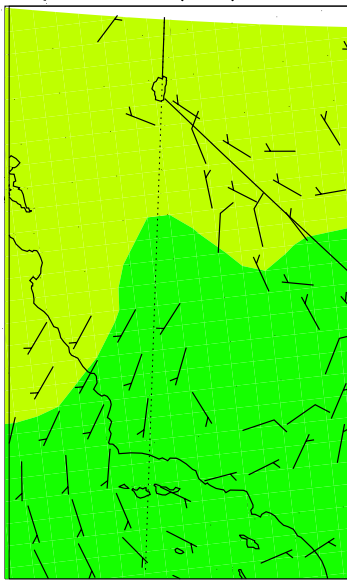
Fig. 3.3: NARR wind fields and geopotential height at the 750 mb and 850 mb at 10am PST on January 31, 2013, displaying typical circulation patterns within the San Joaquin Valley, California.
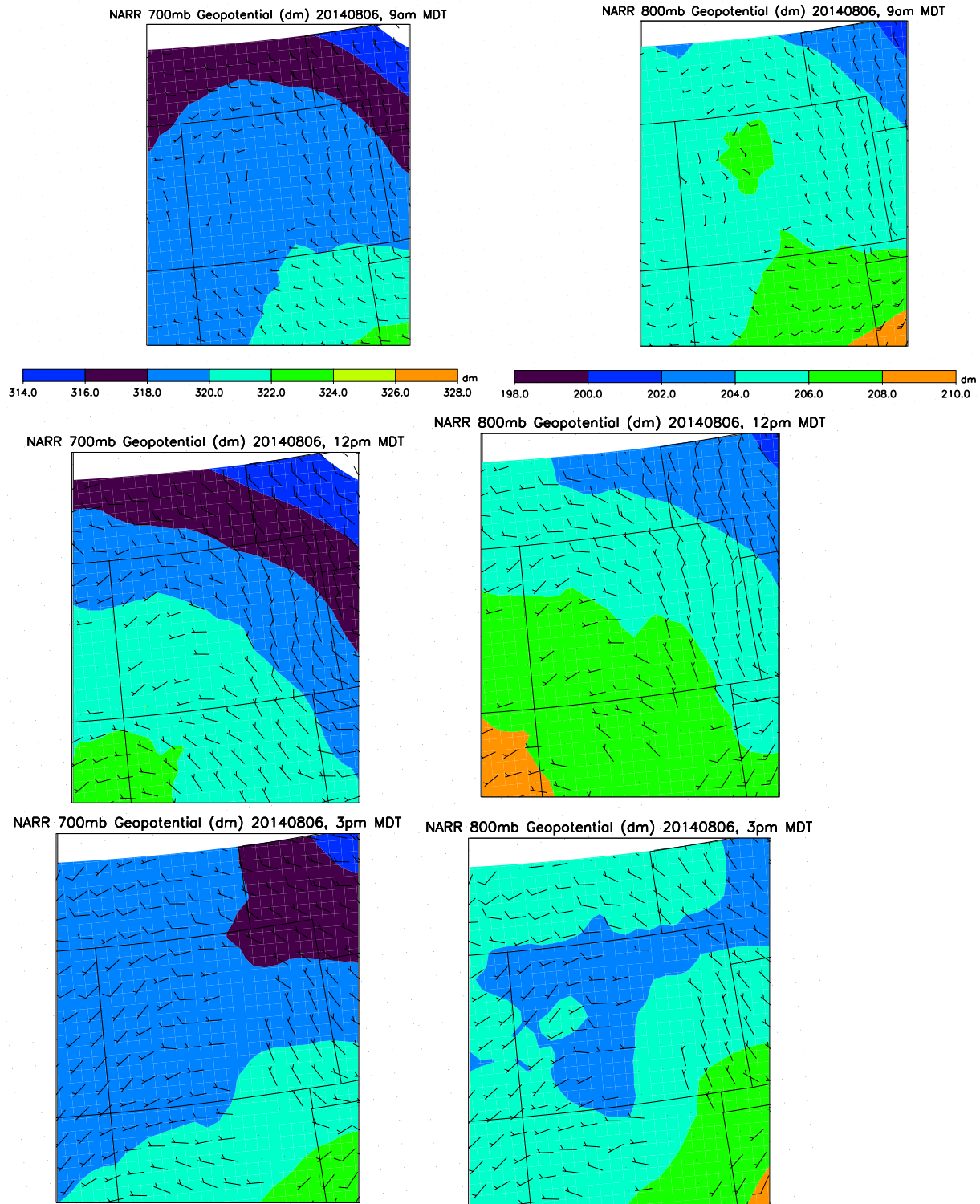
Fig. 3.4: NARR wind fields and geopotential height at the 700 mb and 800 mb at 9am, 12pm, and 3pm MDT on August 6, 2014, displaying typical circulation patterns and changes in wind flow over Colorado.
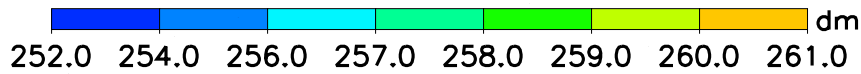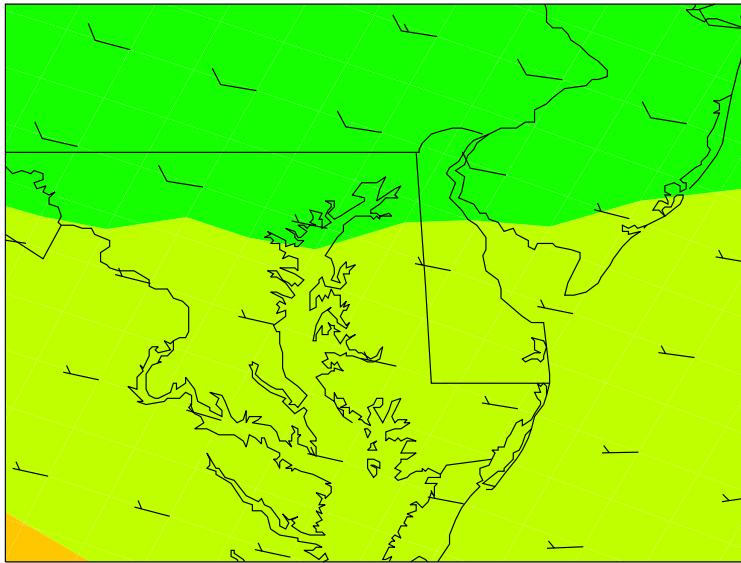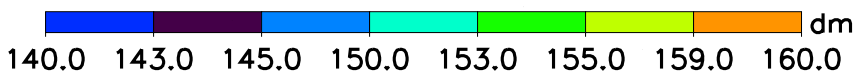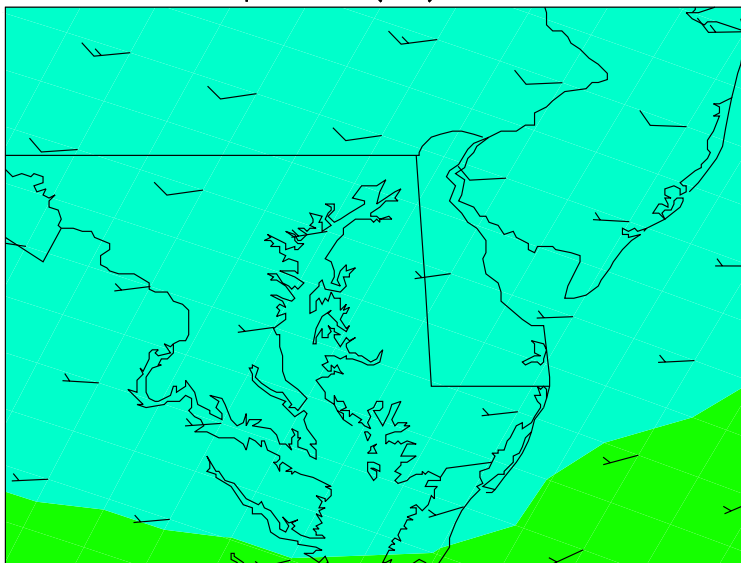
Fig. 3.5: NARR wind fields and geopotential height at the 750 mb and 850 mb at 11am EDT on July 21, 2011, displaying typical westerly flow patterns over Maryland.

Further, differences in profile shapes among clusters for a campaign may be partially

explained by differences in the atmospheric stability encountered during the P-3B spiral sampling

times. Median profiles of potential temperature (θ), an indicator of stability and degree of mixing, for each O_3 profile cluster are displayed in Figure 3.6a)-d). These potential temperature profiles were not themselves clustered, but θ profiles coincident to each O_3 profile included within a cluster were sampled, and the median θ profile computed for that O_3 cluster. Comparison of the median θ profile to the median O_3 profile for each cluster suggests that θ has an influence on these profile shapes for all campaigns; where the θ profile is well mixed within the PBL, the O_3 profile is also relatively well mixed, and vice versa. For example, during the Colorado campaign, Cluster 2 demonstrated a more well mixed θ median profile than Cluster 1, corresponding to a well mixed O_3 median profile (Fig. 3.6d); during the California campaign, Cluster 1 displayed the most well mixed θ profile, also corresponding to a well mixed O_3 profile (Fig. 3.6b). However, the influence of potential temperature was somewhat weaker during the Maryland and Texas campaigns; Clusters 3 and 4 exhibited poorly mixed θ profiles (very stable) but relatively well mixed O_3 profiles during the Texas campaign, for example (Fig. 3.6c).
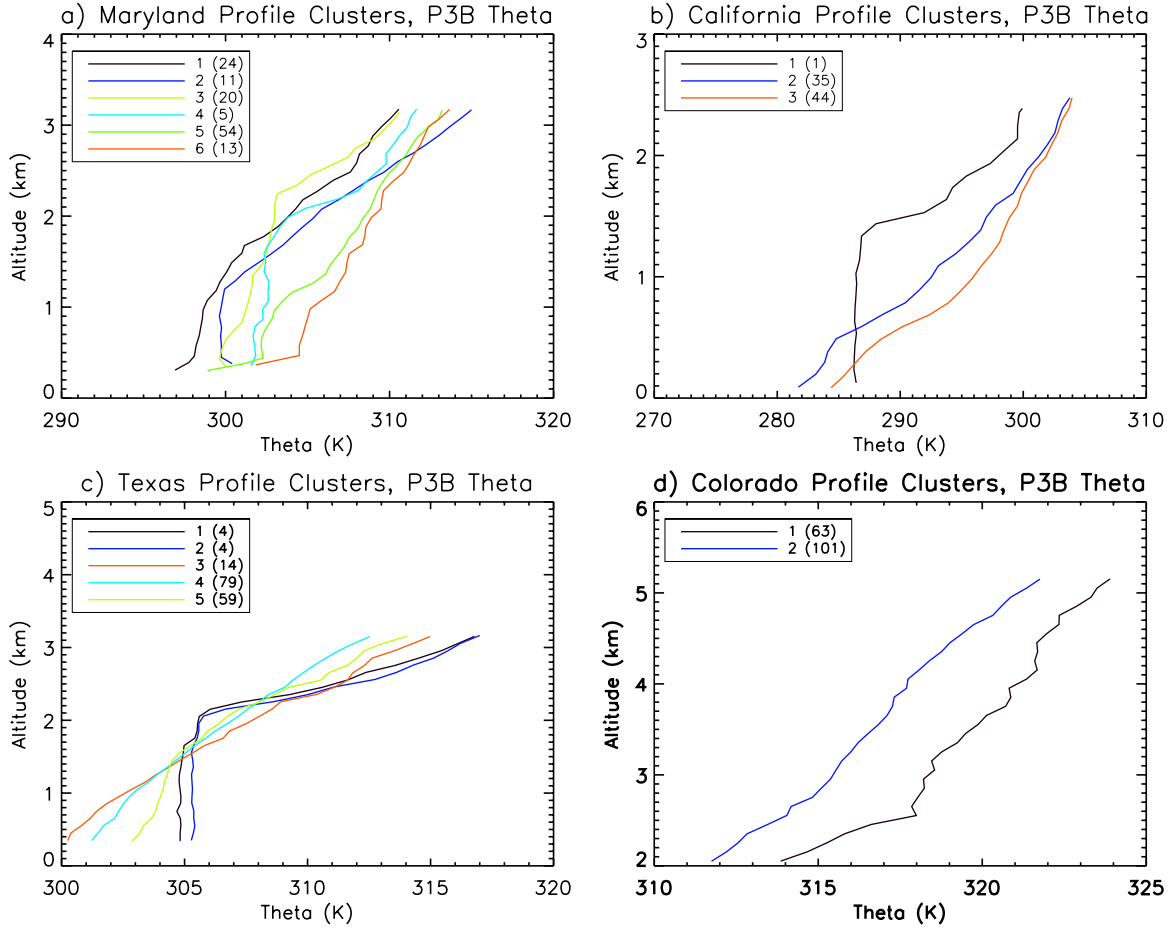
Fig. 3.6: The median potential temperature profiles associated with each *in situ* O$_3$ profile cluster a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. Cluster numbers displayed in legend, with number of profiles in each cluster given in parentheses.

Another indicator of stability, the temperature lapse rate (Γ), defined as

$$\Gamma = {}^{-dT}/_{dz} \qquad\qquad\qquad\text{Eqn. (3.7)}$$

also emerged as an influence on profile shape for all four campaigns. The lapse rate determines the static stability of the atmosphere, and is a local property (i.e., stability is not constant for all portions of the atmosphere). Thus, the lapse rate of an air parcel determines its buoyancy, and thus whether vertical displacement of that parcel is supported or inhibited by the surrounding environment. Three stability regimes are possible: if the parcel's lapse rate is less than the moist adiabatic lapse rate (4 K/km), then the local atmosphere is absolutely stably stratified and

vertical motion is suppressed; if the lapse rate falls between the moist and dry adiabatic (9.8 K/km) lapse rates, the local atmosphere is conditionally unstable, and vertical motion depends on the parcel's degree of saturation; and finally, if the lapse rate is equal to or greater than the dry adiabatic lapse rate, then the local atmosphere is absolutely unstable. Instability encourages vertical motions, which mixes scalar quantities such as potential temperature, water vapor, and pollutants and causing more uniform vertical profiles of these quantities. The lapse rate for each 100 m altitude layer within each O_3 profile included within a cluster was first computed, and then these lapse rates were separated into boundary layer or free tropospheric lapse rates based on the PBL height associated with that O_3 profile. The median values and distributions of lapse rates within the PBL and free troposphere were then compared for each O_3 cluster to determine its influence on profile shape; results are displayed in Fig. 3.7a)-d).

As expected, clusters that exhibited a larger median boundary layer lapse rate value (indicating boundary layers that were more unstable) also exhibited a more well mixed median O_3 profile, while smaller median PBL lapse rate values (indicating more stable boundary layers) were associated with less well mixed O_3 profiles; this is consistent with the influence of the potential temperature profiles. For each campaign, the PBL or free tropospheric lapse rates were often statistically the same among clusters (i.e., overlap of the boxes in Fig. 3.7 representing the 25th and 75th percentile values), which likely dampens differences in profile shapes among clusters within a campaign. Additionally, for some Texas and Colorado clusters (Fig. 3.7b) and 3.7d), the lapse rates were often statistically the same between the PBL and the free troposphere, resulting in smoother median O_3 profiles throughout the entire profile depth than seen for the Maryland or California campaigns (Fig. 3.7a) and 3.7c). More interesting median O_3 profile behavior resulted when differences existed between the PBL and free tropospheric lapse rate

values that were larger than for the other clusters within a campaign. For example, Clusters 1 and 2 of the Texas campaign exhibited such large differences between the PBL and the free troposphere, resulting in O_3 profiles that were well mixed within the PBL and which displayed more layered behavior within the free troposphere; the same profile behavior can be seen for Cluster 2 of the Maryland campaign.
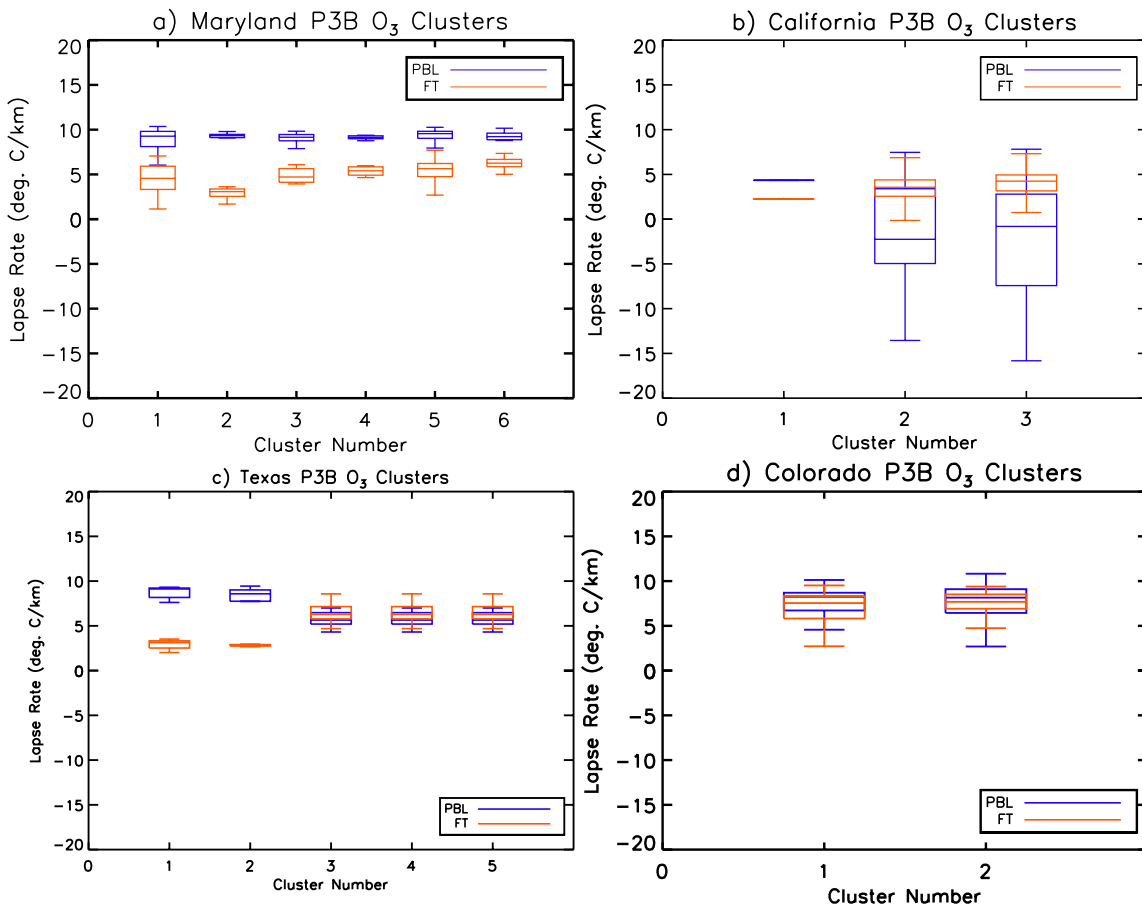
Fig. 3.7: The median lapse rates and distributions associated with each *in situ* O_3 profile cluster, computed separately for the PBL and free troposphere (FT), a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign.

Lastly, flight date during the campaign and spiral sampling time influence the relative magnitudes of the median cluster profiles for the California (Tables 3.1 and 3.2), Texas (Tables

3.3 and 3.4), and Colorado (Tables 3.5 and 3.6) campaigns. The most polluted median profiles were associated with either O_3 pollution episodes or campaign periods and sampling times conducive to O_3 photochemistry. For example, 50.8% of the profiles in Colorado Cluster 1 occurred during the first campaign period (July 17-21, 2014) and were sampled mainly during Spirals 2 and 3 (~11am and ~4pm MDT), while Cluster 2 was sampled mostly during the last campaign period (Aug. 6-10, 2014) and during Spirals 1 and 2 (~9am and ~11am MDT); this is consistent with Cluster 1 demonstrating greater O_3 mixing ratios within the PBL than Cluster 2. It should be noted that July often experienced larger surface O_3 concentrations than did August in Colorado (Mazzuca et al., in prep.). Likewise, Texas Clusters 1 and 2, the most polluted median profiles, contained profiles entirely from the last campaign period (Sept. 24-26, 2013), when a pollution episode occurred, and were primarily sampled during Spiral 3 (~2pm CDT). Sampling time has a greater influence over the relative cluster magnitudes than campaign period for the California campaign, as this campaign took place during the winter. California Cluster 2 sampled primarily during Spirals 1 and 2 (~8am and ~11am PST), while the most polluted Cluster 3 sampled roughly evenly among all three spiral times (~8am, ~11am, and ~1pm PST). The Maryland campaign also demonstrated a weak influence of spiral location for Cluster 4: only Edgewood and Essex afternoon profiles were included in this cluster, such that some influence of the bay breeze may be present.

California Campaign Campaign Period			
Dates	**Cluster 2**	**Cluster 1**	**Cluster 3**
20130116-20130122	48.6%	0%	56.8%
20130130-20130206	51.4%	100%	43.2%

Table 3.1: Percentages of profiles within each cluster that fell within each campaign period, as denoted by the Dates column, for the California campaign O$_3$ clusters. Clusters listed from left to right in order of least to most polluted.

California Campaign Spiral Sampling Time			
Spiral	**Cluster 2**	**Cluster 1**	**Cluster 3**
Spiral 1 (~8am PST)	42.9%	0%	31.8%
Spiral 2 (~11am PST)	40.0%	0 %	31.8%
Spiral 3 (~1pm PST)	17.1%	100%	36.4%

Table 3.2: Percentages of profiles within each cluster that fell within each spiral sampling time, as denoted by the Spiral column, for the California campaign O$_3$ clusters. Clusters listed from left to right in order of least to most polluted.

Texas Campaign Campaign Period					
Dates	Cluster 3	Cluster 4	Cluster 5	Cluster 1	Cluster 2
20130904-20130906	28.6%	21.5%	3.4%	0%	0%
20130911-20130914	0%	65.8%	49.2%	0%	0%
20130924-20130926	71.4%	12.7%	47.5%	100%	100%

Table 3.3: Percentages of profiles within each cluster that fell within each campaign period, as denoted by the Dates column, for the Texas campaign O$_3$ clusters. Clusters listed from left to right in order of least to most polluted.

Texas Campaign Spiral Sampling Time					
Spiral	Cluster 3	Cluster 4	Cluster 5	Cluster 1	Cluster 2
Spiral 1 (~9am CDT)	71.4%	43.0%	23.7%	25%	25%
Spiral 2 (~12pm CDT)	21.4%	32.9%	40.7%	25%	0%
Spiral 3 (~2pm CDT)	7.1%	22.8%	35.6%	50%	75%
Spiral 4 (~3pm CDT)	0%	1.3%	0%	0%	0%

Table 3.4: Percentages of profiles within each cluster that fell within each spiral sampling time, as denoted by the Spiral column, for the Texas campaign O$_3$ clusters. Clusters listed from left to right in order of least to most polluted.

Colorado Campaign Campaign Period		
Dates	**Cluster 2**	**Cluster 1**
20140717-20140721	28.7%	50.8%
20140728-20140803	30.7%	42.9%
20140806-20140810	40.6%	6.3%

Table 3.5: Percentages of profiles within each cluster that fell within each campaign period, as denoted by the Dates column, for the Colorado campaign O$_3$ clusters. Clusters listed from left to right in order of least to most polluted.

Colorado Campaign Spiral Sampling Time		
Spiral	**Cluster 2**	**Cluster 1**
Spiral 1 (~9am MDT)	55.4%	26.9%
Spiral 2 (~11am MDT)	39.6%	33.3%
Spiral 3 (~4pm MDT)	4.9%	39.7%

Table 3.6: Percentages of profiles within each cluster that fell within each spiral sampling time, as denoted by the Spiral column, for the Colorado campaign O$_3$ clusters. Clusters listed from left to right in order of least to most polluted.

Airmass origin, as revealed by HYSPLIT back trajectories, also influenced the relative magnitudes for the Maryland (Table 3.7) and Colorado campaigns (Tables 3.8-3.12). Maryland Cluster 4, the most polluted within the PBL and the only significant Maryland cluster, contained profiles whose corresponding back trajectories originated over the polluted Northwest Canada/Great Lakes region at all altitude levels between 500 m and 2500 m; the less polluted

clusters included back trajectory origins over less polluted regions, such as northern Canada, at all levels. The Colorado profile back trajectories associated with Cluster 1 (most polluted) demonstrated larger percentages of air recirculated over the Denver-Boulder region, thus recirculating urban pollution, than did Cluster 2. Though airmass origin did not emerge as a significant influence on the California cluster magnitudes, it is interesting to note that the back trajectories emphasized recirculation within the San Joaquin Valley, which is consistent with the NARR plots demonstrating significant valley recirculation and horizontal mixing during this campaign.

Maryland Campaign HYSPLIT Back Trajectories						
Airmass Origin	Cluster 1	Cluster 2	Cluster 3	Cluster 5	Cluster 6	Cluster 4
N. Canada	54.2%	27.3%	20%	11.1%	7.7%	0%
NW Canada/Great Lakes	0%	0%	10%	25.9%	0%	100%
Long Range Transport	16.7%	27.3%	0%	7.4%	7.7%	0%
Ohio River Valley/Westerly Flow	25%	45.5%	70%	46.3%	84.7%	0%

Table 3.7: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Maryland campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at all vertical levels included. Percentages computed as percentage of profiles included within a profile cluster that fell within each HYSPLIT cluster. Note that percentages in each column may not total to 100%, as some profiles included in the profile clusters may have been rejected by the HYSPLIT back trajectory clustering algorithm.

Colorado Campaign HYSPLIT Back Trajectories at 500 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	12.9%	19.0%
Southwesterly Flow	29.7%	23.8%
Local Recirculation	22.7%	52.4%
Westerly Flow	1.9%	0%

Table 3.8: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O$_3$ clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 500 m vertical level only. Percentages computed as percentage of profiles included within a profile cluster that fell within each HYSPLIT cluster. Note that percentages in each column may not total to 100%, as some profiles included in the profile clusters may have been rejected by the HYSPLIT back trajectory clustering algorithm.

Colorado Campaign HYSPLIT Back Trajectories at 1000 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	22.8%	15.9%
Southwesterly Flow	37.6%	42.9%
Local Recirculation	11.9%	38.1%
Westerly Flow	1.9%	0%

Table 3.9: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O$_3$ clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 1000 m vertical level only. Percentages computed as percentage of profiles included within a profile cluster that fell within each HYSPLIT cluster. Note that percentages in each column may not total to 100%, as some profiles included in the profile clusters may have been rejected by the HYSPLIT back trajectory clustering algorithm.

Colorado Campaign HYSPLIT Back Trajectories at 1500 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	9.9%	11.1%
Southwesterly Flow	32.7%	23.8%
Westerly Flow	3.9%	11.1%
Local Recirculation	25.7%	44.4%

Table 3.10: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 1500 m vertical level only. Percentages computed as percentage of profiles included within a profile cluster that fell within each HYSPLIT cluster. Note that percentages in each column may not total to 100%, as some profiles included in the profile clusters may have been rejected by the HYSPLIT back trajectory clustering algorithm.

Colorado Campaign HYSPLIT Back Trajectories at 2000 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	28.7%	12.7%
West-northwesterly Flow	23.8%	36.5%
Local Recirculation	16.8%	42.9%

Table 3.11: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 2000 m vertical level only. Percentages computed as percentage of profiles included within a profile cluster that fell within each HYSPLIT cluster. Note that percentages in each column may not total to 100%, as some profiles included in the profile clusters may have been rejected by the HYSPLIT back trajectory clustering algorithm.

Colorado Campaign HYSPLIT Back Trajectories at 2500 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	26.7%	12.7%
West-northwesterly Flow	24.8%	46.0%
Local Recirculation	19.8%	38.1%

Table 3.12: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O$_3$ clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 2500 m vertical level only. Percentages computed as percentage of profiles included within a profile cluster that fell within each HYSPLIT cluster. Note that percentages in each column may not total to 100%, as some profiles included in the profile clusters may have been rejected by the HYSPLIT back trajectory clustering algorithm.

3.8.2 P-3B Ozone Profile Cluster Correlations

Values of R^2 from the simple linear regression analyses between surface mixing ratio and column abundance data for each profile cluster (cluster correlations) are summarized in Tables 3.13-3.14 for the Maryland and California campaigns. The Texas and Colorado campaigns did not produce statistically significant cluster correlations (i.e., R^2 not statistically different from 0.0) and so are not included in the tables. The column-surface correlations over the full set of profiles used with the hierarchical cluster analysis (full correlations) are also included in Tables 3.13-3.14. Representative scatter plots are displayed in Fig. 3.8a)-d). It should be noted that clusters containing fewer than five members contained too few members for a meaningful analysis and are excluded here. The California, Texas, and Colorado campaigns yielded very few statistically significant correlations, while the Maryland campaign yielded many significant and large cluster correlations. It is also notable that, for each cluster during each campaign, the column_air and column_ground correlation analyses typically yielded R^2 values that were

statistically the same, despite the different gap-filling methods (and inherent assumptions about the strength of boundary layer mixing in these methods) used in the column computations. This similarity in R^2 values indicates that O_3 remained generally well mixed horizontally and vertically in the lower troposphere during each campaign, consistent with the findings of Flynn et al. (2014) for the Maryland campaign.
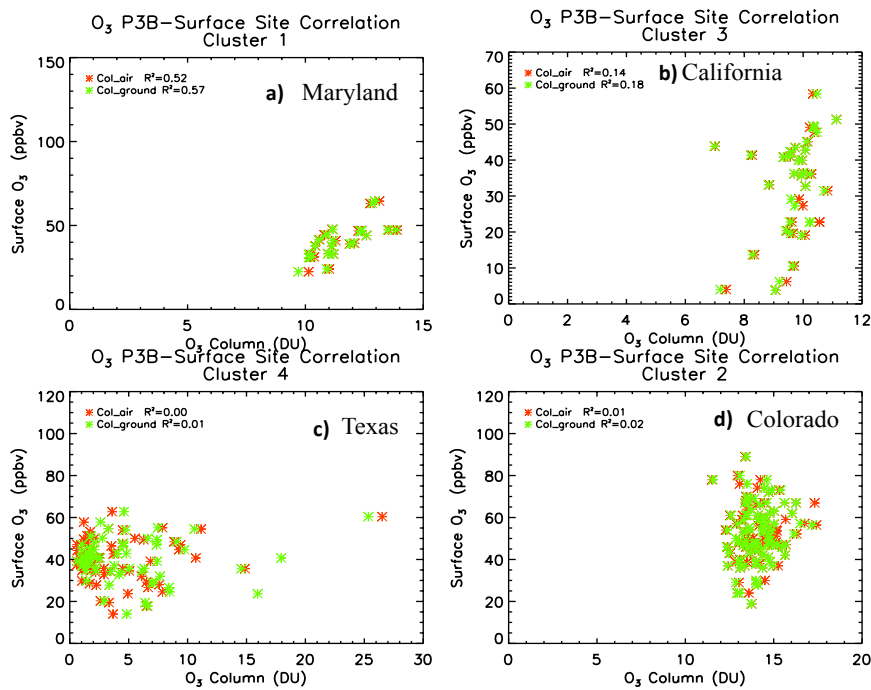
Fig. 3.8. Representative scatter plots for the O_3 column-surface correlations a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. R^2 values for the column_air and column_ground correlations displayed in the legends.

Maryland Campaign O_3 Correlations		
Cluster Number	**Column_Air R^2**	**Column_Ground R^2**
1	0.53	0.57
2	0.87	0.88
3	0.24	0.34
4	N.S.	N.S.
5	0.14	0.23
6	N.S.	0.36
Full	0.71	0.74

Table 3.13: R^2 values for the correlation between column and surface data for each O_3 profile cluster of the Maryland campaign. Cluster correlations denoted in red font indicate clusters that were statistically significantly different from the full correlation.

California Campaign O$_3$ Correlations		
Cluster Number	**Column_Air R^2**	**Column_Ground R^2**
1	--	--
2	N.S.	N.S.
3	0.15	0.19
Full	0.08	0.11

Table 3.14: R^2 values for the correlation between column and surface data for each O$_3$ profile cluster of the California campaign. No clusters presented a statistically different correlation from the full correlation.

Differences in degree of correlation between the Maryland campaign and the others campaigns may be due again to the season in which the campaign took place, as well as the vertical wind shear of the large-scale flow over each study region. Wintertime stagnation during the California campaign interfered with the column-surface connection, as O$_3$ could not be mixed vertically as efficiently as for convective boundary layers, while the Colorado campaign did not experience convective boundary layers as deep as for Maryland or Texas. The California and Colorado campaigns also experienced changes in wind direction and circulation patterns with height, as indicated by the NARR wind fields at several pressure levels, discussed previously. Air was transported from different source locations at different heights, rather than simply being mixed in the vertical, further interfering with the column-surface connection. Again in contrast, as Maryland experienced similar wind patterns at most levels and deep, convective boundary layers, O$_3$ was most well mixed horizontally and vertically during this campaign relative to the other three campaigns, allowing for greater connection between column and surface. Texas was

in a complex chemical environment, such that O_3 production was much more localized, as evidenced by its low correlation with CO over all profiles included in the clustering analysis (Fig. 3.9).
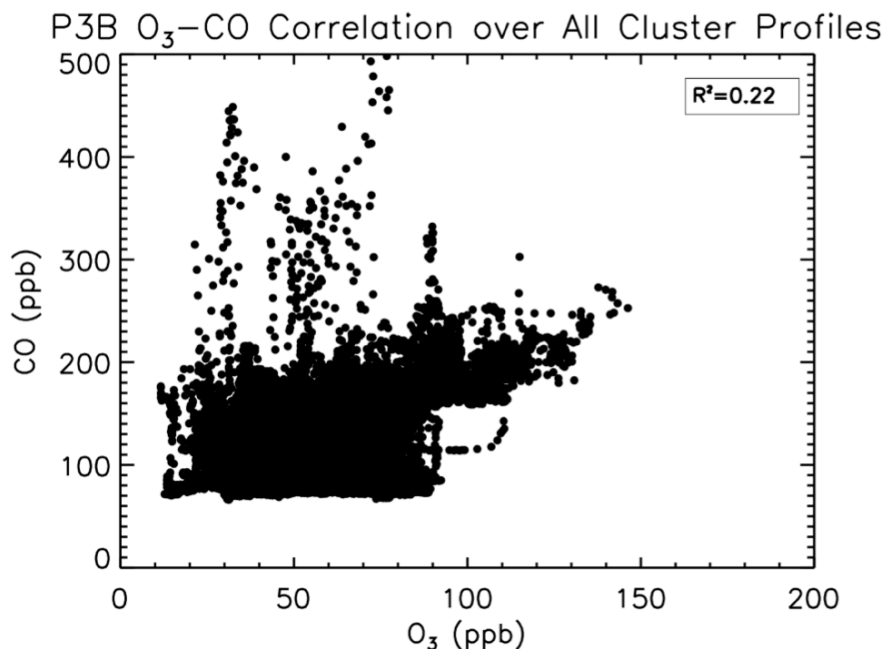
Fig. 3.9: Correlation plot of *in situ* CO vs. O_3 over the spiral depths of the profiles included in the clustering analysis for the Texas campaign.

These cluster correlations were also compared to the full correlation for the Maryland (Table 3.13) and California (Table 3.14) campaigns. As the Texas and Colorado full correlations were not themselves significant, the cluster correlations were statistically the same as the full correlations. Neither California Cluster 2 nor Cluster 3 was found to be statistically significantly different from the full correlation at a confidence level of 95% for both column analyses (Table 3.14). This indicates that no correlation analysis for this campaign was greater than zero. However, half of the Maryland clusters (Cluster 3, 5, and 6; Table 3.13) were significantly

smaller than the full correlation for both analyses; the Cluster 1, 2, and 4 correlations were not significantly different from the full correlation analyses (which also presented a high degree of correlation). It is notable that the median profiles associated with Clusters 1 and 2 were well mixed within the PBL, whereas those for Clusters 3, 5, and 6 were less well mixed and presented more vertical variation within the PBL. Thus, the column-surface correlation for these clusters degraded relative to those for the well mixed Clusters 1 and 2 and the full correlations under conditions of inhibited vertical mixing. This is also consistent with the results for the California campaign: the Cluster 2 median profile was less well mixed than that for Cluster 3, though neither profile could be considered well mixed. The Maryland Cluster 4 correlation presents an exception: this cluster contained only five profiles, fewer than any other Maryland cluster, that may have been influenced by the bay breeze, causing an interruption in the column-surface connection.

The results presented above provide a mixed message for the ability to relate satellite O_3 column observations to surface mixing ratios. The complex meteorological or chemical conditions encountered during the Colorado or Texas campaigns indicate that satellite observations of lower tropospheric column O_3 may not be useful for estimating surface ozone under such conditions. Because the correlations for each cluster and the full correlation analyses were essentially null during the California campaign, this suggests that it matters little if the satellite retrieval correctly approximates the O_3 profile, as the stagnant conditions still prevent there being a connection between column and surface data. The Maryland campaign also indicated degraded column-surface correlation under conditions of poorer vertical mixing, as indicated by the median O_3 and potential temperature profiles for Clusters 3, 5, and 6. However, the correlation analyses suggest that satellites may have the best chance to relate to surface ozone

under the conditions encountered during the Maryland campaign Clusters 1 and 2, which include deep, convective boundary layers and few interruptions to this connection from complex meteorology, chemical environments, or orography.

3.8.3: Ozone Shape Factor Comparisons to CMAQ and GMI

As with the potential temperature median profiles for each cluster, the CMAQ- and GMI-simulated O_3 and NO_2 mixing ratio profiles coincident to each observed profile within a cluster were sampled, and the median shape factor then computed for each model and each cluster. CMAQ well captured the shapes of the median P-3B shape factor profiles for most clusters in each campaign. Shape factors were compared in this analysis because these are used in remote sensing retrievals, rather than mixing ratio profiles; such a comparison elucidates potential implications of model errors for retrievals. Representative comparison plots are displayed in Fig. 3.10a)-d). Interestingly, CMAQ reproduced the complex vertical behavior of the Colorado median shape factor profiles, though CMAQ also performed well relative to the P-3B during the other three campaigns. However, the model placed the upper PBL peak shape factor value approximately 0.5 km above the observed peak within the PBL for many clusters, indicating that, while reproducing the shapes of the median shape factor profiles, CMAQ placed a greater portion of the O_3 column higher in the vertical than seen in the observations during each campaign. This is particularly evident for the Texas and Colorado comparison plots in Fig. 3.10c) and d), and suggests that vertical mixing may be overestimated within the model, as O_3 is a relatively long-lived gas within the PBL. Additionally, CMAQ generally displayed a slight low bias in shape factor values within the PBL and a slight high bias within the free tropospheric portions of the shape factors relative to the P-3B for the Maryland, California and Texas campaigns, and bias

direction varying with altitude during the Colorado campaign. This further indicates that CMAQ placed a greater portion of the O_3 column higher in the vertical than seen in the observations.

CMAQ well captured the magnitudes of P-3B O_3 shape factors for all clusters during the California campaign and for Clusters 1-3 and 5 of the Maryland campaign, with R^2 values of 0.80 and larger (Fig. 3.11a)-b), when correlating CMAQ and observed shape factor profile values for individual model layers across all shape factors within a cluster. However, CMAQ moderately well captured the shape factor values for all clusters during the Texas and Colorado (Fig. 3.11c)-d) campaigns, and Clusters 4 and 6 of the Maryland campaign, with R^2 values between 0.50 and 0.80. It is notable that Maryland Cluster 4 contained profiles only from Edgewood and Essex, and that Cluster 6 contained a large percentage of profiles from Essex. These were the two most polluted median cluster profiles, with PBL mixing ratio values of approximately 80 ppbv or higher; in agreement with this, Texas Clusters 1 and 2, with median profile PBL mixing ratio values in excess of 80 ppbv, also displayed the lowest R^2 values for that campaign. Thus, CMAQ was less able to capture the magnitudes of the shape factors for very polluted profile clusters, and for the two campaigns with complex pollution or meteorological conditions (Texas and Colorado, respectively). This further suggests that CMAQ best captured the O_3 column relative vertical distribution under conditions that were unfavorable to O_3 photochemical production (California campaign), or under conditions of moderate pollution (Maryland Clusters 1-3 and 5). However, the scatter plots presented in Fig. 3.11 (each scatter plot features five altitude ranges, which were chosen to provide one altitude range for the lowermost PBL spiral altitudes, two altitude ranges for the mid-PBL, one altitude range for the upper PBL/lower free troposphere, and one altitude range for the uppermost spiral altitudes in the free troposphere) also indicate that CMAQ often under- or overpredicted the magnitudes of the shape factors relative to the P-3B. CMAQ tended to

overpredict at many vertical levels during the California campaign (Fig. 3.11b), with the exceptions of the 600-1200 m AMSL altitude ranges, which corresponded to the upper PBL and lower free troposphere. Remote sensing retrievals are more sensitive to the upper PBL/lower free troposphere altitude region than near the surface (as determined by the scattering weights), such that errors in the shape factor in this region relative to other altitudes lead to greater retrieval errors. Like the median cluster shape factor profiles, this overprediction in the lower PBL and in the upper PBL/lower free troposphere suggests that CMAQ placed too much of the O_3 column burden within fewer vertical layers (the lower PBL and near the tops of the California P-3B spirals). In contrast, during the Maryland campaign, CMAQ tended to underpredict in the lower PBL (300-800 m AMSL) and less often overpredicted above 2000 m AMSL (free troposphere), and compared well to the P-3B shape factor values in the middle portions of the shape factor profiles (800-2000 m AMSL; Fig. 3.11a). Such underprediction within the lowest PBL portions of the Maryland shape factors, with some compensating overprediction at higher altitudes, again suggests that CMAQ placed too much of the O_3 column burden within fewer vertical layers than seen in the observations. The scatter plots for the Colorado campaign (Fig. 3.11d) are consistent with the model biases suggested by the median shape profiles: CMAQ demonstrated underprediction within the lower PBL (below 2500 m AMSL, or below 1000 AGL) and within the highest spiral altitudes (above 4000 m AMSL or 2500 m AGL), and overprediction for altitudes between 2500-4000 m AMSL. The values of reduced chi squared, given in the legends in each plot in Fig. 3.11, demonstrate the impact of these over- and underpredictions on the ability of CMAQ to capture the P-3B O_3 shape factor values: the simulated and observed magnitudes are correlated for most clusters for each campaign, and clearly demonstrate a linear relationship, but the large errors associated with the under- and overpredicted values lead to large regression errors

and so larger reduced chi squared values. These results and those comparing the median shape factors profiles are encouraging that a regional air quality model such as CMAQ may be able to replicate the shape factors during winter, when photochemical O_3 production is inhibited (California), and the moderate pollution conditions associated with the Maryland campaign Clusters 1-3 and 5.

Additionally, the CMAQ and P-3B CO shape factor magnitudes associated with each O_3 profile cluster were compared as a preliminary investigation of the errors in simulated vertical mixing and the impact of those errors on the simulated O_3 shape factors (Fig. 3.12). Errors in the simulated O_3 shape factors due to erroneous simulated vertical mixing are in addition to errors in the emissions used to drive CMAQ and in the chemical mechanism employed in the model. Indeed, errors in emissions and chemistry in air quality models have been extensively studied using observations over the eastern U.S. For example, Anderson et al. (2014) demonstrated that NO_x emissions were overestimated within the NEI by 51-70% in Maryland, relative to observations from the Maryland DISCOVER-AQ deployment, while CO emissions within the NEI were much more accurate (average overprediction of 15%). Building upon this work, Canty et al. (2015) and Goldberg et al. (2016) examined the impact of updated emissions (mobile NO_x emissions reduced by 50%), as well as changes to the chemical mechanism, on simulations with CMAQ and the Comprehensive Air Quality Model with Extensions (CAMx), respectively. Canty et al. (2015) demonstrated that CMAQ underestimated the NO_2 column over rural areas of the eastern U.S. and overestimated over urban areas, relative to OMI NO_2 tropospheric vertical column observations. After implementing a 50% reduction in mobile NO_x emissions, a reduction of the lifetime of alkyl nitrates in the chemical mechanism from ~10 days to ~1 day, and updates to the biogenic emissions module within CMAQ to better represent isoprene emissions, the

simulated urban/rural NO_2 column ratio improved relative to the same ratio in the OMI observations. Goldberg et al. (2016) implemented the same improvements as Canty et al. into CAMx, as well as increases to the dry deposition velocities of isoprene nitrates, which resulted in improved representations of formaldehyde and NO_y mixing ratio magnitudes over the depth of the P-3B spirals in the model, relative to the Maryland campaign observations. These works demonstrate the clear importance of accurate emissions and chemistry within air quality simulations, and demonstrate impacts of such errors on the magnitude of the simulated O_3 and NO_2 profiles (and thus potentially on the shape factors). However, the impact of erroneous vertical mixing on the simulated shape factors is the focus of this section, as Chapter 2 demonstrated that overestimated vertical mixing within CMAQ may have led to greater column-surface correlation within the model relative to the observations for the Maryland campaign; this is consistent with other works which suggested that vertical mixing within the PBL was too fast relative to observations within CMAQ (Castellanos et al., 2011).

The CO comparisons for the Maryland campaign demonstrated the clearest case of overestimated vertical mixing of all four deployments (Fig. 3.12a). As with the O_3 shape factors, the CO shapes factors were underestimated within the lowermost portion of the profiles (below 1200 m AMSL), with the bias direction transitioning through the middle altitude range of the profiles, leading to general overestimation of the CO shape factor values in the upper PBL and lower free troposphere (1600 m AMSL and above). As CO is a long-lived chemical species (much longer than the time scale of turbulent mixing in the PBL; Zhang et al., 2016), with emissions dominated by surface sources, underestimation in the lower simulated shape factor profile and overestimation aloft suggests that the model mixed too much CO into the upper portions of the profiles, and thus likely over-mixed other pollutants such as O_3. This is consistent

with the results of Chapter 2. The California clusters also suggested overly vigorous vertical mixing relative to the observed CO shape factors. The simulated CO shape factor values tended to be overestimated throughout most of the PBL (50-900 m AMSL), transitioning to a lack of preferred bias direction above 900 m. This is consistent with the overestimation of the O_3 shape factors below 600 m. However, as the depth to which the CO shape factors were overestimated extends further than the depth to which the O_3 values were overestimated, this suggests that, while errors in vertical likely caused errors in the O_3 column vertical distribution, vertical mixing may have played less of a role than for the Maryland campaign. In contrast, the CO comparisons for the Colorado campaign suggest that vertical mixing may have been underestimated relative to the observations, and that CO emissions may have been overestimated; simulated CO shape factor values were overestimated below 2500 m AMSL (1500 m AGL), and tended to be underestimated throughout the remainder of the profile depth. This is does not correspond to the underestimated O_3 shape factor values below 2500 m and above 4500 m AMSL, with overpredicted O_3 between these altitude ranges. Vertical mixing may have exerted the least influence on column values during the Colorado campaign, relative to the other three campaigns, due to complex circulations such as the Denver cyclone or the mountain solenoidal circulations that models struggle to replicate.

Fig. 3.10: Representative shape factor comparison plots for CMAQ vs. P-3B for O₃ a) for the Maryland campaign; b) for the California campaign, c) for the Texas campaign; d) for the Colorado campaign. Computed on CMAQ vertical grid. All altitudes are AMSL.

Fig. 3.11: Representative shape factor comparison scatter plots of modeled and observed shape factor values for CMAQ and the P-3B for O$_3$ a) for the Maryland campaign Cluster 1; b) for the California campaign Cluster 2, c) for the Texas campaign Cluster 3; d) for the Colorado campaign Cluster 1. Computed on CMAQ vertical grid. Scatter plots colored by altitude layers (AMSL) with legend in bottom right displaying the altitude layer ranges for each campaign. Example uncertainty bars also displayed in black for one point; uncertainty for P-3B taken from uncertainty of observed shape factors while uncertainty for CMAQ taken as the standard deviation over the simulated shape factors.

Fig. 3.12: Representative shape factor comparison scatter plots of modeled and observed shape factor values for CMAQ and the P-3B for CO associated with each O_3 profile cluster a) for the Maryland campaign Cluster 1; b) for the California campaign Cluster 2, c) for the Colorado campaign Cluster 1. Computed on CMAQ vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. Example uncertainty bars also displayed in black for one point; uncertainty for P-3B taken from uncertainty of observed shape factors while uncertainty for CMAQ taken as the standard deviation over the simulated shape factors.

The GMI median shape factors were also computed for each cluster of the Maryland and California campaigns (Fig. 3.12a)-b). GMI well captured the altitude variation of the O_3 shape factor profiles for both campaigns, indicating that this global model captured the relative O_3 vertical distributions. However, unlike CMAQ, GMI tended to display a low bias in shape factor

values within the lowermost PBL and a high bias within the mid-PBL, with the preferred bias direction changing for each cluster within the free tropospheric portion of the shape factor profiles, indicating that GMI did not mix O_3 as high in the vertical as in the observations and placed too much of the O_3 column burden within the PBL and lower portions of the profiles. Comparisons of observed and GMI-simulated CO shape factors (Fig. 3.13) typically display overestimates in CO shape factor magnitudes between 800 m and 1600 m for the Maryland campaign, and 300-900 m for the California campaign (the middle portion of the PBL for each campaign), with underestimates in the lowermost PBL. This suggests that mixing into the middle portion of the PBL from the lowermost profile is overestimated in GMI for both campaigns, though not as severely as for CMAQ. A global CTM such as GMI likely also suffers from similar deficiencies in emissions estimates and in the simulated chemical mechanisms as described for regional models by Canty et al., Anderson et al, and Goldberg et al. Such errors would also impact the simulated shape factors, though vertical mixing remains the focus of this analysis. GMI performed similarly well as CMAQ for each campaign relative to the P-3B O_3 cluster median shape factors, while CMAQ was also able to capture the vertical structure of the Colorado campaign shape factors to some extent. This implies that a regional model may be able to estimate O_3 shape factors accurately enough for use in remote sensing retrieval algorithms. It is also initially surprising that a global model performed well relative to the observations just as the regional model did; however, this may be explained by the vertical resolutions of both models, particularly within the lowermost 2 km AGL of the shape factors. The vertical resolution of CMAQ varied with campaign, as described previously, though CMAQ contained at least 13 layers within the lowermost 2 km; the Texas campaign contained 18 layers within the lowermost 2 km. In contrast, GMI contained only 13 layers for both the Maryland and California campaigns

within the lowest 2 km AGL, a coarser resolution than most CMAQ grids. Computing the P-3B shape factors on the GMI vertical grid, rather than the CMAQ grid or some other vertical resolution, likely smoothed out some of the variability within the observations, thereby improving the ability of the global model to capture the shape factors. This speaks further to the potential better ability of CMAQ ozone profiles to perform well within retrieval algorithms, as CMAQ does not smooth out as much variability as GMI and yet still produced realistic profiles that compared well with the P-3B.

GMI moderately well captured the P-3B O_3 shape factor values for all clusters during the California campaign, and for Clusters 1-3 during the Maryland campaign (Fig. 3.14a)-b). Values of R^2 for the correlation between P-3B and GMI shape factor values ranged between 0.60 and 0.90. However, GMI compared more poorly for Clusters 4-6 of the Maryland campaign, with R^2 values less than 0.60 (Fig. 3.14c-d). The values of reduced chi squared also support these results; as the relationship between the simulated and observed shape factor values departed from linearity (such as in Fig. 3.14b), the value of reduced chi squared increased, reflecting the greater errors associated with the over- and underpredictions that caused the relationship to depart from linearity. The contrast in performance for these two groups of clusters is again explained by the difference in degree of pollution for the cluster median *in situ* profiles: Maryland Clusters 4-6 were more polluted than Maryland Clusters 1-3 and all California clusters, with median PBL mixing ratios greater than 65 ppbv. This suggests that GMI performed better for clean to moderately polluted conditions. The scatter plots for GMI vs. P-3B shape factor values demonstrate that GMI often underpredicted the shape factor values within the lowermost portions of the profiles, while above this altitude range, GMI well predicted or slightly overpredicted the shape factor values relative to the P-3B; GMI again demonstrated a tendency

to underpredict in the lower free troposphere during the Maryland campaign and overpredict during the California campaign It is these underpredictions of small shape factor values that likely drive down the correlations between simulated and observed shape factors. These altitude biases coupled with the smooth median shape factor profiles indicate that the GMI relative O_3 vertical distributions were too homogeneous relative to the observations, as GMI placed more of the O_3 column burden approximately evenly within the PBL and the lower free troposphere than seen in the observations, causing GMI to underpredict the O_3 distribution within most of the free tropospheric portions of the profiles. The global GMI model is thus less adequate than CMAQ to accurately simulate the lower tropospheric O_3 shape factors.

Fig. 3.13: Representative shape factor comparison plots for GMI vs. P-3B for O_3 a) for the Maryland campaign; b) for the California campaign. Computed on GMI vertical grid.
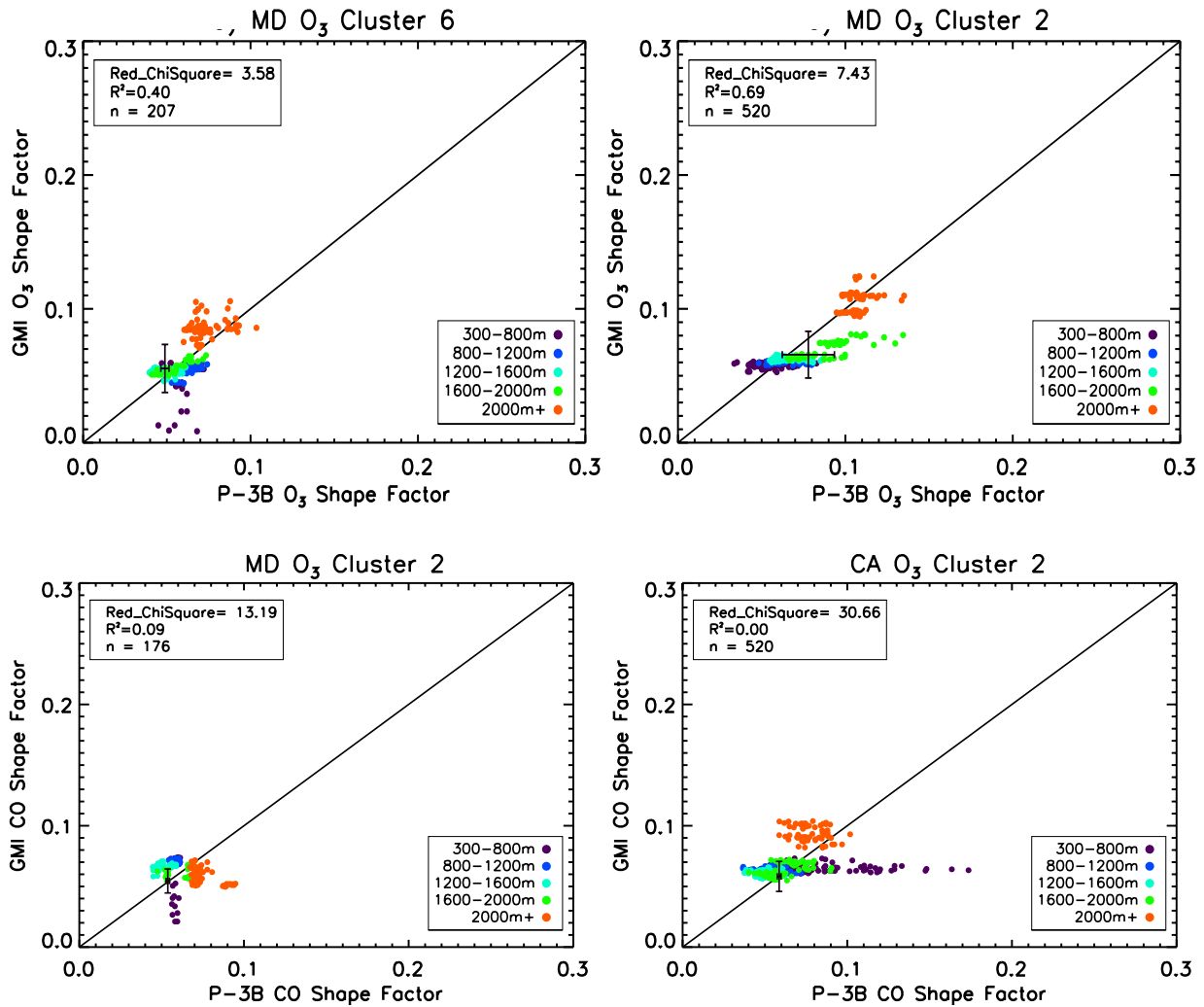
Fig. 3.14: Shape factor comparison scatter plots for O$_3$ and CO of model bias values for GMI vs. P-3B for scatter plot for the Maryland campaign Cluster 6 and 2; and for the California campaign Cluster 2. Computed on GMI vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. Example uncertainty bars also displayed in black for one point; uncertainty for P-3B taken from uncertainty of observed shape factors while uncertainty for GMI taken as the standard deviation over the simulated shape factors.

3.9: Nitrogen Dioxide P-3B *In Situ* Profile Clusters and Comparison to Models

3.9.1: P-3B NO$_2$ Profile Clusters

The median profiles for the P-3B NO$_2$ clusters are displayed in Fig. 3.15a)-d). The Maryland campaign produced the greatest number of NO$_2$ clusters, though none of these were

significant (i.e.; all median profiles were statistically the same; Fig 3.15a). The Texas campaign also produced a large number of clusters yet only two significant clusters, each containing only one profile; all other clusters were statistically the same (Fig. 3.15c). These results indicate that the NO_2 profiles displayed relatively uniform behavior during these two campaigns, which may be expected given that these were warm season campaigns and flights were generally conducted on sunny days with convective boundary layers and with conditions conducive for O_3 production. On the other hand, all three clusters obtained for the California and Colorado campaigns were significant (Figs. 3.15b) and 3.15d). Distinction among clusters was again found only within the PBL rather than the free troposphere, suggesting influence of both chemistry and vertical mixing on NO_2 profile shapes. However, as Cluster 1 contained the vast majority of profiles, the Colorado results again indicate that the NO_2 profiles displayed relatively uniform behavior throughout the campaign period; only the California campaign indicated any NO_2 profile variability. The largest ranges of mixing ratio values were also encountered during the California campaign. The uniqueness of the California NO_2 profile clusters may have been due to the inhibited O_3 formation during this campaign; less NO_2 was converted to O_3 during the day, allowing a greater variation in profile shapes than the other three summer campaigns, in which NO_2 was more rapidly converted to O_3 during the day. The significant Texas and Colorado clusters may be considered as outliers, as they each contained few profile members. No clear meteorological conditions emerged as influences on the NO_2 clusters, further indicating that NO_2 profile variability is influenced by photochemical loss of NO_2 as well as vertical mixing, as may be expected given its short lifetime; this result is consistent with Zhang et al. (2016), which found that the observed NO_x vertical gradient during the Maryland campaign is sensitive to both chemistry and boundary layer mixing.
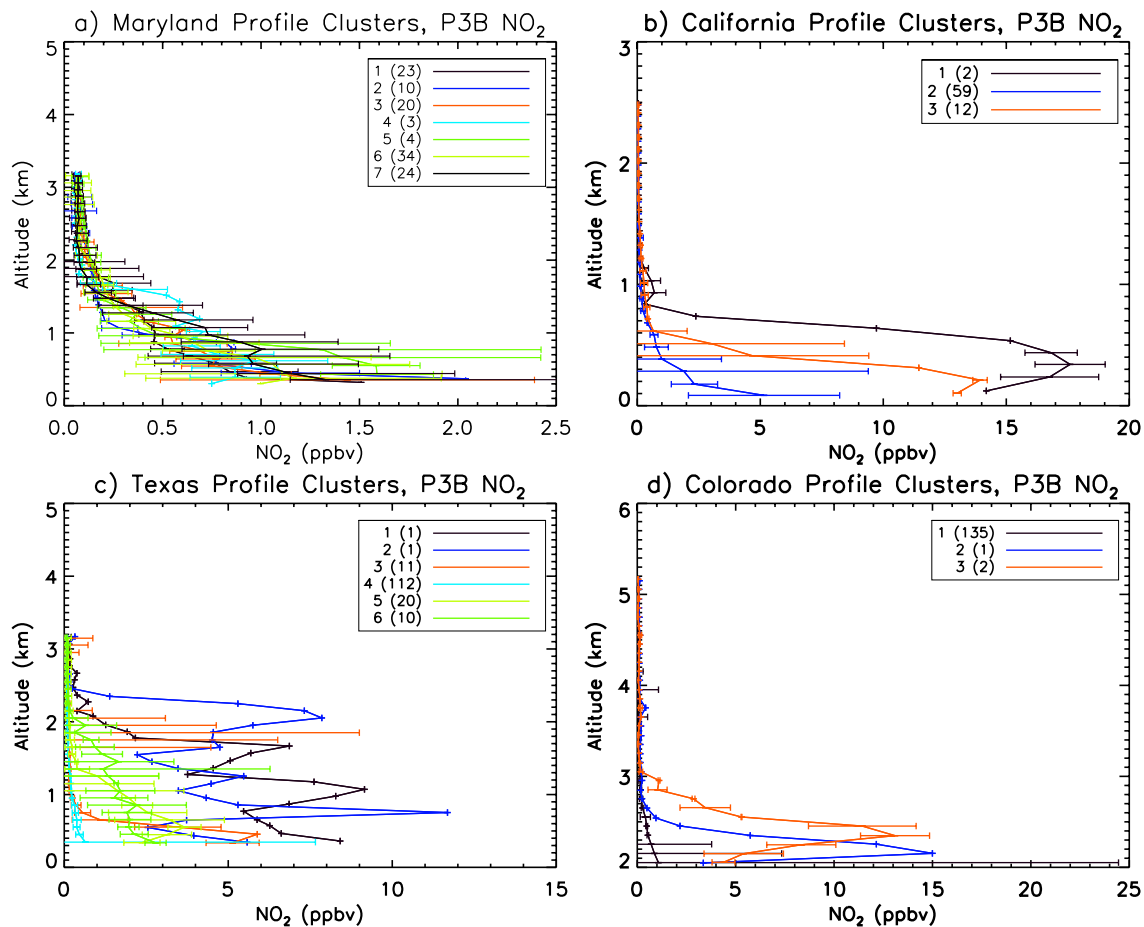
Fig. 3.15: The median profiles for each *in situ* NO$_2$ profile cluster a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. Cluster numbers displayed in legend, with number of profiles in each cluster given in parentheses. Error bars represent 25th and 75th percentile values.

3.9.2: P-3B NO$_2$ Profile Cluster Correlations

Values of R^2 from the simple linear regression analyses between surface mixing ratio and column abundance for each profile cluster during the California and Texas campaigns that presented a statistically significant correlation (i.e., statistically significantly different from 0.0) are summarized in Table 3.15. The remainder of the Texas and California clusters, as well as

Maryland and Colorado clusters, did not produce significant correlations, and are not summarized in the tables. Representative scatter plots are shown in Fig. 3.15a)-d). Again, clusters containing fewer than five members are excluded here. Flynn et al. (2014) demonstrated that, for NO_2, the column_air estimate better represented the "true" lower tropospheric column than did the column_ground estimate during the Maryland campaign; thus, only the column_air correlations will be analyzed here. Much fewer NO_2 cluster correlations were significant than for the O_3 clusters across all campaigns, which is not unexpected given that NO_2 is generally less well mixed than O_3. No cluster correlations were significantly different from the full correlations for the Maryland, Texas, and Colorado campaigns; only the California Cluster 3 correlation was significantly smaller than the full correlation (as it was not a significant correlation).

It is also notable that California Cluster 2 presented a significant, albeit low, correlation, and was not significantly different from the full correlation for California (which was itself statistically significant). The key differences here between California Cluster 2 and Cluster 3 discussed above may lie in the differences in the shapes of the median cluster profiles. California Cluster 2 displayed a more well mixed NO_2 profile, in which NO_2 mixing ratios decreased less sharply in the PBL, than did Cluster 3. In contrast, Cluster 3 displayed a much smoother, exponential decay profile shape, indicative of a less well mixed profile and more stable atmospheric conditions. Cluster 3 also displayed an increase in NO_2 mixing ratios within the lowermost portion of the profile not present in the Cluster 2 profile. This comparison in median profile shapes is consistent with the greater correlation between column and surface data obtained for Cluster 2. However, as Cluster 2 contained the majority of NO_2 profiles during the California campaign, the correlations for this cluster were not statistically significantly different from the full correlation for California. Additionally, the Cluster 4 column_air correlation of the

Texas campaign was significant, contained the vast majority of NO_2 profiles, and displayed the most well mixed median profile shape (NO_2 mixing ratios did not display as great a vertical gradient as for the other clusters). These comparisons further suggest that the vertical distribution, and, by proxy, vertical mixing, has an influence on NO_2 profile variability, as well as photochemical loss of NO_x and production of O_3. Furthermore, as a small number of clusters contained the majority of profiles for each campaign, these results again indicate that NO_2 behavior did not change much over the course of each of these campaigns, further supporting the lack of NO_2 profile variability, and that the ability to relate column observations to surface concentrations depends upon NO_2 exhibiting a more uniform profile vertical gradient in mixing ratio.
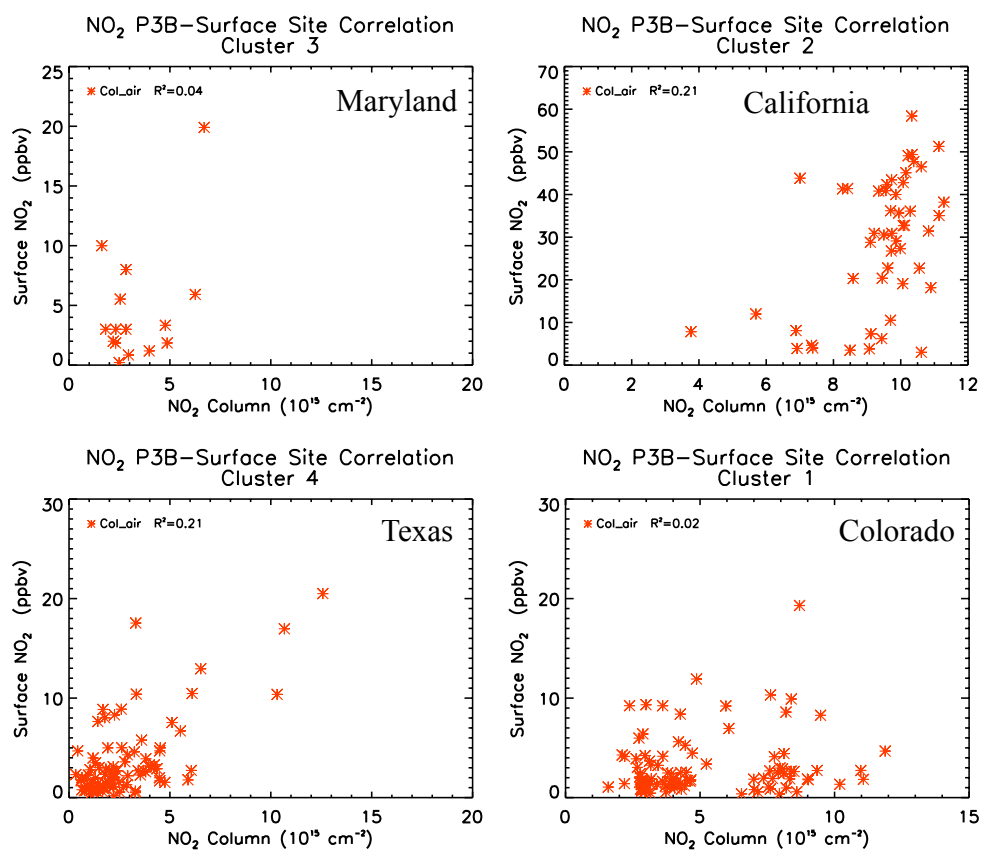
Fig. 3.16: Representative scatter plots for the NO$_2$ column-surface correlations a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. R^2 values for the column_air and column_ground correlations displayed in the legends.

California And Texas Campaign NO$_2$ Correlations	
Cluster Number	Column_Air R^2
California #2	0.21
California Full	0.11
Texas #4	0.21
Texas Full	0.53

Table 15: R^2 values for the correlation between column and surface data for each NO$_2$ profile cluster of the California campaign. No cluster correlations were statistically significantly different from the full correlation.

3.9.3: NO$_2$ Shape Factor Comparisons to CMAQ and GMI

Potential errors in satellite NO$_2$ retrievals are greater when there is bias in the shape factor profile at altitudes where the instrument is more sensitive. Instruments such as OMI are more sensitive to NO$_2$ in the upper PBL and lower free troposphere than near the surface (as determined by profiles of the scattering weights used in the retrieval), so that errors in the shape factors will increase retrieval error should they occur in the upper PBL/lower free troposphere. The median CMAQ NO$_2$ shape factor profiles for each cluster within each campaign compared poorly to the median P-3B shape factor profiles, indicating that CMAQ did not capture the NO$_2$ relative vertical distribution on the CMAQ vertical grid (Fig. 3.16a)-d). While CMAQ often displayed a peak in shape factor values within the PBL when the corresponding P-3B cluster also displayed a peak, the model commonly misplaced the vertical location of that peak, as seen for Colorado Cluster 1 (Fig. 3.16d) and California Cluster 2 (Fig. 3.16b). Additionally, these simulated peaks were often too broad relative to the P-3B shape factor peaks, as seen for California and Texas (Figures 3.16b, c). These peaks often extended through the upper PBL and the lower troposphere, the altitude

regions to which instruments such as OMI are more sensitive for NO_2 than near the surface, and thus where retrieval errors would be greater. The CMAQ shape factors missed the PBL peaks entirely and were too smooth relative to the P-3B for each cluster of the Maryland campaign (Fig. 3.17a). These issues further indicate a general over-mixing within CMAQ, as the NO_2 relative vertical distribution was spread too evenly among the vertical levels relative to the observations.

CMAQ typically struggled to simulate the shape factor values for most clusters during the Maryland, Texas, and Colorado campaigns, with R^2 values below 0.40 and very large values of reduced chi squared, which indicate large over- and underpredictions of the NO_2 shape factor values and that a nonlinear relationship between the simulated and observed NO_2 shape factors exists (Fig. 3.18a)-d). CMAQ struggled most to capture the California shape factors, demonstrating no correlation between the simulated and observed values for any cluster (R^2 values between 0.01 and 0.05); the scatter plots of the simulated shape factor values plotted against the P-3B values demonstrated both over- and underprediction with no consistent bias. These plots also demonstrate that CMAQ often randomly and severely over- or underpredicted relative to the P-3B throughout the depths of the shape factor profiles for the other campaigns. This likely drove down the correlations between simulated and observed values for these campaigns, and further indicates that CMAQ struggled to reproduce the NO_2 vertical distributions under the conditions associated with the California, Texas, and Colorado campaigns. Though CMAQ struggled to capture the shape factor values for the Maryland campaign, CMAQ did perform best relative to the observations for this campaign. In fact, CMAQ compared moderately well to the P-3B for Clusters 4 and 5 (R^2 values of 0.44 and 0.55, respectively). The scatter plots for the Maryland campaign also demonstrate the least scatter of

all four campaigns, as well as consistent, severe overestimation of the NO_2 shape factors above 1200 m (upper PBL/lower free troposphere) and underprediction below this altitude (lower PBL). These patterns of bias are consistent with the patterns of bias in the associated CO shape factor comparison plots (Fig. 3.19), suggesting that overestimated vertical mixing contributes to the errors in the simulated Maryland NO_2 shape factors. This NO_2 comparison, in addition to the comparisons for O_3, further suggests that simulated vertical mixing was most vigorous for the Maryland campaign within CMAQ, and thus likely most erroneous, relative to the other three campaigns. As the greatest correlations between observed and simulated values, and smallest scatter, occurred for the Maryland campaign, this suggests that CMAQ was best able to capture the NO_2 relative vertical distributions under the conditions conducive to O_3 photochemical production and NO_2 photochemical loss that occurred during this campaign, despite the overestimation in simulated vertical mixing that contributed to the errors in the median shape factors for NO_2 for each cluster. Use of CMAQ NO_2 profiles in satellite retrievals over Maryland would produce less error than in the other DISCOVER-AQ deployment regions. It should be especially noted here that the emissions and chemical mechanism errors described by Canty et al., Anderson et al., and Goldberg et al. would also impact the magnitudes of the simulated NO_2 shape factors, as these errors concern overestimates in NO_x emissions as well as inadequate representations of nitrogen chemistry; errors in vertical mixing are one category of several that likely impact the simulated NO_2 shape factor profile.
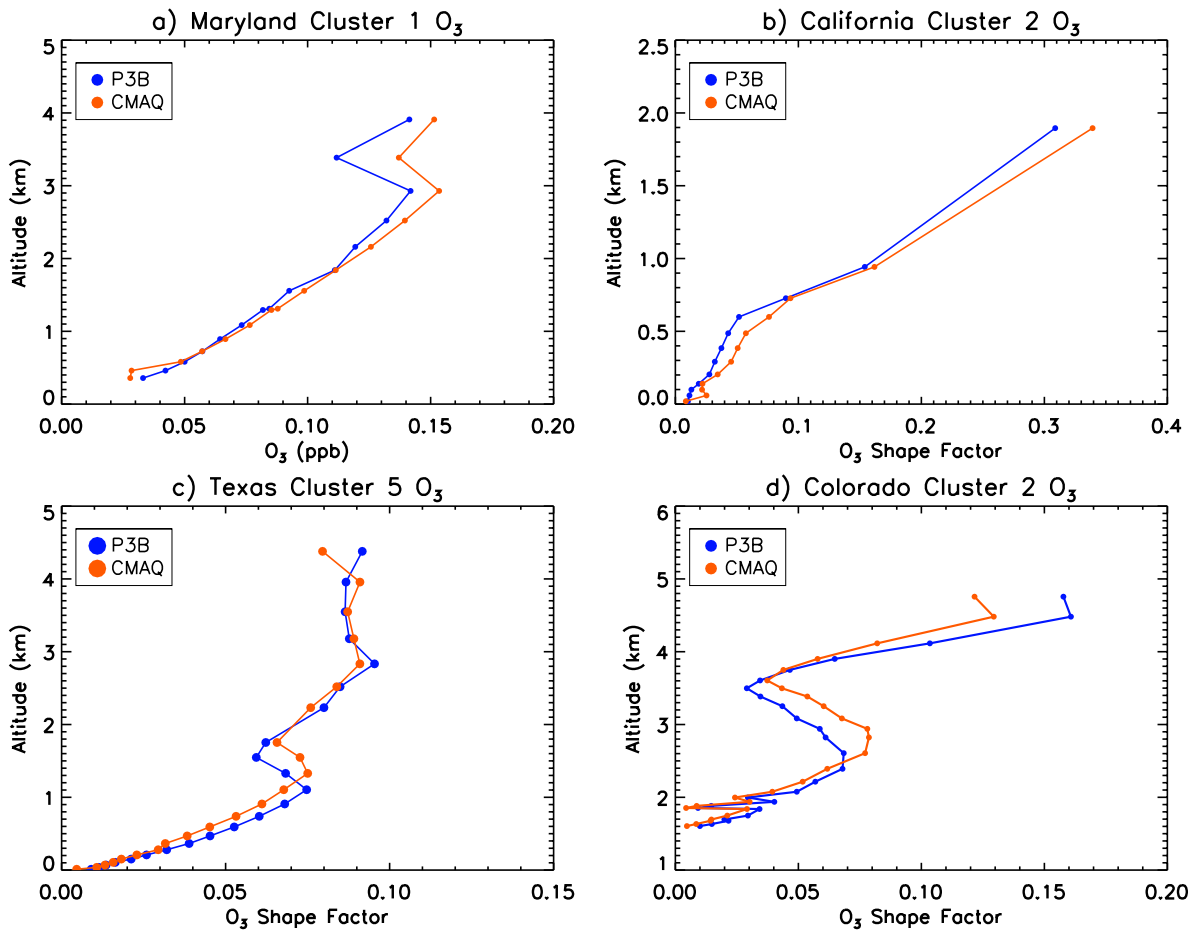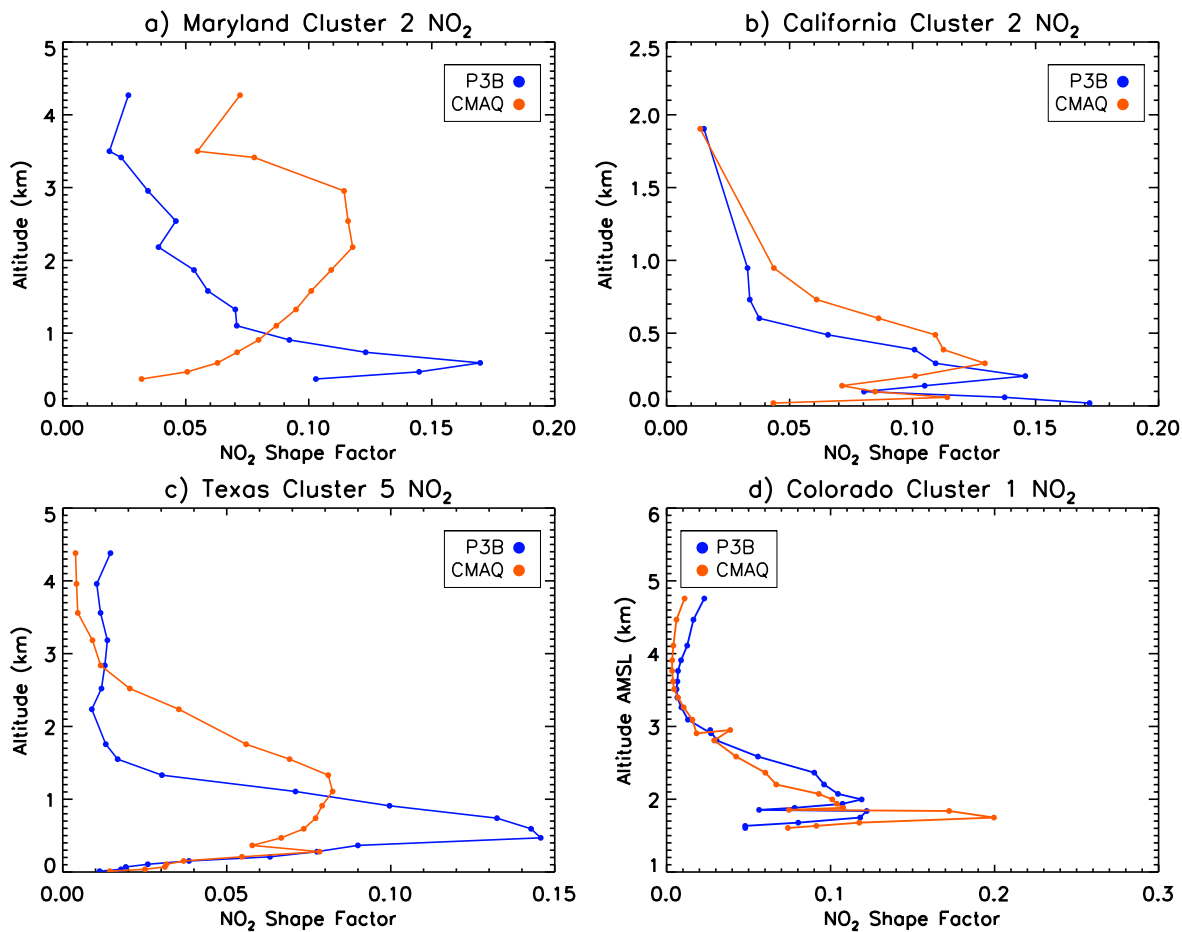
Fig. 3.17: Representative shape factor comparison plots for CMAQ vs. P-3B for NO₂ a) for the Maryland campaign; b) for the California campaign; c) for the Texas campaign; d) for the Colorado campaign. Computed on CMAQ vertical grid.
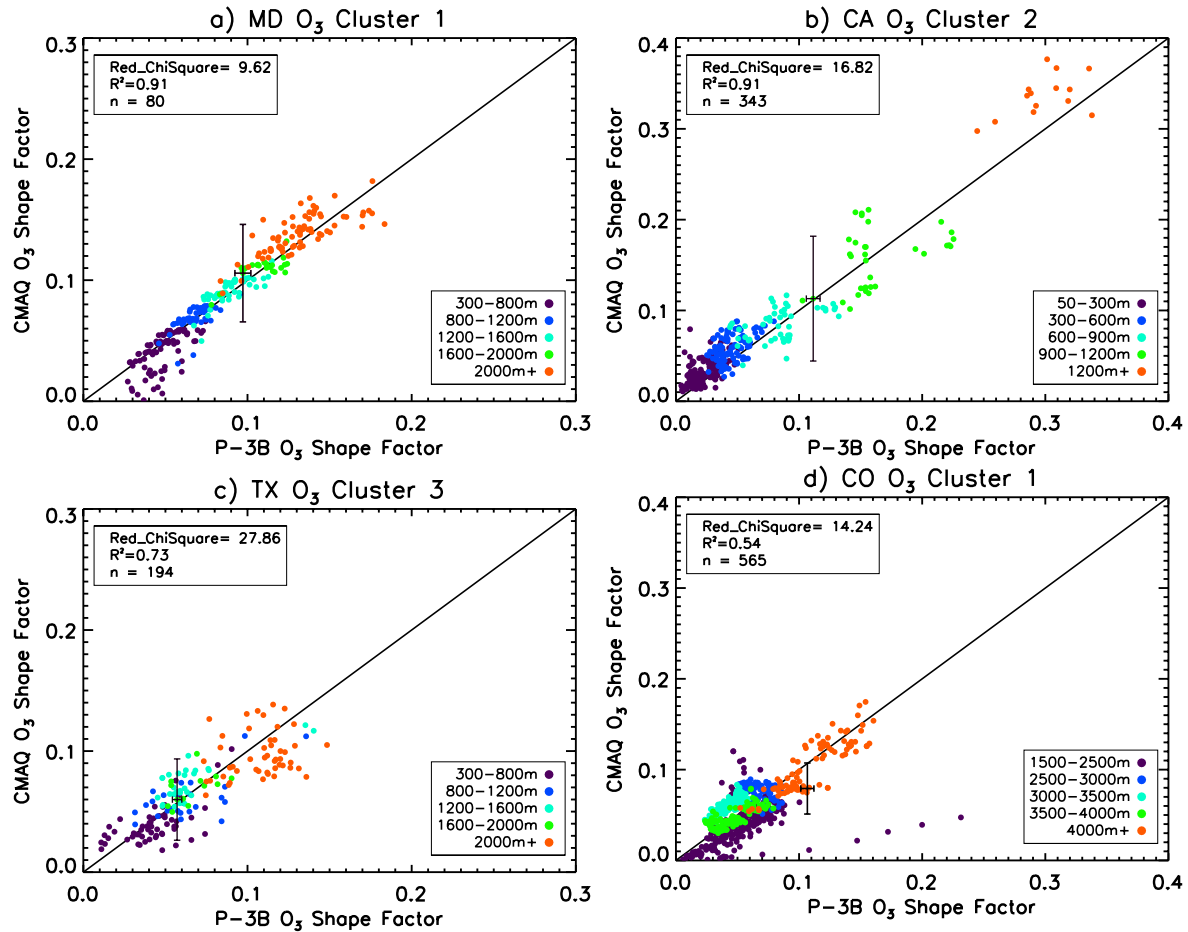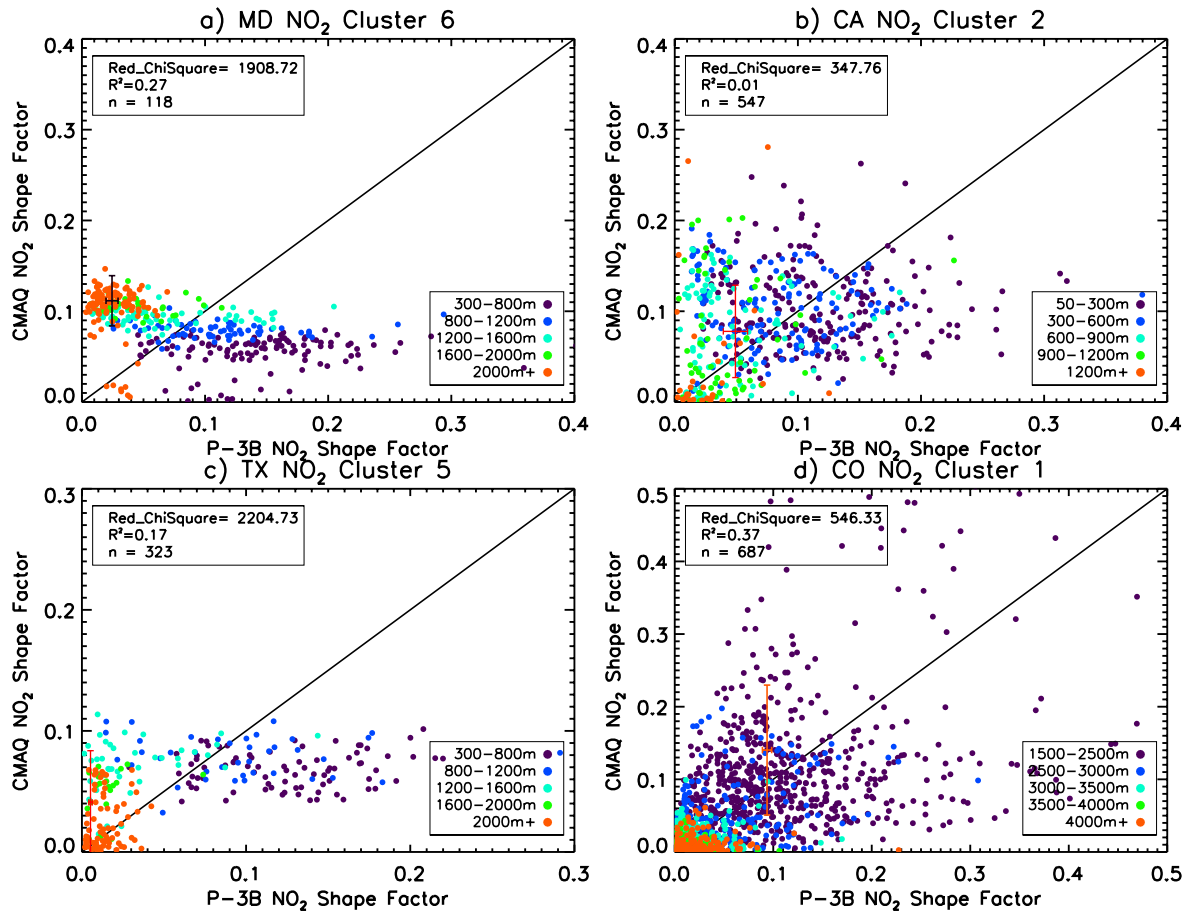
Fig. 3.18: Representative shape factor comparison scatter plots of modeled and observed shape factor values for CMAQ and P-3B for NO₂ a) for the Maryland campaign Cluster 6; b) for the California campaign Cluster 2; c) for the Texas campaign Cluster 5; d) for the Colorado campaign Cluster 1. Computed on CMAQ vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. Example uncertainty bars also displayed in black for one point; uncertainty for P-3B taken from uncertainty of observed shape factors while uncertainty for CMAQ taken as the standard deviation over the simulated shape factors.
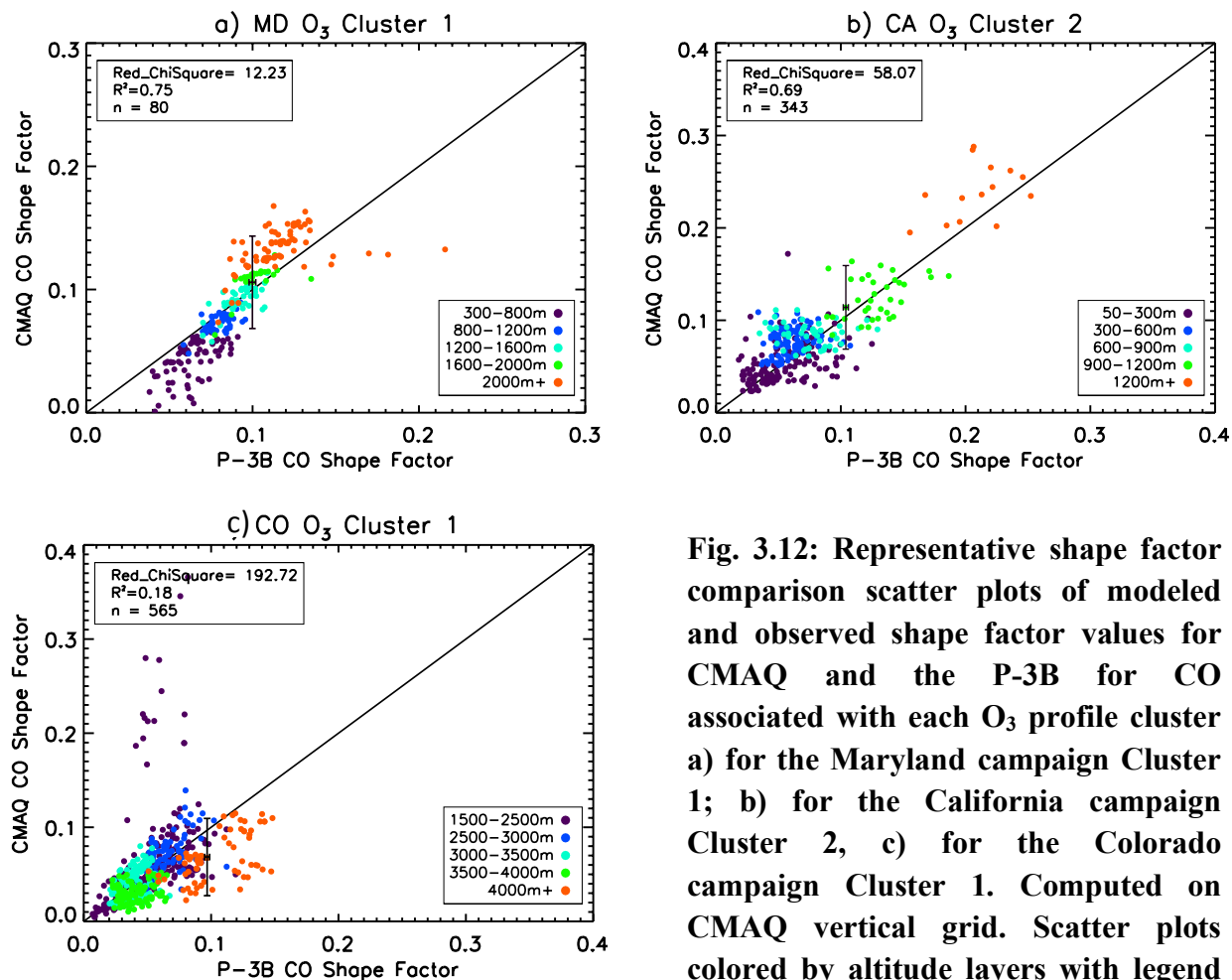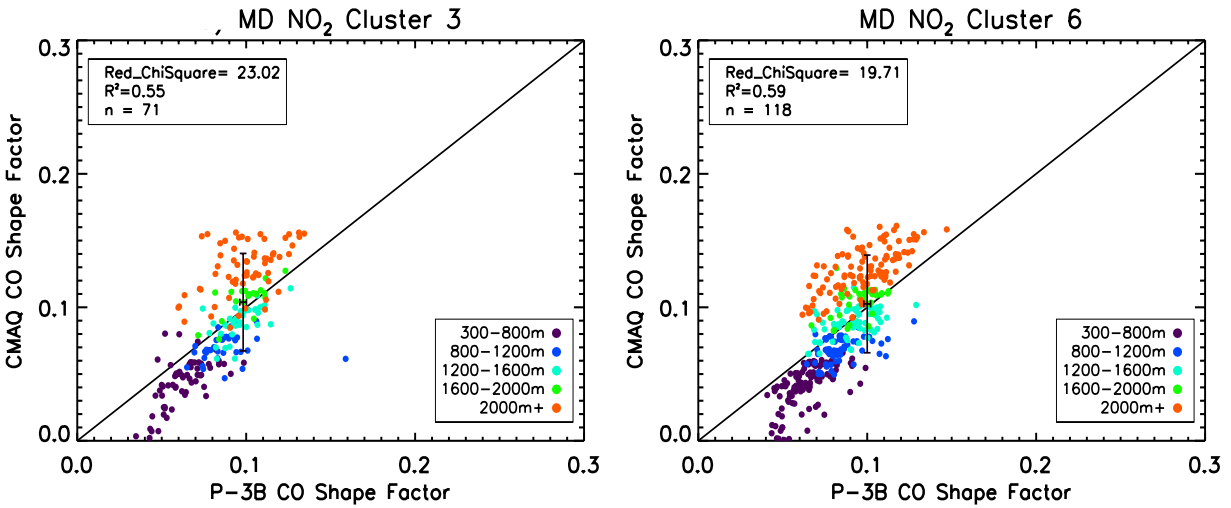
Fig. 3.19: Representative Maryland CO shape factor comparison plots for CMAQ vs. P-3B associated with the NO$_2$ clusters Maryland campaign Cluster 3 and Cluster 6. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. Example uncertainty bars also displayed in black for one point; uncertainty for P-3B taken from uncertainty of observed shape factors while uncertainty for CMAQ taken as the standard deviation over the simulated shape factors.

The GMI NO$_2$ shape factors compared well to the observations for Maryland, but compared poorly for the California campaign (Fig. 3.20a)-b). GMI displayed approximately the same smoothly decaying shape factors as seen for the P-3B during the Maryland campaign. However, GMI often also displayed a high bias in NO$_2$ shape factor values within the upper portion of the PBL, indicating that GMI placed a greater column burden higher in the PBL than was observed. However, during the California campaign, the GMI NO$_2$ shape factors resembled the GMI O$_3$ shape factors much more closely than they did the P-3B NO$_2$ shape factors. As with O$_3$, much of the good comparison between GMI and the P-3B during the Maryland campaign may be due to the model smoothing out the vertical variability when computing the shape factors on the GMI vertical grid. This grid is likely too coarse within the PBL to capture the stagnant conditions in the San Joaquin Valley, and thus to capture the very stable, stratified, and poorly mixed NO$_2$ profile shapes and shape factors. Additionally, these median shape factor profiles

indicate that the NO_2 column burden was distributed too evenly within GMI relative to the observations, as with O_3. Again, the GMI vertical grid likely smoothed out much of the observed NO_2 variability within the PBL, giving rise to a somewhat spurious good comparison between the P-3B and GMI.

GMI poorly captured the shape factor values for all clusters of the California campaign, with R^2 values between 0.10 and 0.30 (Figure 3.20a)-b), and large values of reduced chi squared that indicate errors in the GMI-simulated shape factor values, as with CMAQ for this campaign; GMI compared moderately well for most clusters of the Maryland campaign, with R^2 values above 0.30 (Fig. 3.21). GMI compared exceptionally poorly to the P-3B shape factor values for Maryland Cluster 2 (R^2 = 0.18), but compared exceptionally well for Maryland Clusters 4 and 7 (R^2 = 0.82 and 0.63, respectively). The poor comparison of the GMI shape factor values to the observed California shape factors is particularly evident in the scatter plots, as they display both extreme over- and underprediction for all three clusters; most simulated shape factor values fell between approximately 0.05 and 0.09, regardless of the coincident observed values. This indicates that GMI is both under- and overpredicting. Much of this error in the GMI-simulated NO_2 shape factors may be due to overestimated vertical mixing, as the associated CO shape factor comparison plots display a consistent underpredicted in CO in the lower PBL, with overprediction in the upper PBL/lower free troposphere (Fig. 3.21). This is consistent with the results of the comparisons to the observed O_3 (that vertical mixing within GMI is not as overestimated as within CMAQ), and further highlights the influence of atmospheric stability and mixing on NO_2 mixing ratio profiles and shape factors. The scatter plots for the Maryland clusters indicate that a few extreme over- and underpredicted values typically undermined the correlation between the simulated values and the observed shape factor values. The CO shape

factor comparison plots for Maryland do not indicate the same errors in vertical mixing as for California. These results suggest that GMI performs better for the conditions associated with the Maryland campaign, similarly to CMAQ, in which NO_2 was more rapidly converted to O_3, than for the California campaign, where NO_2 concentrations were able to build up and the PBL was stably stratified. Though the correlations between simulated and observed shape factor values were similar between CMAQ and GMI, GMI produced somewhat higher correlations likely due to the difference in vertical resolution within these models. The coarser GMI vertical grid within the lower troposphere, compared to the CMAQ vertical grid, likely smoothed out much of the observed NO_2 variability within the PBL, giving rise to a somewhat spurious good comparison between the P-3B and GMI. However, the finer vertical resolution further supports the superiority of a regional model such as CMAQ for use within remote sensing retrievals. The relative performances of GMI and CMAQ for the NO_2 shape factors reveals in starker relief the impact on model vertical resolution on the simulated shape factors under various conditions.
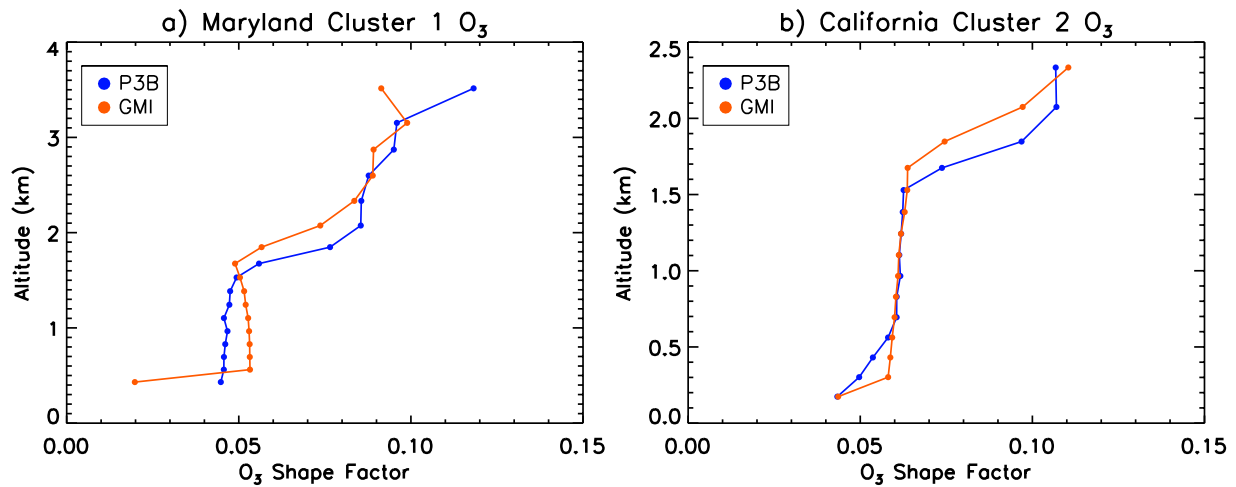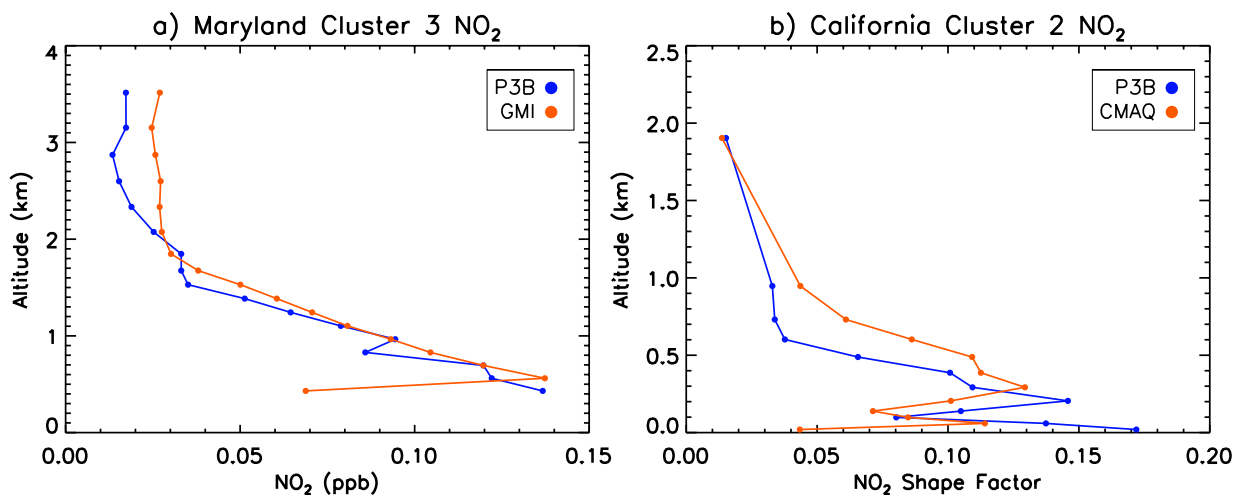

Fig. 3.20: Representative shape factor comparison plots for GMI vs. P-3B for NO_2 a) for the Maryland campaign; b) for the California campaign. Computed on GMI vertical grid.

a) MD NO$_2$ Cluster 4

b) CA NO$_2$ Cluster 3

CA NO$_2$ Cluster 3
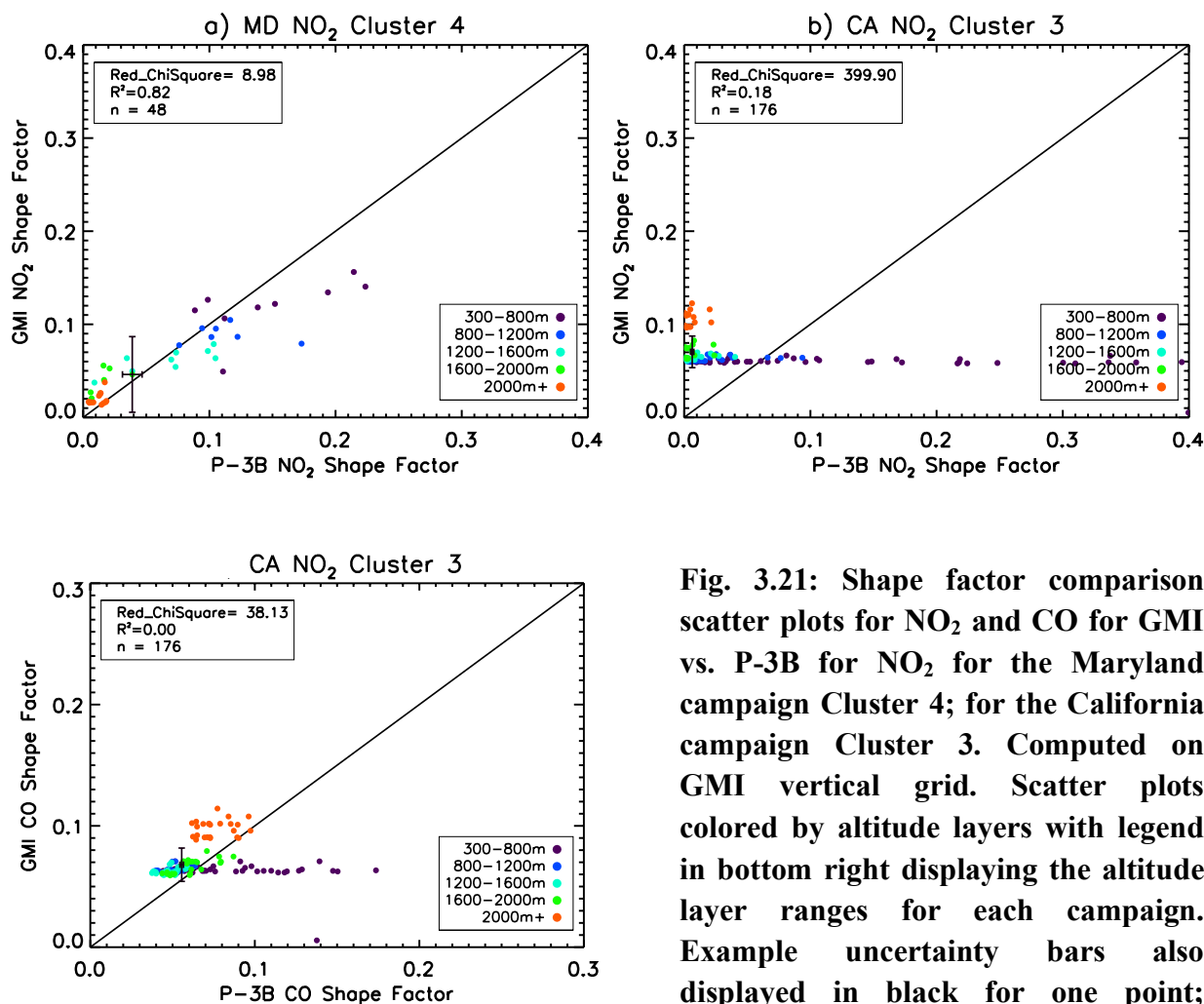
Fig. 3.21: Shape factor comparison scatter plots for NO$_2$ and CO for GMI vs. P-3B for NO$_2$ for the Maryland campaign Cluster 4; for the California campaign Cluster 3. Computed on GMI vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. Example uncertainty bars also displayed in black for one point; uncertainty for P3B taken from uncertainty of observed shape factors while uncertainty for GMI taken as the standard deviation over the simulated shape factors.

3.10: Conclusions

All *in situ* O$_3$ profile clusters produced by the agglomerative hierarchical cluster analysis were significant for the California, Texas, and Colorado DISCOVER-AQ deployments, with Texas producing the greatest number of distinct clusters (five). Only one distinct cluster (Cluster

4) emerged for the Maryland deployment. The season in which the campaign took place, air mass history as delineated by HSYPLIT, and the relative degrees of vertical mixing and horizontal mixing, drove the differences in profile variability, as denoted by the number of significant clusters that emerged, among campaigns. Further, atmospheric stability, as indicated by the lapse rate and potential temperature profiles, played a key role in modulating the differences in median O_3 profile shape among clusters within each campaign, while time period during the campaign, spiral sampling time of day, and airmass origin influenced the relative magnitudes of the clusters. In contrast, very few *in situ* NO_2 profile clusters were distinct for each campaign, and those that were distinct contained very few profiles, indicating that NO_2 profile behavior remained relatively uniform throughout the course of each campaign; the interplay of NO_2 photochemical loss and vertical mixing likely played the key role in determination of the amount of NO_2 profile variability.

When significant, the correlations between the corresponding P-3B column and surface data for each cluster were generally larger for O_3 than NO_2, which is not surprising given that O_3 is more well mixed in the vertical and horizontal than NO_2. However, many cluster correlations and correlations over the full set of profiles used in the clustering analysis were not significant for both O_3 and NO_2, suggesting that, even if the model profiles used in a satellite retrieval algorithm were to correctly represent the actual lower tropospheric profiles for these trace gases, meteorological or chemical conditions may prevent a strong column-surface connection. For O_3, column observations may be most representative of surface concentrations under the conditions of deep, convective boundary layers, reduced wind shear, and few terrain influences associated with the Maryland deployment. The degree of vertical mixing was also found to have an influence on the ability to relate column and surface quantities for NO_2, consistent with the

results of Chapter 2, as greater column-surface correlation was found for clusters with more well mixed median NO_2 cluster profiles than for those with less well mixed median profiles. However, a typically low degree of correlation was associated with these well mixed median cluster profiles, indicating that accurate representation of the lower tropospheric NO_2 profile in a satellite retrieval does not guarantee the ability to connect column and surface.

Lastly, the regional CMAQ model and global GMI model simulated lower tropospheric O_3 shape factors compared moderately well to the P-3B shape factors corresponding to most clusters for the Maryland and California campaigns. These results suggest that a regional air quality model that captures the observed shape factor variability would be potentially useful in remote sensing retrievals for O_3. However, the effects of vertical grid resolution within both models are more pronounced on the NO_2 shape factors than for O_3, likely because O_3 is generally more well mixed than the short-lived NO_2. In the case of NO_2, the regional model (with finer vertical resolution) produced more realistic profiles of shape factors than the global model. However, both models performed better relative to the observations for the Maryland campaign than for any other campaign. These results demonstrate that models may be best able to capture trace gas profiles under the conditions of convective boundary layers and O_3 photochemical formation associated with the Maryland campaign. These results also demonstrate the necessity of sufficient vertical resolution within the PBL, where the majority of the NO_2 burden resides.

Chapter 4: Diurnal Cycles of O_3 and NO_2 Column Amounts and Surface Mixing Ratios in Observations and Model Output during DISCOVER-AQ

4.1: Introduction

Current satellite air quality instruments are flown onboard polar-orbiting low earth orbiting (LEO) satellites, such as the Ozone Monitoring Instrument (OMI) onboard the Aura satellite. These LEO satellites generally observe the entire globe during a 24-hour period, providing observations for regions lacking other observation networks, and at varying horizontal resolutions. However, the temporal resolution of these satellites is quite low, with one or, at most, two passes over any given surface site during daylight hours from any one satellite. However, some recent works have taken advantage of observations from multiple satellite instruments or over multiple years to elucidate temporal variability of satellite column measurements. Boersma et al. (2009) examined tropospheric NO_2 columns from the SCIAMACHY instrument onboard Envisat (overpass ~10:00 LT) and the OMI instrument (overpass ~13:30 LT) over Israeli cities during 2006. This work found a diurnal cycle in NO_2 columns during summer months as shown by higher morning columns measured by SCIAMACHY than the afternoon OMI columns; the opposite diurnal pattern was found for winter months. These comparisons indicate that the diurnal cycle of column NO_2 is controlled by the diurnal cycle of anthropogenic NO_x emissions and loss of NO_2 through reaction with OH, while the seasonal variation in the diurnal cycle between summer and winter can be explained through the seasonal variation in photochemistry (OH concentrations are greater in summer than winter due to greater water vapor concentrations and greater UV flux, leading to greater magnitudes in the chemical loss rate of NO_2 with respect to OH in summer than winter and a larger diurnal cycle in this loss rate in summer than winter). A similar diurnal variation in NO_2 tropospheric column amounts was found over southeastern Europe for 1996-2001 based on

examination of GOME, GOME-2, SCIAMACHY and OMI observations, when these observations were available (Zyrichidou et al., 2009). Lamsal et al. (2008) also found that *in situ* surface NO_2 measurements at OMI overpass time were 36% smaller than annual 24-hour concentrations, again suggesting a diurnal cycle in NO_2 abundances. However, despite these cases, the limited LEO overpasses restrict the analysis of diurnal variability and evolution of O_3 and NO_2 column amounts throughout the day.

Some of the limitations of LEO satellites will be resolved through the development of satellites orbiting in geostationary earth orbit (GEO). Upcoming geostationary satellite missions highlight the importance of understanding the diurnal variability of column abundances. TEMPO (part of GeoCAPE) will provide hourly observations of species such as O_3, NO_2, CH_2O, SO_2, and aerosol optical depth (AOD) during all daylight hours, allowing them to capture the diurnal variability of column amounts (Chance et al., 2012; Fishman et al., 2012). If an IR instrument is funded as part of GeoCAPE or otherwise, the diurnal variability of CO could be observed. This temporal resolution is much greater than for any current LEO instrument. The geostationary air quality satellite constellation (TEMPO over North America, Sentinel-4 over Europe, and GEMS over East Asia) will also cover much of the Northern Hemisphere, including data-sparse regions, at finer horizontal resolution than the current LEO instruments. Thus, geostationary satellites will open a new opportunity to observe column abundances of key trace gases at high temporal and spatial resolutions, to better understand pollutant abundances over the Northern Hemisphere as well as changes in pollutant abundance during daylight hours (Chance et al., 2012; Fishman et al., 2012).

Of particular interest is the possibility to characterize the diurnal variability of near-surface or surface pollution conditions from geostationary observations (Fishman et al., 2008;

Martin, 2008). This may be especially useful for regions lacking sufficient surface air quality monitors. Furthermore, if such observations well capture near-surface or surface pollutant variability, then satellite observations may well correlate with or be easily transformed into surface pollutant concentrations, providing a measure of surface air quality directly relevant to policy makers and the public. As shown in Chapters 2 and 3, lower tropospheric column amounts correlate well with surface mixing ratios under certain meteorological or chemical conditions, supporting this potential use for geostationary observations. Thus, understanding of the diurnal behavior of O_3 and NO_2 lower or full tropospheric column amounts, as well as of surface concentrations, may indicate when satellite observations may be most representative of surface conditions. Insights into how the column-surface relationship changes throughout daylight hours would also be provided, giving an indication of when satellites best represent surface air quality.

By extension, understanding how well satellite observations capture surface variability necessitates understanding how well the assumed pollutant profiles used in the satellite retrieval process capture both observed column and surface variability, as well as profile shape. The simulated pollutant profiles are currently provided by global chemical transport models, such as GMI. The ability of these models to capture observed column magnitudes and variability, as well as observed surface variability, directly impacts the ability of satellite retrievals to capture such observations and their relevance to surface and near-surface quantities. Comparison of the performance of the global model with a regional air quality model, such as CMAQ, will also indicate if the use of a regional model within retrievals may provide an advantage. Thus, analysis of the simulated column diurnal cycle for O_3 and NO_2 provides an assessment of the models used in retrievals, and insight in addition to the correlation analyses and profile cluster analyses

presented in previous chapter into how well the assumptions inherent in these simulations impact the ability of satellites to capture observed variability.

This chapter focuses on analysis of the diurnal behavior of O_3 and NO_2 column amounts from the P-3B aircraft (lower tropospheric columns) and from remote sensing instrumentation (full tropospheric column amounts) at each of the spiral sites from each of the four DISCOVER-AQ deployments. Several remote sensing column datasets are available: Pandora spectrometers were available at all spiral sites during all four campaigns, observing O_3 and NO_2 columns; the Airborne Compact Atmospheric Mapper instrument (ACAM; Lamsal et al., 2016) was flown onboard the UC-12 aircraft during the Maryland campaign, and measured O_3 and NO_2 column amounts; and the Geostationary Trace gas and Aerosol Sensor Optimization instrument (GeoTASO; Nowlan et al., 2016) was flown onboard the NASA Falcon aircraft during the Texas campaign, and measured NO_2 column amounts. These column diurnal cycles will be compared to each other, to determine if the remote sensing observations capture the lower tropospheric variability, as well as to the diurnal behavior of O_3 and NO_2 surface mixing ratios, to determine which column type, if any, captures the surface diurnal variability. Lastly, the observed column diurnal time series will be compared to the simulated column diurnal time series from the CMAQ and GMI models, to assess the performances of these regional and global air quality models. Such comparisons among observational data sets and to different models across all campaigns will provide insight into which times of day, which types of column abundance (lower or full, observed or simulated), and under which meteorological and pollution conditions column variability best captures surface variability. Thus, the work presented here will build upon that presented in Chapters 2 and 3, and provide insight into the column-surface correlations presented previously.

4.2: Description of Observational Column Datasets

4.2.1: P-3B Column Air and Column Ground Lower Tropospheric Columns

The P-3B column_air and column_ground O_3 and column_air NO_2 values were computed as described previously in Chapter 2, Section 2.2.1, and Chapter 3, Section 3.3, for all four DISCOVER-AQ campaigns. It is again noted that the P-3B actively avoided clouds during flights, and flight days were chosen to minimize cloud cover. Column_ground NO_2 is not discussed, as results from Chapter 2 indicate that column_air more accurately represents the true NO_2 lower tropospheric column amounts.

Measurements of the *in situ* NO_2 photolysis frequency j(NO_2) for the photochemical loss of NO_2 from the reaction

$$NO_2 + h\nu \rightarrow NO + O(^3P) \hspace{4cm} \text{(Eqn. 4.1)}$$
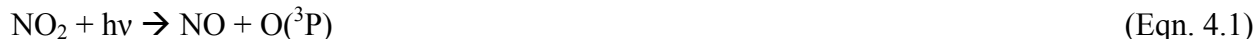
were also available for the P-3B spirals. J(NO_2) itself is defined as the integral of the product of the actinic flux I, the NO_2 absorption cross section σ, and the NO_2 photolysis quantum yield φ over the 280-420 nm wavelength (λ) range. These data were provided by the NSERC P-3B data acquisition and distribution system as nadir and zenith measurements, so these two values were summed to obtain the total photolysis frequency. Lastly, the total j(NO_2) values were averaged over the time of each P-3B spiral, to obtain an average value that corresponded to each P-3B column abundance.

4.2.2: P-3B Col_Air_Sonde Lower Tropospheric Columns

Measurements of the lower tropospheric NO_2 column were also computed from tethersonde measurements at the Huron, CA, and Smith Point, TX, sites (col_air_sonde). Tethersonde NO_2 data from Edgewood were found to be of insufficient quality (due to the extensive and random presence of negative mixing ratio values), and the data from Golden, CO,

are not yet available. These column amounts were computed analogously to the column_air computation, except that the NO_2 tethersonde profile, rather than the NO_2 measurement at the lowest aircraft measurement level, was used to the fill the gap between the P-3B spiral and ground level. The NO_2 tethersonde profiles were observed from the ground to approximately 500 m AGL. The Huron or Smith Point P-3B and tethersonde profiles were first co-located in time, requiring the two types of profiles to be coincident within 30 minutes of each other. The P-3B and tethersonde profiles were then combined, and the lower tropospheric column computed. This column type represents an alternative to the column_air and column_ground abundances, using other profile data rather than P-3B or surface measurements to fill the gap between the P-3B and the ground.

4.2.3: Ozonesonde Full Tropospheric Columns

Ozonesondes were launched at the Beltsville, MD, Edgewood, MD, Porterville, CA, Smith Point, TX, Moody Tower, TX, and the Platteville, CO, sites, covering the full depth of troposphere and much of the stratosphere. The ozonesonde full tropospheric O_3 column amounts were thus computed for these sites. The tropopause was placed at the first altitude above 9 km where the temperature lapse rate was positive, such that data within the stratosphere were excluded before computing the ozonesonde column.

4.2.4: Pandora Full Tropospheric Columns

The ground-based Pandora UV/Vis spectrometers (Herman et al., 2009) were located at each spiral site for the four campaigns. Pandora full tropospheric column amounts for O_3 and NO_2 were computed for each campaign as described previously in Chapter 2, Section 2.2.2.

4.2.5: ACAM and GeoTASO Full Tropospheric Columns

The Airborne Compact Atmospheric Mapper instrument (ACAM; Lamsal et al., 2016; Liu et al., 2015) and the Geostationary Trace gas and Aerosol Sensor Optimization instrument (GeoTASO; Nowlan et al., 2016) NO_2 vertical column densities were provided in the DISCOVER-AQ data archive as tropospheric vertical column abundances; the ACAM O_3 column densities were provided as total vertical column abundances. For both the ACAM and GeoTASO retrievals, the slant columns were first retrieved through fitting a modeled spectrum to the observed radiance spectrum. Slant columns were converted into vertical columns through use of the air mass factor (AMF), as defined in Chapter 3. Scattering weights were provided by the VLIDORT radiative transfer model, while the *a priori* O_3 or NO_2 vertical profiles used to compute the shape factors for the AMF computation were provided by the CMAQ model, at 4 km horizontal resolution. The use of a regional air quality model rather than a global air quality model such as GMI in both retrievals represents a departure from satellite retrievals, which employ global models. ACAM data are currently available only for the Maryland campaign, while GeoTASO data are available for the Texas campaign. Uncertainty for the ACAM NO_2 columns were approximately 20-30%, while uncertainty for the GeoTASO column was more variable (between approximately 10-50%). To obtain tropospheric ACAM O_3 vertical columns, the OMI stratospheric column was subtracted from the ACAM total column; uncertainty in the tropospheric columns were approximately 2-10% (Lamsal et al., 2016; Liu et al., 2015; Nowlan et al., 2016).

4.2.6: OMI Full Tropospheric Columns

Tropospheric columns from OMI were retrieved with the Version 2.1 Goddard tropospheric NO_2 retrieval algorithm (Buscela et al., 2013) and the ozone profile algorithm by Liu et al. (2010) with modifications as described in Kim et al. (2013). These data were screened

for cloud fraction (effective cloud fraction less than 30%, as denoted by the "CRF" flag provided in the data files), the instrument row anomaly (screened by Lok Lamsal and Xiong Liu during creation of the OMI data files), and distance from the surface site (pixel center less than 100 km distant; "distance" flag in the data files) for each campaign. The OMI tropospheric O_3 column was typically approximately twice as large as P-3B O_3 column_air and column_ground, indicating a significant tropospheric burden above the tops of the P-3B spirals. The bias relative to P-3B NO_2 column_air and column_ground varied more than for O_3, though the OMI NO_2 tropospheric column was typically approximately 10-40% larger than the P-3B column amounts, indicating some tropospheric NO_2 burden not measured by the P-3B.

4.3: Description of Surface Volume Mixing Ratio Datasets

The available surface O_3 and NO_2 mixing ratio datasets available for each campaign are detailed in Chapter 2, Section 2.3, and Chapter 3, Section 3.4. However, only those sites for which photolytic or corrected molybdenum converter NO_2 surface measurements were used in the comparison between campaign-average surface and column diurnal timeseries, to avoid comparison to surface NO_y observations. These sites include Edgewood, Essex, Fair Hill, and Padonia, MD; Bakersfield, Fresno, Huron, and Porterville, CA; Galveston, Moody Tower, and Manvel Croix, TX; and the BAO Tower, Chatfield Park, Fort Collins, and Golden, CO, spiral sites.

4.4: Description of CMAQ and GMI Tropospheric Column Computations

4.4.1: Loughner et al. and NOAA CMAQ Simulations

The Loughner et al. CMAQ simulations and NOAA CMAQ forecasts used in the analyses presented below have been described previously in Chapter 2, Section 2.4.1, Chapter 2, Section 2.4.2, Chapter 3, Section 3.5.1, and Chapter 3, Section 3.5.2.

4.4.2: NASA GMI Simulations

The GMI simulations used in the analyses below have been described previously in Chapter 3, Section 3.5.3.

4.4.3: Model Tropospheric Column Amounts

4.4.3.1: CMAQ Partial Tropospheric Column Amounts

For computation of model biases relative to the P-3B lower tropospheric column amounts, CMAQ O_3 and NO_2 partial tropospheric columns were computed hourly for each campaign over the depth of the P-3B spirals. The procedure to compute these CMAQ partial columns was previously detailed in Chapter 2, Section 2.4.3, and was applied to the CMAQ output for all four campaigns.

4.4.3.2: CMAQ and GMI Full Tropospheric Column Amounts

For computation of model biases relative to the observational full tropospheric column data, and for use in the time series analysis, CMAQ O_3 and NO_2 columns were computed hourly over the full depth of the model troposphere. Model layers within the stratosphere were first excluded before computation of the column amounts for both models. The first CMAQ layer within the stratosphere was taken to be the first layer above 8 km where the O_3 mixing ratio exceeds 100 ppbv. These stratospheric layers were then excluded before integration of the CMAQ simulated profile to obtain the full tropospheric CMAQ column amounts.

The GMI column output files (the "column" files) provided column amounts at 12 Z directly, computed over the full depth of the troposphere only. The GMI profile output (the "daily" and "overpass" files, provided at 8 AM, 12 PM, and 3 PM EDT for July 2011, and at 11 AM and 2 PM PDT for winter 2013) included the simulated tropopause heights, so that layers

above the tropopause were excluded before computation of the full tropospheric column from these simulated profiles.

4.5: Comparison of O_3 Column and Surface Mixing Ratio Diurnal Variability

4.5.1: Observed O_3 Column vs. Surface Mixing Ratio Diurnal Variability

All four campaigns exhibited similar campaign-average diurnal variability of the hourly mean surface O_3 mixing ratio values, as expected. Representative surface and column diurnal time series are displayed in Fig. 4.1a)-d). The O_3 mixing ratios were lowest during early morning, with mixing ratio values quickly increasing throughout the morning, as expected as O_3 formation increases with increasing sunlight and O_3 layers aloft are mixed down as the PBL grows. Typically, though not always, mixing ratio values increased from approximately 10 ppbv during early morning to approximately 60 ppbv by local noon during the Maryland, Texas, and Colorado campaigns. Maximum surface values were reached during the afternoon hours of 12 PM – 3PM local time during each campaign, when both the height of PBL development and abundant sunlight favored photochemical production of O_3. During the Maryland, Texas, and Colorado campaigns, O_3 mixing ratios began to decrease between approximately 4 PM – 6 PM local time, as the boundary layer began to collapse; the California campaign did not exhibit this collapse, likely because the flight day ended before collapse of the boundary layer and sundown. Unsurprisingly, the wintertime California deployment exhibited the smallest diurnal range of mixing ratio values (blue lines in Fig. 4.1). Ozone mixing ratios grew from typically approximately 10 ppbv during early morning to approximately 40 ppbv during afternoon at most spiral sites, due to inhibited O_3 formation encountered in the San Joaquin Valley during winter, relative to the diurnal cycles of the other three deployments. However, despite inhibited formation, the morning minimum and afternoon maximum in surface mixing ratios seen for the

other campaigns is clearly apparent at the California spiral sites, further supporting that surface O$_3$ displays a consistent diurnal cycle in a wide range of pollution and meteorological conditions.
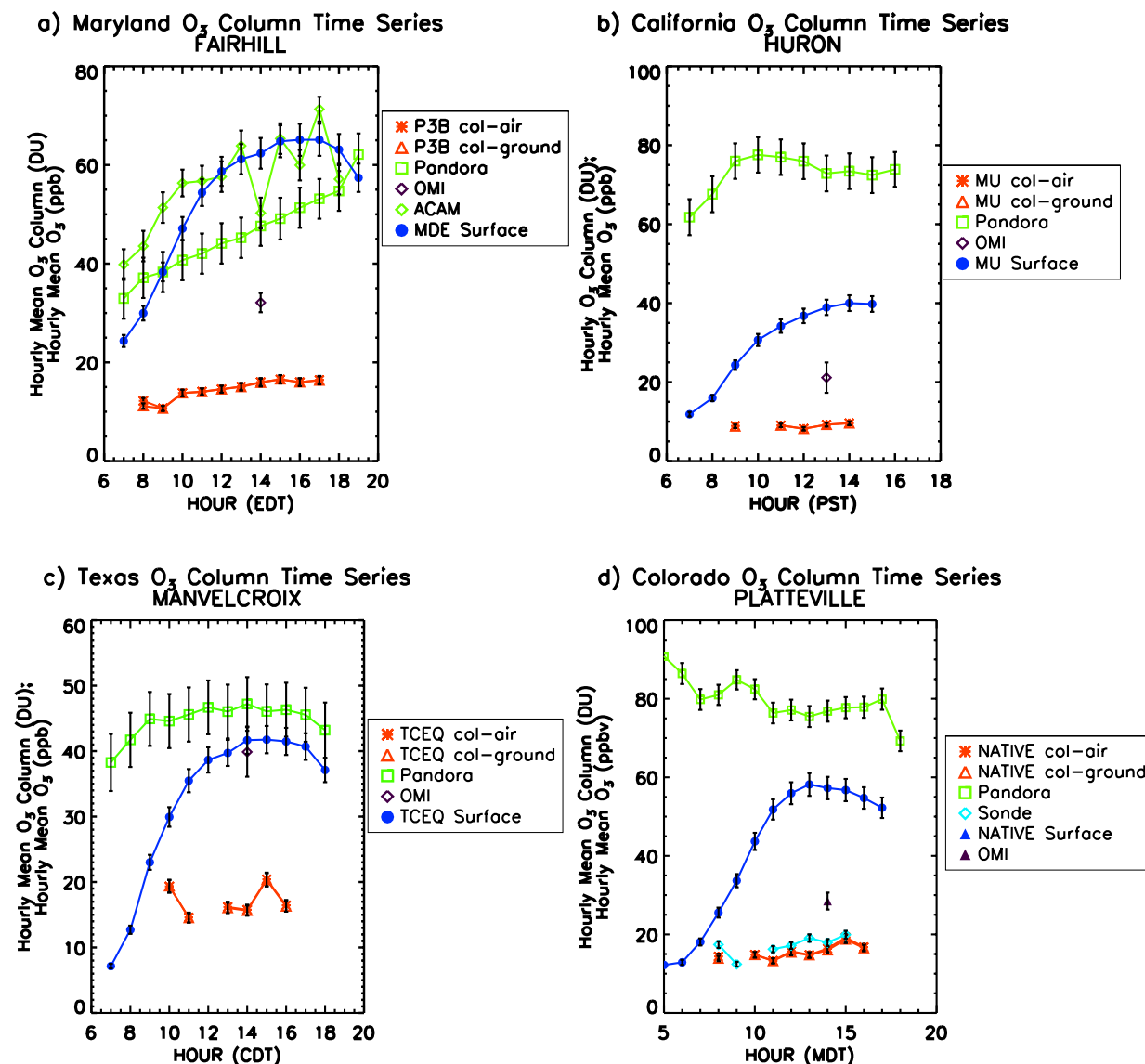
Fig. 4.1: Representative campaign average diurnal timeseries for O$_3$ surface mixing ratios (ppbv) and column abundances (DU) for a) the Maryland campaign, b) the California campaign, c) the Texas campaign, and d) the Colorado campaign. Surface and column values plotted over daylight hours for each campaign.

In contrast to the surface datasets, the P-3B O$_3$ column_air and column_ground lower tropospheric columns, and ozonesonde full tropospheric columns, demonstrated very little diurnal variability during all four deployments, and remained relatively constant in value throughout the

137

day. Additional representative scatter plots are displayed in Fig. 4.2 to highlight the variability of the different column datasets. Column_air and column_ground values typically remained at or below 20 Dobson units (DU) for all campaigns. The ozonesonde columns typically remained between 20 DU and 40 DU for the Maryland and California deployments, while the ozonesonde columns typically remained at approximately 20 DU or below during the Texas and Colorado deployments. It is notable that the ozonesonde and P-3B column magnitudes were similar for these latter two deployments, and even more so that the Platteville, CO, ozonesonde column diurnal behavior very closely resembled that of P-3B column_air and column_ground. Further, it is surprising that the P-3B lower tropospheric and ozonesonde full tropospheric column diurnal cycles were not coupled with the surface diurnal cycles during any campaign, given that these column data were computed from *in situ* data, were integrated from vertical levels near the surface in the case of the ozonesondes, and that the P-3B column datasets represent lower tropospheric column amounts. These results are particularly unexpected for the Maryland campaign, which presented large correlation between O_3 column and surface data for each spiral site and most profiles clusters (Chapters 2 and 3).

The ACAM columns for Maryland (Fig. 4.1a and 4.2a) displayed some diurnal behavior, though the diurnal cycle was much smaller in amplitude than for the surface data. Though less apparent than for the surface data, the ACAM O_3 column burden increased during the morning (approximately 7 AM to 12 PM EDT). This is reminiscent of the ramp up in surface mixing ratio values at all six spiral sites. However, the morning ramp up for ACAM generally occurred over less time than did the surface mixing ratios: maximum column values were often reached by 10 AM EDT, rather than early afternoon. The ACAM columns typically remained between approximately 50 DU and 60 DU during the afternoon, and did not exhibit clear indications of

decrease in magnitude as the PBL began to collapse, as the surface data did, indicating plentiful ozone above the top of the PBL in the late afternoon. The ACAM dataset displayed a damped diurnal variation, with its small morning ramp up in column abundance to maximum values. However, the Pandora tropospheric O_3 columns most often failed to display diurnal variation across all four campaigns (Figs. 4.1 and 4.2), consistent with the lack of P-3B column diurnal variability. Some spiral sites exhibited a damped diurnal variation in Pandora ozone column, similar to ACAM, such as Fair Hill, MD, Galveston, TX, and Huron, CA, in which the column abundance increased at these sites from a minimum value during early morning until approximately 10 AM standard or daylight saving time, when maximum values were reached. Pandora also displayed the largest range of magnitudes of any O_3 column dataset, from approximately 40 DU to 80 DU; typically, the California and Colorado campaigns demonstrated larger Pandora values than the other campaigns. The magnitudes of the campaign-average OMI tropospheric column were often between those of the P-3B and ozonesonde column amounts and between the Pandora and ACAM column amounts. Like the *in situ* columns, the remotely sensed ACAM and Pandora column diurnal variation demonstrated little to no coupling with the surface diurnal variation. Indeed, the Pandora columns during the Colorado deployment often exhibited very different diurnal behavior from the P-3B columns, ozonesonde columns, and surface data, such as seen for Platteville, CO.
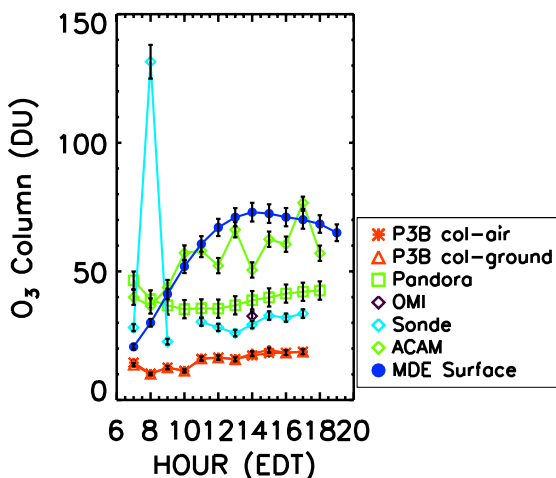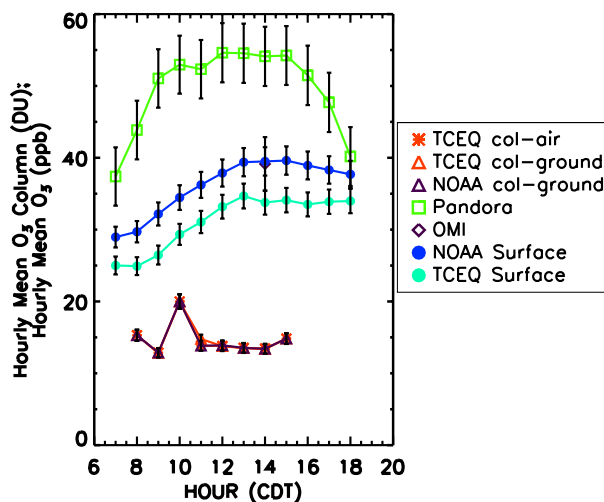
Fig. 4.2: Representative campaign average diurnal time series highlighting the column diurnal variation for the ozonesonde, ACAM, and Pandora column amounts for (left) Edgewood, MD, and (right) Galveston, TX.

The diurnal variability in surface mixing ratios is driven primarily by photochemistry. However, column variability (or lack thereof) may be driven primarily by PBL dynamics. During nighttime, after the convective PBL has collapsed to the shallow, stable boundary layer, O_3 within the PBL is titrated by NO_x to low concentrations. However, collapse of the convective PBL leads to formation not only of the stable layer, but also of a residual layer that exists immediately above the stable layer, which consists of air that had been turbulently mixed the previous day during the height of the CBL. As the NO_x within the residual layer had been depleted by the previous day's photochemistry, O_3 levels may persist within this layer through the night, to be mixed with the stable layer air as the PBL grows again during the next day. Additionally, surface O_3 concentrations do not tend to dramatically differ from day to day over the course of a 30-day deployment, with exception for events such as cold frontal passage or the ozone buildup associated with a pollution episode. Thus, the O_3 within a column of air does not vary much over the course of a day on average, as the column includes both photochemically produced surface O_3

140

and residual layer O_3 that are mixed as the PBL develops into the CBL during daytime. This is consistent with the lack of variability in the lower tropospheric column abundances. The lack of diurnal variability in the full tropospheric column abundances further suggests that column variability is controlled by PBL dynamics, as much O_3 within the free troposphere is O_3 that had been vented from the PBL and transported from its source region, in addition to freshly produced O_3 pollution. The mixing of surface and residual layer air also explains the large correlation between O_3 column and surface quantities for some campaigns, despite the lack of connection between column and surface diurnal cycles. Lastly, due to the lack of column diurnal variation, these results do not suggest a time of day when satellite observations may be most representative of surface O_3 conditions.

4.5.2: O_3 Observed vs. Simulated Column Diurnal Variability

The CMAQ and GMI O_3 full tropospheric columns displayed very little diurnal variability across all four campaigns, consistent with the observational column datasets. Representative diurnal time series are displayed in Fig. 4.3a)-d). In fact, the simulated columns displayed very little deviation in magnitude throughout most daylight hours at each spiral site for each campaign. The diurnal variability of the simulated column amounts best matches that of the P-3B lower tropospheric columns and ozonesonde columns for both models rather than the ACAM or Pandora columns, though the P-3B columns demonstrated somewhat more variation in column magnitude throughout the day than did CMAQ or GMI. Thus, though the models overestimated the lack of diurnal variability, CMAQ and GMI adequately simulated the diurnal cycle of O_3 column amounts.

The CMAQ full tropospheric columns were greatly overpredicted relative to the ozonesonde column estimates for the Maryland and Texas campaigns, as denoted by the positive

median model bias value and the time series plots (bias defined as the percent difference of the model hourly-mean column value relative to the coincident hourly-mean observational value) amounts, displayed in Fig. 4.4a)-d); the simulated and ozonesonde columns were integrated to approximately the same altitude. CMAQ overestimated relative to the ozonesondes for the California and Colorado campaigns as well, but not as greatly as for the other two campaigns. The CMAQ full tropospheric columns were underestimated relative to Pandora during the California and Colorado campaigns, while CMAQ compared well to the Pandora columns during the Texas campaign and somewhat overestimated during the Maryland campaign; CMAQ slightly underestimated relative to ACAM during the Maryland campaign. The CMAQ partial tropospheric column amounts, computed over the depth of the P-3B spirals only, compared well to P-3B column_air and column_ground for all four campaigns (the Col_Air and Col_Grd boxes in Fig. 4.4). This result is encouraging that CMAQ may adequately capture lower tropospheric column burdens. Further, with the exception of the median biases relative to the ozonesondes, these results suggest that CMAQ was most able to capture the observed O_3 columns for the conditions associated with the Maryland and Texas campaigns, which included deep, convective boundary layers and adequate temperatures and sunlight for O_3 formation.

Fig 4.3: Representative campaign average diurnal timeseries for observational and simulated O_3 column abundances (DU) for a) the Maryland campaign, b) the California campaign, c) the Texas campaign, and d) the Colorado campaign.

Fig. 4.4: Box-and-whisker plots of model O$_3$ bias median values and distributions for a) the Maryland campaign, b) the California campaign, c) the Texas campaign, and d) the Colorado campaign. Model bias computed as percent difference relative to each observational dataset. Tops and bottoms of boxes represent the 75th and 25th percentile bias values, respectively, while whiskers represent outliers.

4.6: Comparison of NO$_2$ Column and Surface Mixing Ratio Diurnal Variability

4.6.1: Observed NO$_2$ Column vs. Surface Mixing Ratio Diurnal Variability

Surface NO$_2$ mixing ratios exhibited a general decrease from maximum values (due to shallow PBL, morning traffic emissions, and the nighttime build up of NO$_2$ due to lack of photolysis) during the early morning hours to minimum values by approximately 12 PM to 1 PM standard or daylight saving time, and remained at minimum values through most of the remainder of the day during all four campaigns, as displayed in the representative timeseries in Fig. 4.5a)-d). Some sites during these campaigns exhibited a slight increase in NO$_2$ mixing ratios beginning at

approximately 6 PM standard or daylight saving time, indicative of the collapse of the CBL, titration of O_3 to produce NO_2, and evening traffic emissions. Maximum values were typically between 5 ppbv and 10 ppbv, decreasing to less than 5 ppbv during the day, for the Maryland and Texas campaigns, while maximum values for the California campaign were often approximately 20 ppbv, decreasing to approximately 10 ppbv. This is again indicative of the inhibited O_3 formation (cold temperatures and less sunlight available for photolysis relative to the other campaigns) and shallow PBL encountered during the California campaign. Surface magnitudes were smallest during the Colorado campaign, and typically remained between approximately 1 ppbv and 4 ppbv. The campaign-average diurnal variation in NO_2 is consistent with that of O_3 discussed above: NO_2 levels are greatest during early morning, when NO_2 is produced from NO_x emissions and O_3 formation is inhibited, with NO_2 levels decreasing as it is photochemically converted to O_3 as the sun rises and temperatures increase during the day. The Colorado diurnal cycle plots, as they begin at an earlier than the other campaigns, clearly demonstrate the build up of NO_2 as rush hour begins, followed by decreases in NO_2 as it is converted to O_3 during late morning and afternoon.
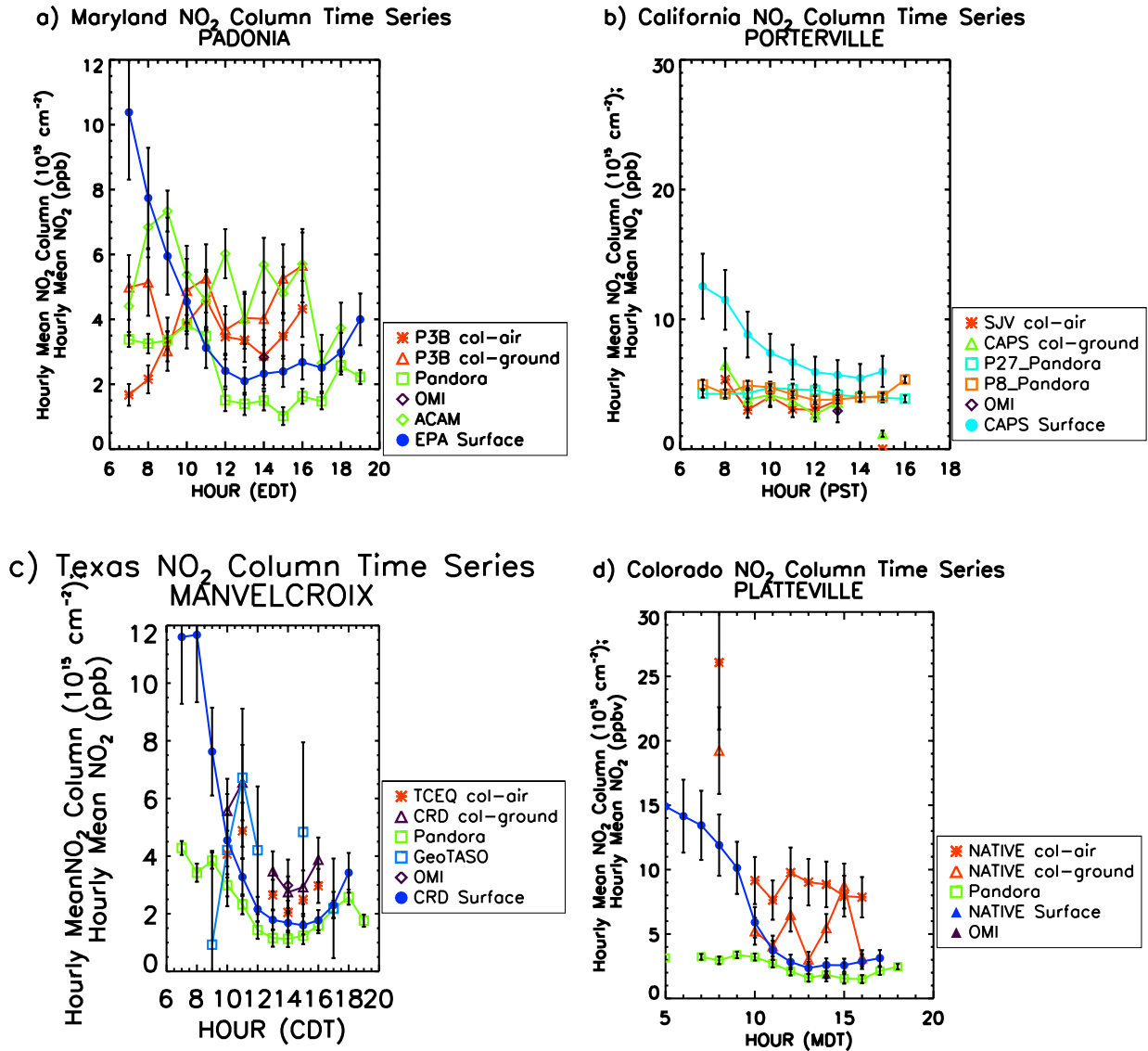
Fig. 4.5: Representative campaign average diurnal timeseries for NO₂ surface mixing ratios (ppbv) and column abundances (10^{15} cm^{-2}) for a) the Maryland campaign, b) the California campaign, c) the Texas campaign, and d) the Colorado campaign. Surface and column values plotted over daylight hours for each campaign.

Unlike O_3, the NO_2 column datasets presented mixed results for diurnal variability. Additional plots highlighting col_air_sonde and the remotely sensed column abundances are displayed in Fig. 4.6. P-3B column_air exhibited diurnal variation, though this variation did not always follow that of the surface, and was damped relative to the amplitude of the surface

variation. The column_air diurnal cycle was least consistent during the Colorado campaign, during which the Chatfield Park and Golden spiral sites displayed some diurnal variation while the BAO Tower and Fort Collins sites did not. During the Maryland, California, and Texas campaigns, and at the Chatfield Park and Golden sites, the column_air diurnal cycle typically followed that of the surface, but offset by several hours from the surface cycle; peak and minimum columns amounts typically occurred later than those for the surface mixing ratios. Not surprisingly, it took some time for surface-based emissions to mix upward and have significant effects on the column amount. It is notable that the Fair Hill, MD, column_air cycle does not closely resemble that of the surface until after 4 PM EDT, when the column_air curve suggests an increase in column abundance coincident with an increase in surface mixing ratio. Column_air abundances generally remained at or below approximately 6×10^{15} cm^{-2} during the Maryland deployment, at or below approximately 10×10^{15} cm^{-2} during the Texas and Colorado deployments, and between approximately 2×10^{15} cm^{-2} and 20×10^{15} cm^{-2} during the California deployment; maximum column values for California are again consistent with the inhibited O_3 photochemistry during this campaign. The diurnal variation of the col_air_sonde column amounts at the Huron, CA, and Smith Point, TX, spiral sites is in agreement with that of column_air: col_air_sonde displayed a small cycle somewhat offset in time that followed that of the surface. The magnitudes of col_air_sonde were similar to column_air throughout the day at Huron, and were slightly larger than column_air at Smith Point, suggesting further that column_air is representative of the true NO_2 lower tropospheric column.
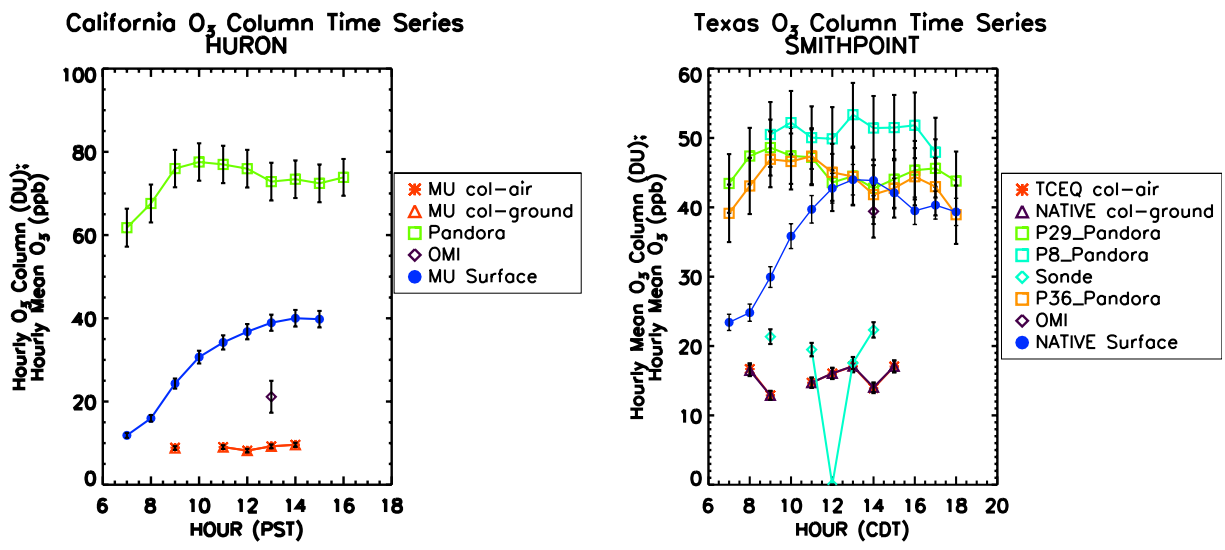
Fig. 4.6: Representative campaign average diurnal time series highlighting the column diurnal variation for the col_air_sonde, ACAM, GeoTASO, and Pandora column amounts for (left) Huron, CA, and (right) Smith Point, TX.

The ACAM NO_2 full tropospheric column amounts for the Maryland campaign and the GeoTASO column amounts for the Texas campaign displayed clear diurnal variation at most spiral sites. As with column_air and col_air_sonde, these cycles were offset in time from the surface cycle. ACAM exhibited a peak in column abundance between 9 AM and 10 AM EDT, as the surface cycle had already begun to decay, with another, smaller peak at approximately 3 PM EDT. GeoTASO displayed peak column amounts between 10 AM and 12 PM CDT, with another, smaller peak occurring at 3 PM CDT or later, when data were available. Thus, both aircraft-based remotely sensed column abundance datasets for NO_2 exhibited similar diurnal cycles to each other; these datasets were also somewhat offset from the P-3B diurnal cycles. The Pandora columns, however, presented a damped diurnal variation at most sites during the Maryland, California, and Texas campaigns, while during the Colorado campaign, the Pandora data exhibited a clear diurnal cycle that was similar to those for ACAM and GeoTASO at each site. Again, the Pandora cycle tended to be offset from that of the surface, as well as column_air and col_air_sonde. Remotely sensed column amounts and those based on *in situ* data appear to have

somewhat similar diurnal cycles, but offset in time from each other and the surface. The ACAM, GeoTASO, and Pandora campaign-average hourly-mean column magnitudes fell within the same ranges as for column_air and col_air_sonde.

The surface NO_2 diurnal cycle is controlled by NO_x emissions, particularly from mobile sources such as automobiles, and photochemical loss of NO_2 (Eqn. 4.1), the rate-limiting step in the O_3 production mechanism, as described in Chapter 1. In accordance with the established literature, maximum surface NO_2 mixing ratio values occurred during early morning, due to a combination of NO_x emissions during the morning rush hour and the nighttime buildup of NO_2 from lack of photolytic destruction; increases in surface NO_2 were again encountered during late afternoon at many spiral sites, as evening rush hour NO_x emissions commenced. Minimum surface values occurred during afternoon, when sufficient sunlight was available to drive O_3 photochemical production and NO_2 photolytic loss.

As the majority of the NO_2 column resides in the lower troposphere (Chapter 1), emissions and photochemistry should also play important roles in the regulation of the NO_2 column diurnal variability, in addition to the surface diurnal cycle. This is consistent with the late morning maxima in column abundances followed by afternoon minima, when a diurnal cycle could be discerned, during each campaign. As NO_2 is produced near the surface, it is also mixed up into the column through turbulent mixing within the PBL, as the convective PBL, or CBL, develops during morning. The column_ground diurnal cycles at most sites particularly reflected the influence of emissions, with larger early morning peaks than other column datasets, due to the incorporation of a surface concentration in this column computation. Comparisons of the column diurnal variation to the diurnal variation of j(NO_2) suggest the influence of photochemical loss on the NO_2 columns as sufficient sunlight becomes available to drive such

reactions, particularly during the Maryland and Texas campaigns; representative plots are displayed in Fig. 4.7. The campaign-average diurnal variation of $j(NO_2)$ during each campaign typically displayed larger values during late morning and early afternoon, and smaller values during early morning and late afternoon, following the times of day of greatest sunlight, as expected. Again, the P-3B actively avoided cloudy conditions during flights as much as possible. Decreases in column abundance occurred as the $j(NO_2)$ hourly mean values increased. Some sites, such as Galveston (Fig. 4.7c) and Golden (Fig. 4.7d) displayed immediate responses in column abundances to changes in $j(NO_2)$; for example, a sudden decrease in $j(NO_2)$ occurred at 12 PM CDT at Galveston, indicating a greater campaign-average cloud fraction at this site, that coincided with increased in column_air, column_ground, and GeoTASO column abundances due to decreased photolytic loss. The California campaign demonstrated a weaker connection between the diurnal cycles of $j(NO_2)$ and column abundances than did the other three campaigns, likely because $j(NO_2)$ values were often not as large as for the other campaigns (larger $j(NO_2)$ values were encountered over a shorter time span for the California spiral sites as seen in Fig. 4.7b, relative to the broader peaks displayed in the time series for the Maryland or Texas sites, for example). Thus, $j(NO_2)$ demonstrated less effect on NO_2 column variability during the California deployment relative to other influences, such as emissions.

However, PBL mixing likely contributed to the offset in time of the diurnal variations of the column and surface data for each campaign; boundary layer mixing delays the growth and decay of the NO_2 column relative to the surface. Though turbulent mixing communicates surface and near-surface NO_2 to the column above that surface site, as mixing from the surface into the PBL takes time, the column may continue to grow during late morning even as the surface abundance decays. During late morning and afternoon, sufficient sunlight is available for

photochemistry that NO_2 within the column also decays, and the column amounts begin to decrease. Just as importantly, this offset in diurnal variation between column and surface amounts may also explain the poor correlations obtained between column and surface data for each campaign: the column abundance may be more representative of surface concentrations at a previous time, rather than surface concentrations coincident to that column observation.
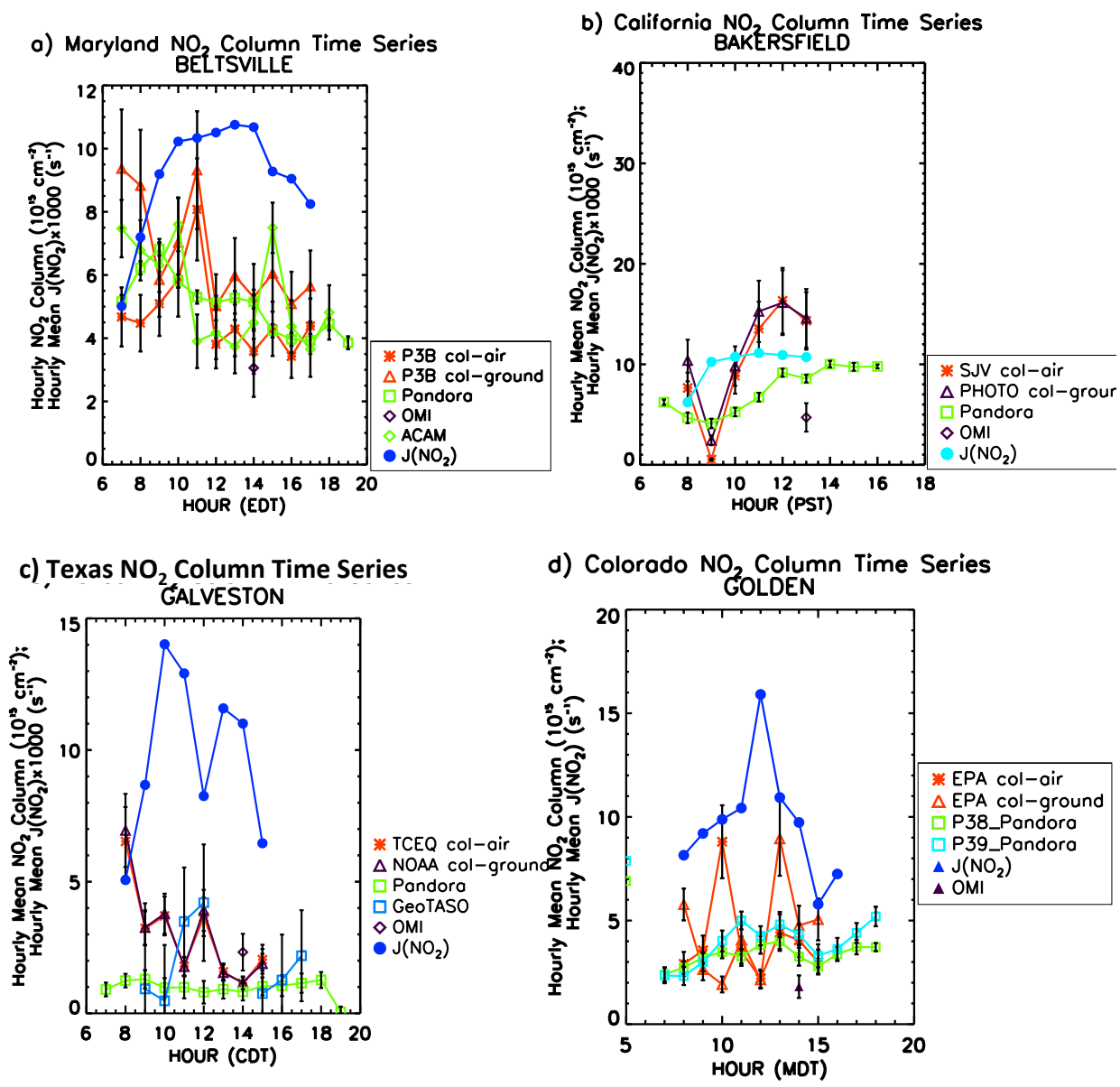
Fig 4.7: Representative campaign average diurnal timeseries for observational column abundances (10^15 cm^-2) and photolysis frequency j(NO_2) (s^-1) for a) the Maryland campaign, b) the California campaign, c) the Texas campaign, and d) the Colorado campaign.

4.6.2: NO_2 Observed vs. Simulated Column Diurnal Variability

The CMAQ and GMI simulated full tropospheric NO_2 column amounts exhibited clear diurnal variability at most sites during all four campaigns. Representative diurnal timeseries comparing the observational and simulated column amounts are displayed in Fig. 4.8a)-d). During

the Maryland, California, and Texas campaigns, CMAQ exhibited maximum column abundances during morning hours, decreasing during late morning to reach minimum values during afternoon, much as the observational column datasets. However, the CMAQ column variability during the Colorado campaign (Fig. 4.8d) is unlike that of the observations or any other campaign; the cycles at each Colorado spiral site also often did not resemble each other. For example, at Fort Collins (Fig. 4.8d), CMAQ exhibited a minimum in column abundance between 10 AM and 12 PM MDT, with maxima during early morning and mid-afternoon, while CMAQ exhibited a minimum during early morning at Golden, with column abundance increasing throughout the day. Though much fewer GMI column amounts were available, these simulated diurnal cycles also exhibited a small decrease between the early morning GMI column value and the afternoon GMI column values during the Maryland campaign, when both morning and afternoon column values were available; only afternoon columns were available for the California campaign, limiting analysis of the GMI column diurnal variability. It is encouraging that the CMAQ and GMI diurnal cycles resemble each other, indicating that both the regional and global models qualitatively captured the observed column variability.
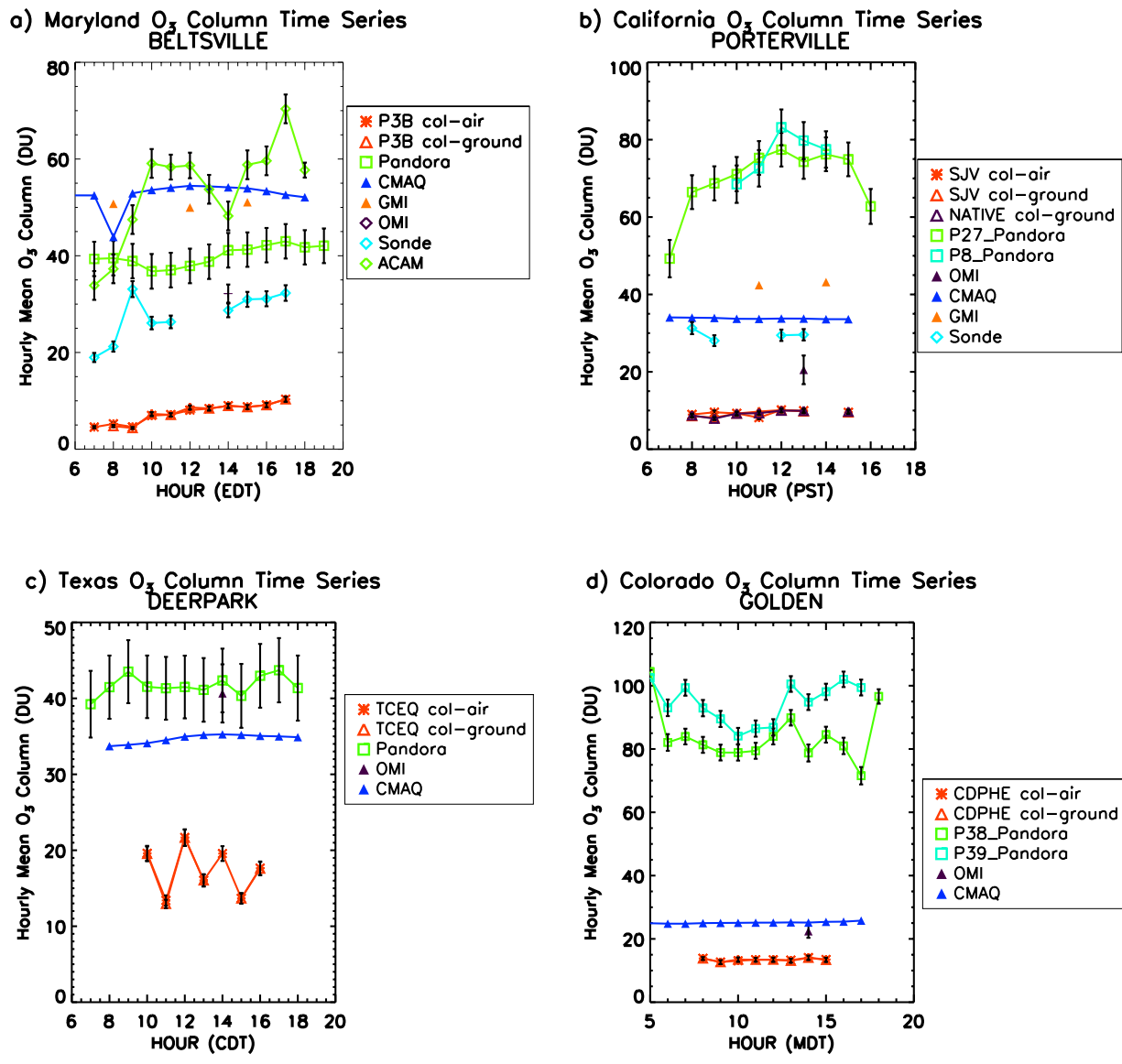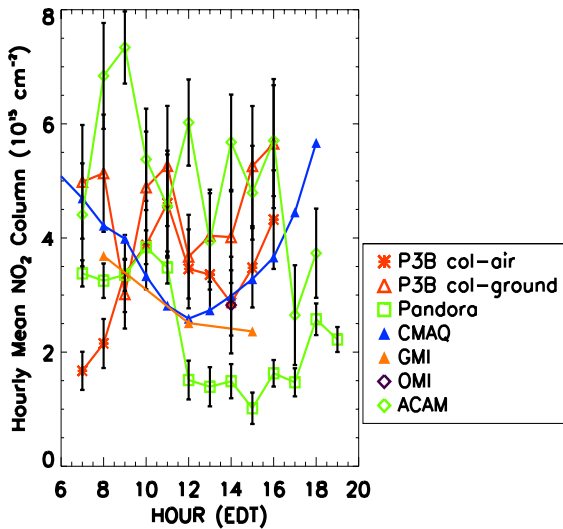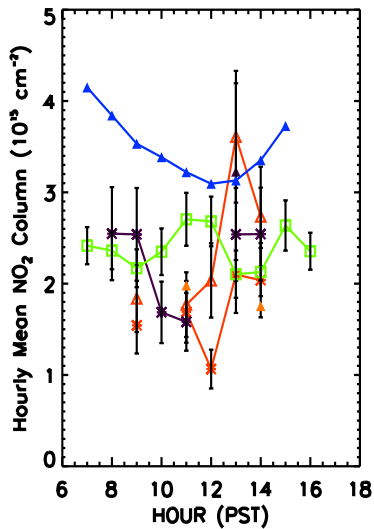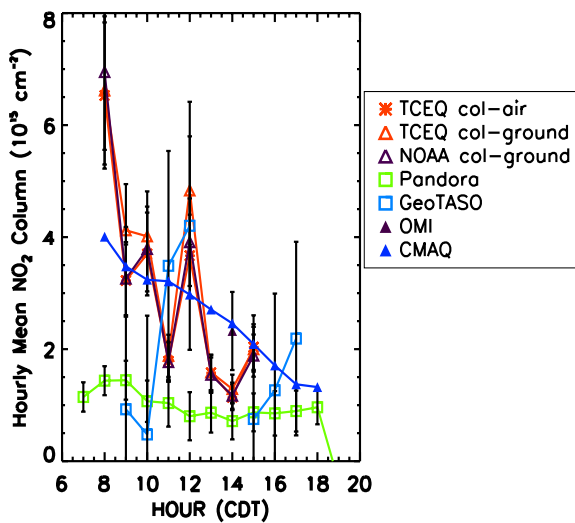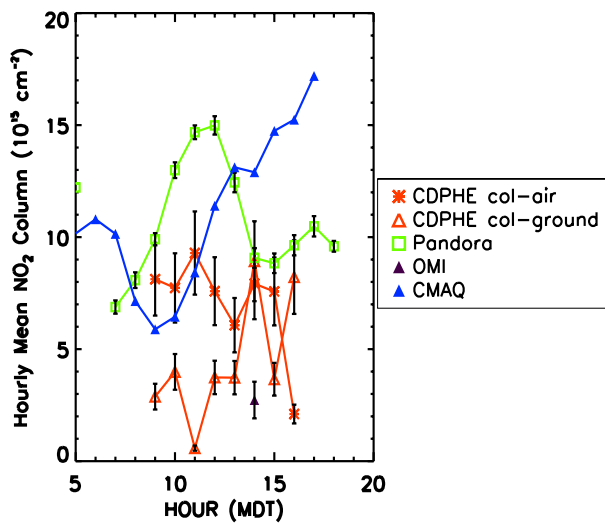
Fig 4.8: Representative campaign average diurnal timeseries for observational and simulated NO₂ column abundances (10^{15} cm⁻²) for a) the Maryland campaign, b) the California campaign, c) the Texas campaign, and d) the Colorado campaign.

CMAQ full tropospheric columns typically displayed the greatest campaign-median model bias values relative to the observational column datasets for the California campaign, and compared well to most datasets for the Maryland, Texas, and Colorado campaigns. Bias values

are given as Box-and-whisker plots displaying the campaign-median biases relative to each observational dataset are displayed in Fig. 4.9a)-d). CMAQ tended to somewhat overpredict relative to Pandora during all campaigns, and to GeoTASO during the Texas campaign, in the median, though CMAQ demonstrated no preferred bias direction relative to ACAM during the Maryland deployment. The CMAQ lower tropospheric column amounts were also slightly overpredicted relative to column_air for these three campaigns, and slightly underpredicted relative to col_air_sonde during the Texas campaign in the median. Model bias values of 50% difference or greater were obtained for the California campaign, which was the campaign that experienced the greatest amount of NO_2 pollution. Thus, CMAQ more adequately captured NO_2 column magnitudes for the campaigns that experienced significant O_3 production, rather than the campaign that experienced inhibited O_3 formation. However, the Texas campaign presented the smallest model biases across most observational datasets, with the smallest range of model bias values (as denoted by the box and whisker plots), indicating that CMAQ performed best in terms of column magnitude under the complex chemical conditions associated with this campaign for NO_2.
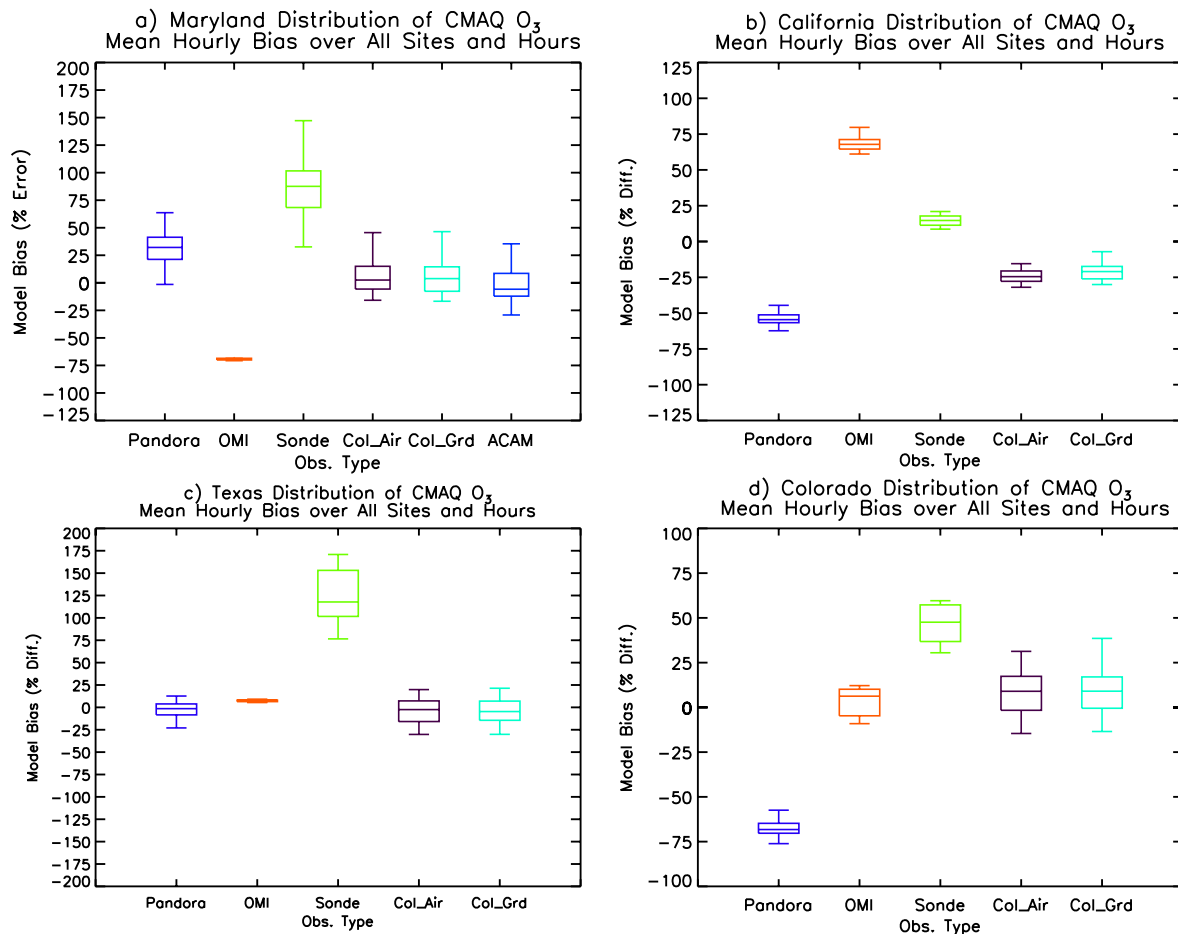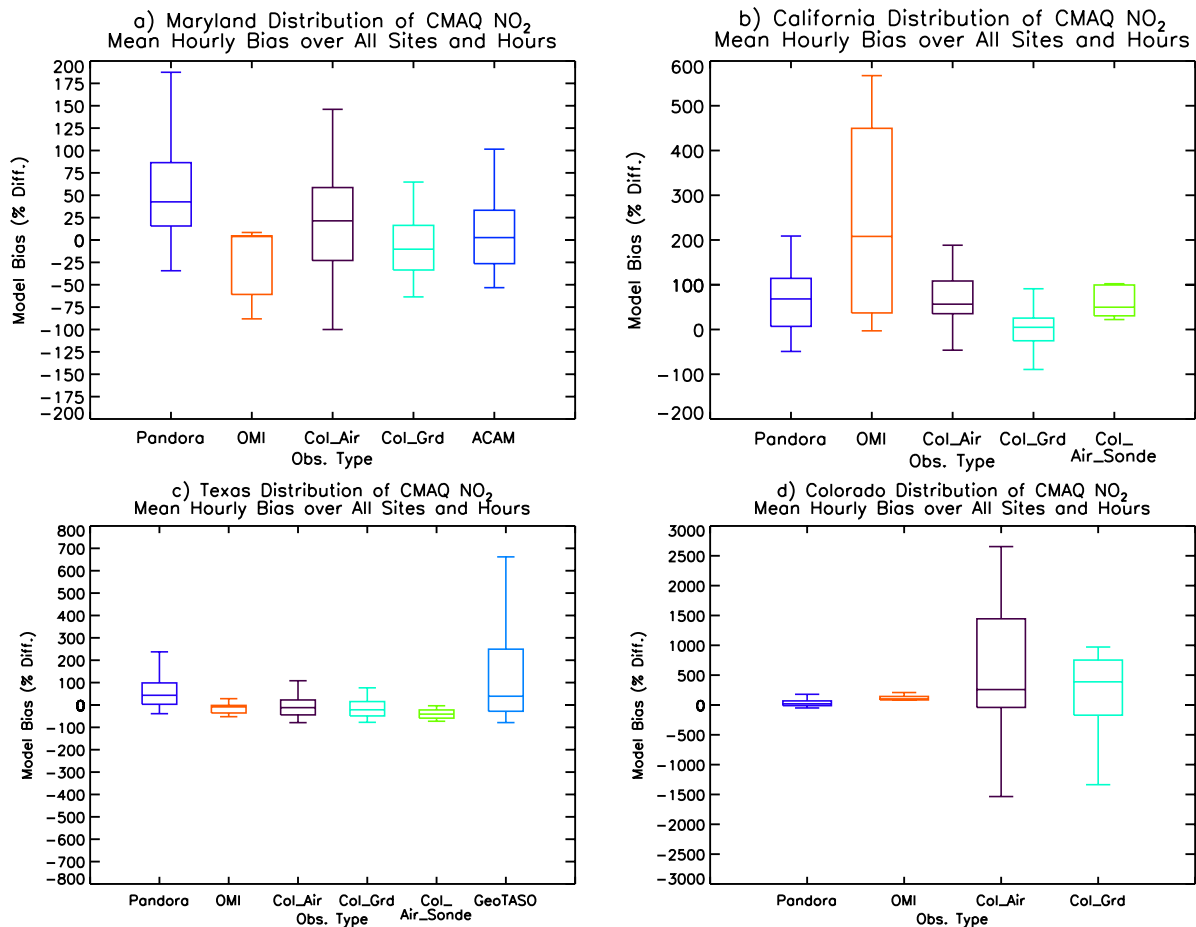
Fig. 4.9: Box-and-whisker plots of model NO$_2$ bias median values and distributions for a) the Maryland campaign, b) the California campaign, c) the Texas campaign, and d) the Colorado campaign. Model bias computed as percent difference relative to each observational dataset. Tops and bottoms of boxes represent the 75th and 25th percentile bias values, respectively, while whiskers represent outliers.

4.7: Conclusions

The campaign-average diurnal cycles of surface O$_3$ and NO$_2$ mixing ratios did not vary significantly across spiral sites or campaigns, indicating a consistent diurnal cycle for these trace gases across meteorological and pollution conditions, in agreement with established literature. In general, surface O$_3$ exhibited a minimum during morning, increasing to an afternoon maximum, while surface NO$_2$ exhibited a morning maximum, decreasing to an afternoon minimum. In contrast, neither full tropospheric nor lower tropospheric O$_3$ column abundances exhibited a clear

diurnal cycle for any spiral site or campaign. This suggests that PBL dynamics play an important role in the regulation of the variability of these columns: O_3 tends to be well mixed within the PBL during daytime and persists in the residual layer during nighttime, such that though the surface concentrations of O_3 vary, the abundance of O_3 within the column does not vary significantly throughout the course of the day. The NO_2 full tropospheric and lower tropospheric column abundances, however, did display diurnal variability at most spiral sites during all campaigns, though the column diurnal variability was smaller in amplitude and offset in time relative to the surface diurnal variation. The NO_2 column datasets presented a consistent diurnal variation across all four campaigns and all datasets. These results suggest that NO_2 column variability is controlled by surface production of NO_2 and boundary layer mixing of NO_2 into the lower troposphere. Neither set of results for O_3 and NO_2 suggest a time of day when satellite column observations may be most representative of surface concentrations.

Comparison of the simulated column diurnal variation within the regional CMAQ model and global GMI model demonstrated that both models replicated the lack of O_3 column diurnal variation, although these models overestimated the lack of variation. CMAQ demonstrated the smallest campaign-median model bias relative to the P-3B O_3 column_air and column_ground abundances across all four campaigns, indicating that CMAQ well captured the magnitudes of the lower tropospheric columns. Additionally, the smallest biases relative to the observational columns were typically obtained for the Maryland and Texas campaigns, suggesting that CMAQ was most able to capture the observed O_3 columns for the conditions associated with these campaigns, which included deep, convective boundary layers and adequate temperatures and sunlight for O_3 formation. In agreement with the magnitudes and shapes of the observed NO_2 column diurnal variability, CMAQ and GMI exhibited clear diurnal variation in the simulated

NO$_2$ column amounts. CMAQ also best captured the observed NO$_2$ column magnitudes for all campaigns except California, which experienced greater NO$_2$ pollution and inhibited O$_3$ photochemistry.

Chapter 5: Evaluation of Vertical Mixing and Pollutant Shape Factors in WRF-Chem during DISCOVER-AQ July 2011

5.1: Introduction

Boundary layer mixing exerts an important influence on the connection between O_3 and NO_2 column and surface data (Chapters 2 and 3). For example, Flynn et al. (2016, submitted) found greater correlation between lower tropospheric O_3 column amounts and surface mixing ratios during the DISCOVER-AQ Maryland campaign in July 2011 than during the California campaign in winter 2013, due in large part to the relative strength of the vertical mixing encountered during these campaigns. Other works demonstrated improvements in correlation between NO_2 tropospheric column amounts and surface mixing ratios after inclusion of PBL height (PBLH) information. Knepp et al. (2013) used model-derived PBLH values to convert ground-based Pandora UV/Vis spectrometer NO_2 tropospheric columns into average surface mixing ratios, demonstrating high correlation between converted columns and surface data. Flynn et al. (2014) similarly normalized NO_2 lower tropospheric column amounts by observed PBLH estimates. Greater correlation between these normalized column amounts and surface data was obtained than between the original column amounts and surface data. These studies suggest that the degree and height of PBL mixing adds useful information to the column-surface relationship.

Vertical mixing, and, by extension, atmospheric stability, exert an influence on the column-surface relationship through its impact on the vertical distribution of pollutants. Turbulent eddies communicate momentum, heat, and passive scalars such as moisture and pollutants between the surface and the atmospheric boundary layer (Holtslag and Boville, 1993; Stull, 1988). Under unstable conditions, eddies mix the PBL more efficiently and to greater depths than possible under stable conditions. Stronger vertical mixing, associated with greater

atmospheric instability, thus acts to produce more uniform pollutant vertical profiles, as pollutants are distributed more evenly among different levels within the PBL and are mixed to greater depths; pollutant profiles often exhibit a layered structure under stable conditions and weaker mixing (Holtslag and Boville, 1993; Lin et al., 2008; Lin and McElroy, 2010). Further, the shape of pollutant vertical profiles determines which altitude layers contribute most to the column amount associated with that profile, which in turn impacts how well that column amount relates to the surface. A column associated with a uniform profile within the PBL, indicative of strong vertical mixing, may relate better to the surface than a column associated with a profile exhibiting enhanced concentration in the upper or lower PBL, indicative of less vigorous vertical mixing as well as transport of pollutants or production of pollutants, for example. This effect is apparent in recent work relating OMI NO_2 tropospheric column amounts to estimated *in situ* NO_2 columns through use of assumed profile shapes. Boersma et al. (2009) employed an assumption of vertically uniform boundary layer NO_2 profiles, with negligible mixing ratios above the PBL, to convert observed surface mixing ratios to *in situ* column amounts, an assumption that was verified by the GEOS-Chem global chemical transport model (CTM). Ordóñez et al. (2006) computed the portion of the NO_2 column within each vertical layer from the MOZART-2 global CTM, which also assumes a well mixed PBL NO_2 profile, to scale observed near-surface NO_2 mixing ratios into vertical column amounts. These estimates compared well to the OMI tropospheric NO_2 columns. Addtionally, these studies and that of Knepp et al. (2013) emphasize the need for model-derived information on PBL mixing for relating surface and column data, due to lack of sufficient observations. In contrast to these works, Zhang et al. (2016) examined potential temperature and NO_2 vertical gradients during the Maryland deployment, and demonstrated that the NO_x vertical gradient (and thus profile shape) is highly sensitive to

atmospheric stability: NO_2 profiles exhibited an exponential decay shape under stable conditions, with mixing ratios decreasing sharply from a maximum near the surface to minimum values above the PBL, while the proifle shape becomes more well mixed (smaller vertcial gradient in mixing ratio) as the PBL becomes more unstable (see Zhang et al., 2016, Fig. 1). Further, Zhang et al. found that the use of a well-mixed boundary layer NO_2 profile could lead to errors of ~45% in estimated boundary layer O_3 production and that variabiltiy in NO_2 profiles may account for ~5-15% of variablity in retrieved NO_2 tropospheric vertical columns.

Understanding of the interplay between boundary layer mixing and vertical profile shapes contributes to understanding of the degree of correlation between column and surface data, with implications for the applicability of observations from upcoming geostationary air quality satellites such as TEMPO (Chance et al., 2013) to surface or near-surface air quality conditions. As described in Chapter 3, assumed profiles derived from global chemical transport models (CTMs) provide the *a priori* information in the conversion of satellite-retrieved slant column amounts to vertical columns for NO_2, and O_3 when retrieved with a differential optical absorption spectroscopy (DOAS) technique, such as the OMI O_3 total column retrieval. Other retrievals of O_3 vertical columns employ simulated *a priori* partial O_3 column amounts at each model vertical layer. However, the assumed profile shapes are given as shape factors, or the fraction of the tropospheric column within each vertical layer, rather than profiles of mixing ratio values. It is again emphasized that the shape factor depends on the mixing ratio profile, and thus partially on vertical mixing, as it influences the distribution of pollutants within the profile. Potential errors in satellite retrievals maximize when there is bias in the shape factor profile at the altitudes at which the satellite instrument is most sensitive, as determined by the retrieval scattering weights. For example, OMI is more sensitive to NO_2 in the upper PBL and free

troposphere than near the surface, so that errors in the NO_2 shape factor at these levels will maximize retrieval errors here (Bhartia et al., 2002; Chance, 2002; Lamsal et al., 2014; Liu et al., 2009; Palmer et al., 2001). The models used within satellite retrievals or used to relate column and mixing ratio data must therefore accurately simulate the observed O_3 and NO_2 profiles and column amounts. Therefore, they must also accurately simulate boundary layer mixing. Additionally, there is interest in the use of regional rather than global air quality models in remote sensing retrievals for finer horizontal resolution simulated profiles (Nowlan et al., 2016; Lok Lamsal, NASA GSFC, personal communication).

Given the difficulty in computation of boundary layer turbulence, regional and global meteorology and chemical transport models employ PBL parameterization schemes to simulate boundary layer mixing and compute the PBLH. These PBL schemes are classified as local, in which mixing occurs only between model vertical levels adjacent to each other, and nonlocal, in which mixing occurs between adjacent and nonadjacent vertical levels; a fuller description of local and nonlocal PBL schemes is given below in Sections 5.6.2-5.6.4 (Holtslag and Boville, 1993; Hu, et al., 2010; Lin and McElroy, 2010; Shin and Hong, 2011; Stull, 1988; Tang et al., 2011). The choice of a local or nonlocal PBL scheme can have important impacts on the vertical structure of temperature and moisture, as demonstrated by Holtslag and Boville (1993) within the NCAR Community Climate Model, version 2 (CCM2). Hu et al. (2011) examined the Yonsei University (YSU; Hong et al., 2006), the Asymmetric Convective Model version 2 (ACM2; Pleim, 2007a), and the Mellor-Yamada- Janjić (MYJ; Janjić, 1990, 1994) PBL schemes within the WRF model over the south-central U.S. The YSU and ACM2 schemes, both nonlocal schemes, produced smaller biases in temperature and moisture during daylight hours than the MYJ scheme, the local scheme, and hence produced more accurate simulations than the local

162

scheme. Generally, nonlocal PBL schemes better represent the vertical structures of the PBL under unstable conditions, while both local and nonlocal schemes struggle to represent these structures under stable conditions; stable conditions remain difficult to accurately simulate (Shin and Hong, 2011).

Differences in simulated vertical mixing within different PBL schemes affects the accuracy of air quality simulations. Global CTMs often employ a full mixing assumption, in which the boundary layer is uniformly mixed, or offer a choice among local or nonlocal PBL schemes. Use of a nonlocal mixing scheme within MOZART and GEOS-Chem improved biases in surface O_3 concentrations and improved the shape of the lower tropospheric NO_2 relative to observations (Lin et al., 2008; Lin and McElroy, 2010). Comparison of the correlation between O_3 or NO_2 column and surface data within CMAQ output and within DISCOVER-AQ observations (see Chapters 2 and 3) revealed that CMAQ overestimated the degree of correlation for both trace gases during the Maryland campaign, suggesting that vertical mixing is overestimated within this model and carrying implications for the use of a regional air quality model in relating column and surface quantities. However, though CMAQ ingests simulated mixing parameters from a meteorology model (which can use a variety of PBL schemes) after preprocessing by the MCIP routine, the CMAQ model itself is capable of using only the ACM2 PBL scheme to simulate vertical mixing of pollutants. Recent work has thus found that surface O_3 concentrations were relatively insensitive to the choice of PBL scheme within the meteorology model used to drive CMAQ (Mao et al., 2006; Tang et al., 2011). Zhang et al. (2016) using a one-dimensional CTM, found that accuracy of simulated NO_2 profiles is relatively insensitive to the choice of land surface model, but that choice of PBL scheme had some impact. These studies present mixed results for the ability of air quality models or CTMs to accurately

capture vertical mixing and thus pollutant profiles. However, few studies examined pollutant profile sensitivity to the choice of vertical mixing scheme in the CTM, or the implications of uncertainties in such sensitivities to satellite retrievals. Additionally, none of these studies examined coupled meteorology-chemistry models. Uncertainties thus remain in the understanding of the interplay of vertical mixing and pollutant profiles.

The Maryland deployment of the DISCOVER-AQ project (Chapter 2, Section 2.1) presents an excellent opportunity to study the impact of boundary layer mixing on the O_3 and NO_2 vertical profiles and column-surface relationships. As discussed previously, the greatest column-surface correlation for O_3 was during the Maryland deployment, suggesting that satellite observations may be most relevant to surface air quality under the conditions of deep, convective boundary layers, reduced wind shear, and few terrain influences associated with this region and deployment. The WRF model with coupled Chemistry (WRF-Chem) was chosen for use in this work, as chemistry and meteorology are computed in the same time step, eliminating the need for preprocessing of meteorology model output to drive the air quality simulation. Results are presented for WRF-Chem simulations with five different PBL schemes (Sections 5.6.3-5.6.4) of the July 26-29, 2011, period, during which an O_3 pollution episode occurred. The ability of this regional, coupled chemistry-meteorology model to effectively simulate the interplay between boundary layer mixing and O_3 and NO_2 vertical profiles, and the associated impacts on the column-surface correlations for these trace gases, is investigated. Further, the relevance of the WRF-Chem model profiles for use in remote sensing retrievals will be evaluated.

5.2: P-3B *In Situ* Profile Measurements

The P-3B *in situ* O_3 and NO_2 profile measurements for the Maryland campaign have been described in Chapter 3, Section 3.2

5.3: P-3B Column Air and Column Ground Lower Tropospheric Columns

The P-3B O_3 and NO_2 column_air and column_ground data sets for the Maryland campaign have been described in Chapter 2, Section 2.2.1.

5.4: of Surface Volume Mixing Ratio Datasets

The surface volume mixing ratio data sets available for the Maryland campaign have been described in Chapter 2, Section 2.3.

5.5: Observational PBLH Datasets

5.5.1: P-3B PLBH Dataset

Donald Lenschow (NCAR, retired) provided boundary layer height analyses based on the P-3B potential temperature profiles during the Maryland campaign. The PBL top was located where the potential temperature lapse rate exceeded approximately 3 K/km, with a relatively constant potential temperature lapse rate from the surface to the PBL top. The potential temperature profiles were also analyzed manually to ensure the algorithm accurately diagnosed the PBLH. Water vapor and ozone profiles were examined in addition to the potential temperature profiles. These PBLH estimates have an uncertainty of approximately 20%.

5.5.2: HSRL and MPL Mixed Layer Height Datasets

Ground-based MicroPulse Lidar instruments (MPL; Welton et al., 2002) were available at the Beltsville, Edgewood, and Fair Hill spiral sites and were operated during all days in July 2011; the NASA UC-12 aircraft provided continuous remote sensing observations of aerosols from the High Spectral Resolution Lidar (HSRL; Hair et al., 2008) during flight days over all spiral sites and over the transects between spiral sites. The Haar covariance wavelet transform (Brooks, 2003) was used to estimate mixed layer height (MLH) values from the aerosol backscatter profiles observed by the MPLs and the HSRL (Ruben Delgado, http://www-

air.larc.nasa.gov/cgi-bin/ArcView/discover-aq.dc-2011?ANALYSIS=1#ML.HEIGHT/; Scarino et al., 2014; both datasets have an associated uncertainty in PBLH of approximately 20%). It should be noted that the HSRL MLH values are the "best estimate" values from comparison of the Haar transform results and manual inspection of the backscatter profiles.

5.6: Description of WRF-Chem Simulations and Column Amounts

5.6.1: WRF-Chem Model Options

The WRF-Chem version 3.7.1, with Advanced Research WRF (ARW) core (the latest model version available; Grell et al, 2005; Fast et al., 2006) was used to simulate the July 26-29, 2011, period, when an O_3 pollution event occurred; this pollution event was chosen for simulation rather than the entire July 2011 campaign due to time constraints and the larger amount of computing resources required by WRF-Chem relative to offline air quality simulations. The online chemistry-meteorology WRF-Chem model was used in this study to study the effects of vertical mixing on the chemistry. The traditional, offline WRF-CMAQ system requires that the WRF output be processed and time-averaged (typically hourly averages) before passing to CMAQ in addition to preventing feedback from the chemistry onto the dynamics, while the coupled WRF-CMAQ model cannot be run for domains outside the testbed domain at the time of this writing. The simulations were begun on July 22, to provide three days of model spin-up time. The model options common to all simulations are listed in Table 5.1, while the nested simulation domains are illustrated in Fig. 5.1. WRF-Chem was run with 34 vertical layers from the surface to 100 mb, with 16 layers within the lowest 2 km AGL to capture boundary layer processes. The NARR was used for the meteorological initial and boundary conditions, while a simulation of the MOZART-4 model provided the chemical initial and boundary conditions (Emmons, et al., 2010). WRF-Chem used the unified Noah land surface

model (Niu et al., 2011); the Fast-J photolysis scheme (Wild et al., 2000); the Goddard shortwave radiation scheme (NASA Goddard Space Flight Center); the Rapid Radiative Transfer Model (RRTM) longwave radiation scheme (Mlawer et al., 1997); the Carbon-Bond Mechanism version Z (CBM-Z) chemical mechanism (Zaveri and Peters, 1999); and the Model for Simulating Aerosol Interactions and Chemistry (MOSAIC) aerosol module with four aerosol size bins (Barnard et al., 2010). The Model of Emissions of Gases and Aerosols from Nature (MEGAN; Guenther et al., 2006) was used for biogenic emissions, and the online version employed in WRF-Chem v.3.7.1 is the same as the offline version 2.04. The projected 2012 anthropogenic emissions based on the 2005 NEI described in Chapter 2 were used in these simulations. These older emissions were used rather than the NEI 2011 due to constraints on time available to run simulations, due to the amount of time taken to get a coupled model successfully running. The 12 km horizontal resolution output was used in the following analyses. Five simulations with five different PBL schemes (Table 2) were performed, to be described in the following three subsections.
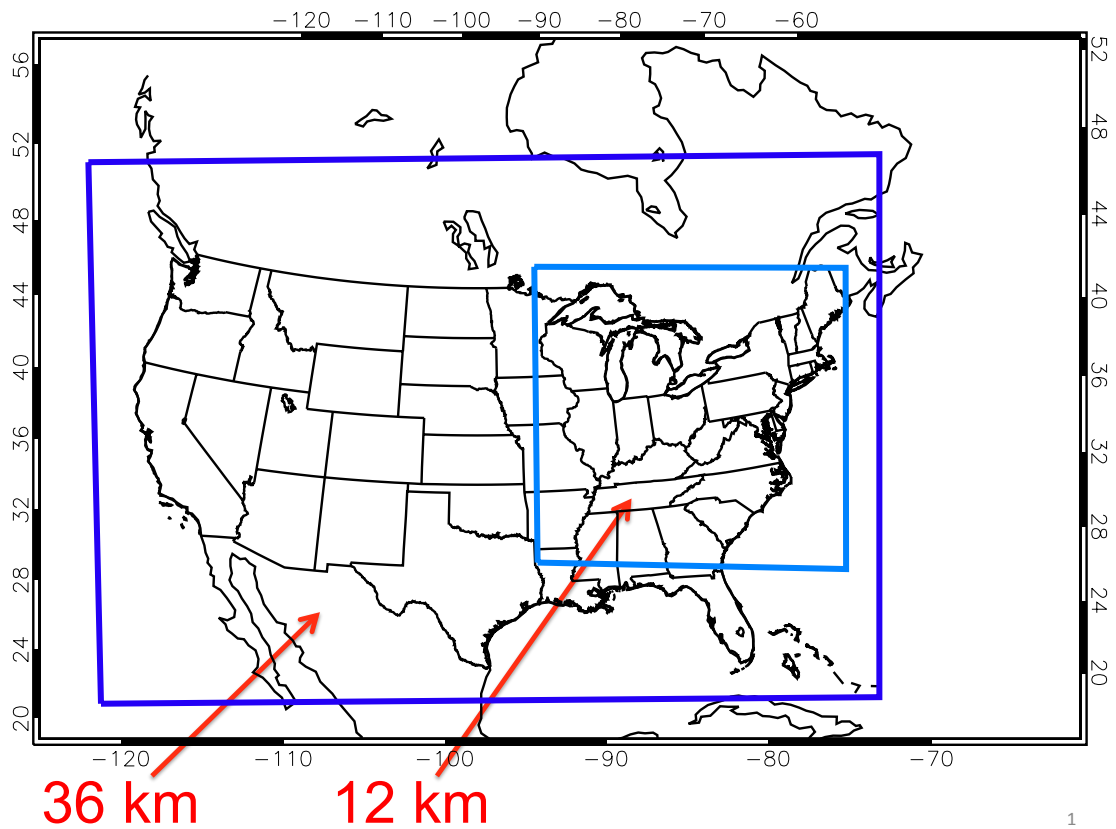
Fig. 5.1: Geographical domains for the 36 km and 12 km horizontal resolution domains. The outermost 36 km simulation covers the continental U.S., while the 12 km simulation covers the eastern U. S.

WRF-Chem v3.7.1 Maryland DISCOVER-AQ Campaign Common Options	
Time Period	July 22-31, 2011 Focus of Analysis: July 26-29, 2011
Chemical mechanism	CBM-Z
Aerosols	MOSAIC with 4 aerosol bins
Radiation	Longwave-RRTM; Shortwave-Goddard
Meteorology and Chemical Inputs	NARR; MOZART-4 CTM
LSM	unified Noah LSM
Photolysis	Fast-J

Table 5.1: WRF-Chem v3.7.1 simulation options common to all simulations.

WRF-Chem v3.7.1 Maryland DISCOVER-AQ Campaign PBL Schemes	
PBL Scheme Simulation	**Surface Layer Scheme**
YSU	MM5 similarity theory (formulated for YSU scheme)
ACM2	Pleim-Xiu surface layer scheme
MYJ	Eta similarity theory
BouLac	Eta similarity theory
QNSE	QNSE similarity theory

Table 5.2: WRF-Chem v3.7.1 PBL schemes tested with corresponding surface layer schemes. Simulations named after the PBL scheme used in that simulation.

5.6.2: Atmospheric Turbulence, the Closure Problem, and Turbulence Parameterizations

A fundamental problem in the statistical description of atmospheric turbulence is the so-called "closure problem." The number of unknown variables will always exceed the number of equations for any finite set of equations used to describe turbulence (Stull, 1988). This problem is not critical within the free troposphere, where the turbulent eddies are resolved by the synoptic observational network; turbulent eddies, however, are of crucial importance within the PBL, because they transport heat, moisture, momentum, and other scalar quantities between the surface and lower portion of the atmosphere. The existence of turbulent eddies at many temporal and spatial scales, which may or may not be resolvable with current instrumentation, creates the closure problem in the description and modeling of boundary layer flow. Thus, the unresolved turbulent fluxes must be parameterized in terms of known mean quantities to compute turbulent transport within the PBL (Holton, 2004; Stull, 1988).

Two broad categories of turbulent parameterizations are commonly used in atmospheric modeling: local closure parameterizations and nonlocal closure parameterizations. In local closure, the turbulent fluxes at one point in space are approximated by mean atmospheric variables or gradients at that same point. In nonlocal closure, the turbulent fluxes at one point are approximated by mean variables or gradients at many points in space (Holtslag and Boville, 1993; Hu, et al., 2010; Shin and Hong, 2011; Stull, 1988). Furthermore, parameterization schemes are characterized by the highest-order prognostic equation retained, which can range from zero-order closure to third-order closure. For example, in a first-order closure scheme, the prognostic equation for the x-component of the mean wind $\overline{U}$, a first-order statistical moment, is retained:

$$\frac{\partial \overline{U}}{\partial t} = -\frac{\partial}{\partial z}\left(\overline{u'w'}\right) \tag{5.1}$$

However, this equation contains a second-moment turbulent flux term, $\overline{u'w'}$, which must be parameterized, typically through use of an eddy diffusivity parameter K. One such possible parameterization is given in Eq. 5.2:

$$\frac{\partial \overline{U}}{\partial t} = -\frac{\partial}{\partial z}(\overline{u'w'}) = \frac{\partial}{\partial z}[K\left(\frac{\partial \overline{U}}{\partial z}\right)] \tag{5.2}$$

No third- or higher moment terms are retained in first-order closure (Shin and Hong, 2011; Stull, 1988). Typically, local first-order or one-and-a-half-order closure parameterizations, or nonlocal first-order closure parameterizations, are used within atmospheric models, and the following discussion will thus focus on PBL schemes employing such parameterizations (Holtslag and Boville, 1993).

5.6.3: Local PBL Schemes Tested in WRF-Chem

Boundary layer and associated surface layer schemes examined in this work are listed in Table 5.2. Three commonly used local, 1.5-order PBL schemes include the Mellor-Yamada-Janjic (MYJ; Janjić, 1990, 1994), Bougeault-Lacarrére (BouLac; Bougeault and Lacarrere, 1989), and the quasi-normal scale elimination (QNSE; Sukoriansky, et al., 2005) PBL schemes. These schemes are classified as 1.5-order because they retain the prognostic equations for the first-moment terms, such as $\overline{U}$, and parameterize the turbulent flux terms in terms of eddy diffusivities, but also require an additional prognostic equation for the turbulent kinetic energy (TKE or e), which is computed in these schemes as

$$\frac{\partial e}{\partial t} = \frac{-1}{\rho}\frac{\partial \rho \overline{w'e'}}{\partial z} - \overline{u'w'}\frac{\partial U}{dz} - \overline{v'w'}\frac{\partial V}{\partial z} + \beta\overline{w'\theta'} - \in \tag{5.3}$$

where the $\beta\overline{w'\theta'}$ term represents buoyant production of TKE and $\in$ represents viscous dissipation. TKE is used to compute the eddy diffusivities within these three schemes:

$$K_C = l\sqrt{e}S_C \tag{5.4}$$

where K_C is the eddy diffusivity term for any mean scalar or vector quantity C, such as heat or momentum, l is the mixing length, and S_C is the proportional coefficient. These schemes employ the local diffusivities within both the convective boundary layer (CBL) and the stable boundary layer (SBL), and make no distinction between boundary layer and free atmospheric mixing (Hu et al., 2010; Shin and Hong, 2011). Each scheme defines l and S_C somewhat differently. Furthermore, the BouLac scheme assumes that the eddy diffusivities for heat and momentum are equivalent (Bougeault and Lacarrere, 1989):

$$K_H = K_M \tag{5.5}$$

whereas the MYJ scheme assumes the following relationship for the heat and momentum diffusivities (Janjić, 1990, 1994):

$$K_H = 1.25 K_M \tag{5.6}$$

The QNSE scheme computes the eddy diffusivities from spectral theory, to account for internal wave generation in the presence of turbulence within the stably stratified PBL, and assumes the following (Sukoriansky et al., 2005):

$$K_H = Pr_{t0}^{-1} K_M \tag{5.7}$$

where Pr_{t0}^{-1} is the inverse turbulent Prandtl number computed at the first model time step. The turbulent Prandtl number is defined as the ratio of the momentum eddy diffusivity to the thermal eddy diffusivity, which represents the ratio of the viscous diffusion ratio to the thermal diffusion rate. The eddy diffusivity for heat (K_H) is also used for other scalars such as pollutants. Lastly, each of these three schemes defines the height of the PBL (PBLH) as the level at which TKE first decreases to a prescribed value of 0.1 m^2/s^2.

5.6.4: Nonlocal PBL Schemes Tested in WRF-Chem

Nonlocal PBL schemes may be more appropriate than local schemes for convective or

unstable conditions, in which the size scale of the transporting eddies may be of the same order as the depth of the boundary layer (Holtslag and Boville, 1993). Two nonlocal, first-order PBL schemes are currently available for use in atmospheric models: the Asymmetric Convective Model version 2 (ACM2; Pleim, 2007a and recent revisions) and the Yonsei University (YSU; Hong et al., 2006) schemes. These schemes include local mixing, and account for nonlocal mixing through use of either a nonlocal gradient adjustment term or a transilient matrix of mixing coefficients. The YSU applies a gradient adjustment term (γ_c) to the local gradient of each prognostic mean quantity to implicitly express nonlocal mixing within the heat and momentum prognostic equations only:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z}[K_C(\frac{\partial C}{\partial z} - \gamma_c) - \overline{w'c'}_h \left(\frac{z}{h}\right)^3]$$ (5.8)

where C represents any mean heat, moisture, chemical species, or momentum scalar or vector quantity and h is the height of the PBL; the gradient adjustment term is also shown in Eqn, 5.8, though it is applied only to the heat and momentum prognostic equations. The $(- \overline{w'c'}_h \left(\frac{z}{h}\right)^3)$ term represents the asymptotic entrainment flux within the inversion layer that frequently caps the PBL. The gradient adjustment term in the heat and momentum prognostic equations is computed as

$$\gamma_c = b \frac{\overline{(w'c_0')_0}}{w_{s0}h}$$ (5.9)

where $\overline{(w'c')_0}$ represents the surface flux for heat or momentum quantity c, w is the vertical velocity, w_{s0} is the vertical velocity scale at 0.5h, and b is constant of proportionality (Hong et al., 2006). The ACM2 scheme, however, employs the transilient matrix technique to explicitly model local and nonlocal mass fluxes, for better representation of passive quantities such as chemical tracers or humidity. ACM2 also makes use of a weighting factor f_{conv} in its governing

equation:

$$\frac{\partial C_i}{\partial t} = \text{f}_{conv} Mu C_1 - \text{f}_{conv} Md_i C_i + \text{f}_{conv} Md_{i+1} C_{i+1} \frac{\Delta z_{i+1}}{\Delta z_i} + \frac{\partial}{\partial z}[K_C(1 - f_{conv})\frac{\partial C}{\partial z}]$$ (5.10)

where C is any mean scalar quantity, such as a pollutant, C_1 is the magnitude of the quantity in layer 1, Mu is the upward convective mixing rate from the lowest vertical layer, Md_i is the nonlocal downward mixing rate from layer i to i-1, and Δz_i is the layer thickness (Pleim, 2007a; Shin and Hong, 2011). The f_{conv} parameter is further related to the eddy diffusivity for heat (K_h) and the gradient adjustment term for heat (γ_h) through

$$f_{conv} = \frac{K_h \gamma_h}{K_h \gamma_h - K_h \frac{\partial \theta}{\partial z}}$$ (5.11)

where K_h is the eddy diffusivity for heat and γ_h is gradient adjustment term, defined similarly to the corresponding term in the YSU scheme. It is this parameter that controls the degree of nonlocal vs. local mixing within the ACM2 scheme, as Eqn. 5.11 demonstrates that f_{conv} is the ratio of the nonlocal flux to the total flux (nonlocal and local). As a coefficient in the governing prognostic equation, f_{conv} partitions the upward and downward mixing rates Mu and Md_i from the total mixing rates to the nonlocal mixing rates. Thus, the YSU and ACM2 schemes adjust for stable or neutral flows through adjustments to the gradient adjustment terms and the f_{conv} parameter (Hong et al., 2006; Hu et al., 2010; Pleim, 2007a; Shin and Hong, 2011).

The YSU and ACM2 schemes diagnose the PBLH based on the bulk Richardson number (Ri_b) rather than TKE, which is defined as

$$Ri_b = \frac{g[\theta(h) - \theta_s](h - z_{mix})}{\overline{\theta_v}[U(h) - U(z_{mix})]^2}$$ (5.12)

The PBLH is determined as the first level at which the bulk Richardson number (Ri_b) first exceeds a critical Richardson number (Ri_{crit}), set to 0.25 in both schemes. However, some differences exist between YSU and ACM2 in determination of that first level. The YSU scheme

computes Ri_b over all model levels, and then compares that value to the critical value for each level. Under unstable conditions, the ACM2 scheme assumes that the PBLH is the sum of the top of the convectively mixed layer (z_{mix}) and the thickness of the entrainment layer, so that ACM2 first computes z_{mix} based on the virtual potential temperature profile and then computes Ri_b over the depth of the entrainment layer only, using a prognostic equation similar to that for Ri_b within YSU. The value of z_{mix} is thus used in the computation of Ri_b within the entrainment layer, after Eqn. 5.12. The entrainment layer top is then diagnosed as the first level at which Ri_b is equal to Ri_{crit}. Thus, PBLH is the sum of the entrainment layer thickness and z_{mix} (Hong et al., 2006; Hu et al., 2010; Pleim, 2007a; Shin and Hong, 2011). It is again noted that the ACM2 scheme exclusively is used within CMAQ to simulate vertical mixing, regardless of the choice of PBL scheme used in the meteorological model simulation used to drive CMAQ, and thus the ACM2 scheme is of particular importance in the following evaluations.

5.7: Analysis of PBL Mixing in WRF-Chem

5.7.1: Comparison of Observed and Simulated PBL Heights

5.7.1.1: Comparison to P-3B PBL Heights

     None of the five PBLH schemes tested compared well to the P-3B PBLH estimates in terms of magnitude of the boundary layer depth, as shown in Fig. 5.2a)-e). However, the P-3B dataset represents instantaneous estimates of PBLH, whereas the WRF-Chem PBLH values are provided as hourly averages. Though the WRF-Chem estimate most coincident in time to each P-3B spiral was sampled for use in these analyses, the mismtach in temporal resolution likely prevented good correlation between simulated and observed P-3B PBLH values. The values of χ^2_{red} (displayed on each plot in Fig. 5.2, and similar for all schemes) support the possiblity of temporal mismatch, errors in the model leading to errors in thes simulated PBLH values, as well

as a potential underestimation of the uncertainty in the P-3B PBLH values, while the lack of correlation also indicates a large degree of scatter due to the model overestimating by differing amounts relative to the P-3B PBLH observations. Some generalizations can be made from the scatter plots displayed in Fig. 5.2a)-e). The QNSE scheme most often overpredicted the PBLH relative to the P-3B estimates (percentage of observed PBLH values overpredicted by each scheme shown in Table 5.3); the ACM2 scheme also often overpredicted (Table 5.3), though not as severely as for the QNSE scheme. However, YSU, MYJ, and BouLac also overpredicted at approximately half of the data points (Table 5.3). The mean bias values, defined as the difference between the coincident simulated and observed PBLH values (absolute values were not taken), for each scheme are displayed in Table 5.4, and support the results of the scatter plots. The QNSE and ACM2 schemes produced the largest and second-largest mean bias values, respectively, of all five schemes. The mean PBLH biases produced by these schemes were also statistically significantly larger than those for the YSU, MYJ, and BouLac schemes; further, the QNSE mean bias was statistically significantly larger than that for the ACM2 scheme. The mean biases for the YSU, MYJ, and BouLac schemes were not statistically significantly different from each other. The values of the average perpendicular distance of the data points from the 1:1 line (displayed on the plots in Fig. 5.2) were also similar for the BouLac, MYJ, and YSU schemes, and larger for the ACM2 and QNSE schemes, indicating that these three schemes demonstrated less scatter about the 1:1 and demonstrated less over- or underprediction relative to the observations than the ACM2 and QNSE schemes. Thus, these three schemes performed equally well as each other relative to the P-3B dataset, while the lack of correlation between simulaed and observed PBLH values indicates that these three schemes, while systematically overpredicting relative to the P-3B, often overpredicted by vastly different amounts.

Additionally, the MYJ, BouLac, and YSU schemes performed better relative to the P-3B observations than the ACM2 and QNSE schemes, and the QNSE scheme performed worst of all five schemes.
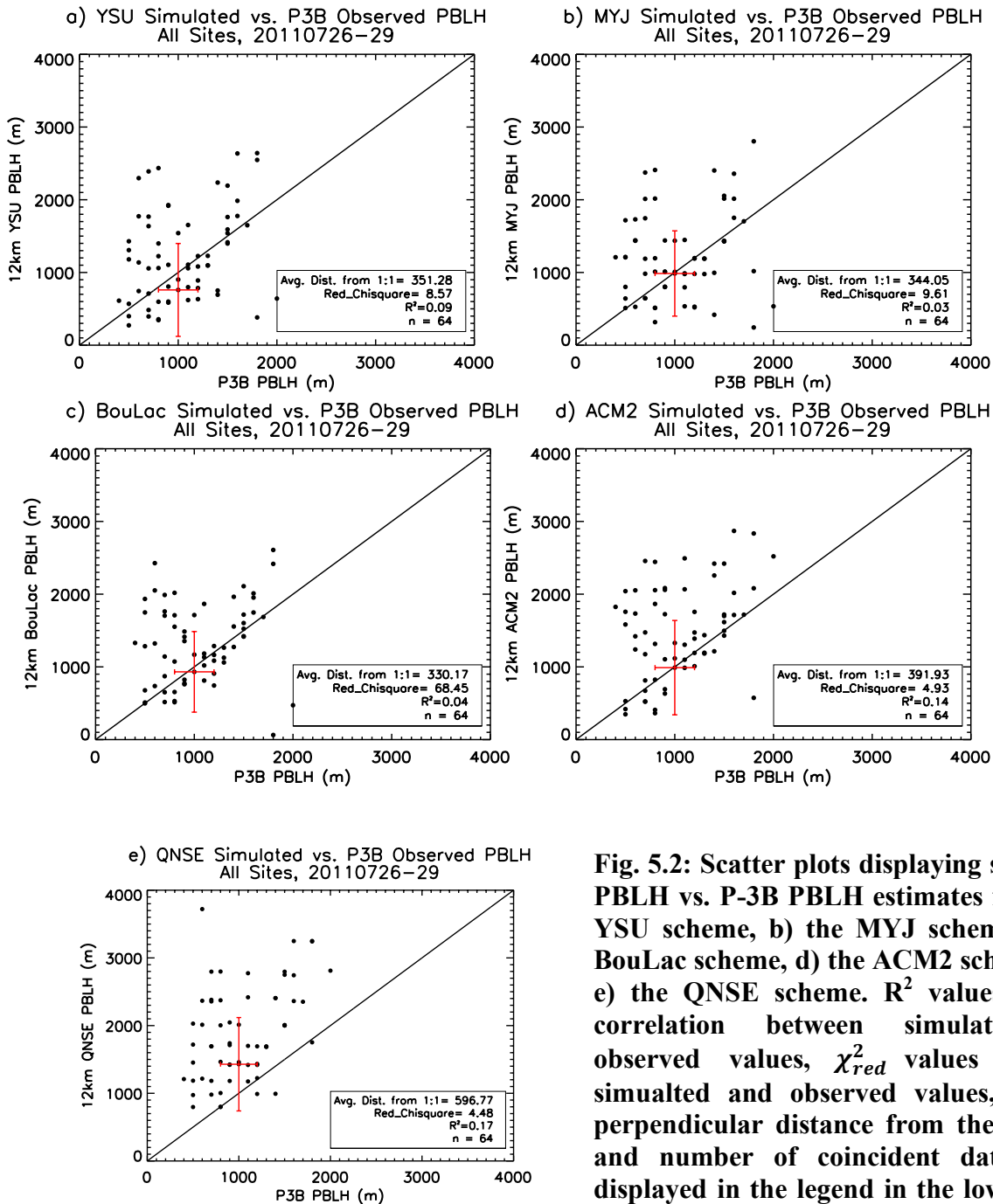
Fig. 5.2: Scatter plots displaying simulated PBLH vs. P-3B PBLH estimates for a) the YSU scheme, b) the MYJ scheme, c) the BouLac scheme, d) the ACM2 scheme, and e) the QNSE scheme. R^2 values for the correlation between simulated and observed values, χ^2_{red} values between simualted and observed values, average perpendicular distance from the 1:1 line, and number of coincident data points displayed in the legend in the lower right. Example uncertainty bars displayed on median data point in red: uncertainty in measurement used for observed PBLH and standard deviation used for simulated PBLH.

177

Frequency of Overpredictions in PBLH			
	P-3B (67 Total Points)	MPL (167 Total Points)	HSRL (110 Total Points)
ACM2	65.7%	53.3%	26.4%
BouLac	58.2%	50.9%	27.3%
MYJ	49.3%	49.7%	25.5%
QNSE	88.1%	83.8%	54.5%
YSU	52.2%	44.9%	22.7%

Table 5.3: Frequency of overprediction in PBLH for each PBL scheme. Total number of points within each dataset listed under the dataset name.

Mean PBLH Bias – P-3B Dataset		
Scheme	Mean Bias (m)	1 Standard Deviation (m)
ACM2	430.9	622.5
BouLac	259.11	616.7
MYJ	166.5	650.3
QNSE	823.1	637.5
YSU	181.9	642.3

Table 5.4: Mean PBL scheme bias values in PBLH relative to the P-3B dataset, and the 1 standard deviation value. All values in meters.

5.7.1.2: Comparison to HSRL and MPL MLH

The MPL and HSRL MLH estimates were computed as hourly averages; thus, the MLH observational datasets and WRF-Chem PBLH values are provided at the same temporal

resolution; the associated χ^2_{red} values (displayed on the plots in Figs. 5.3 and 5.4) were also smaller than those for the comparison to the P-3B, indicating less of a mismatch or underestimate in uncertainty of the observed values. Each PBL scheme compared moderately well to the MPL (Fig. 5.3a)-e) or HSRL (Fig. 5.4a)-e) observations, with R^2 values between 0.49 and 0.66 for the correlations between simulated and observed estimates for each PBL scheme and observational dataset. The correlations between simulated and observed PBLH values for MPL and HSRL were not statistically significantly different from each other, indicating that WRF-Chem performed the same in comparison with both datasets for each PBL scheme. Additionally, the MPL and HSRL scatter plots demonstrate the same clear, consistent overprediction by the QNSE scheme as relative to the P-3B PBLH estimates. However, the ACM2 does not as clearly overpredict relative to the MPL and HSRL as it did relative to the P-3B: the average orthogonal distances to the 1:1 displayed, also displayed on the plots in Figs. 5.3 and 5.4, were similar for all schemes relative to both the MPL and HSRL datasets, with the exception of the QNSE scheme. The average distances suggest that the QNSE scheme least well captured the observed MPL and HSRL MLH values, while BouLac demonstrated a small advantage in capturing the observations over the other four schemes. Most schemes underpredicted low observed MLH values, between approximately 400 m and 600 m, and overpredicted high MLH values relative to the MPL dataset; however, the QNSE scheme did not exhibit much underprediction of low PBLH values. This underprediction is not apparent in the HSRL scatter plots. Additionally, all schemes except QNSE overpredicted relative to the MPL at approximately half of the data points; QNSE overpredicted most frequently (Table 5.3) The mean model biases are consistent with this overprediction relative to the MPL for most schemes: the MPL mean biases were statistically significantly larger than those for the HSRL. The YSU and MYJ schemes present an exception to

this pattern, with absolute values of MPL mean bias values smaller than the HSRL mean bias values. As with the P-3B mean biases, the QNSE scheme demonstrated the largest mean biases relative to the MPL and HSRL datasets, indicating significant overprediction. However, the ACM2, MYJ, YSU, and BouLac scheme mean biases were not statistically different from each other relative to the MPL dataset (Table 5.5), which, coupled with the similar average orthogonal distance values, suggest no scheme clearly performed better than the others. The ACM2 and BouLac mean biases were statistically the same relative to the HSRL dataset, while the YSU and MYJ mean biases were also statistically the same as each other (Table 5.6). As the absolute values of the mean biases for the ACM2 and BouLac schemes were smallest, this suggests that these two schemes best captured the HSRL MLH observations. These results suggest that the ACM2 and BouLac schemes are most adequate to capture boundary layer height observations relative to lidar observations.
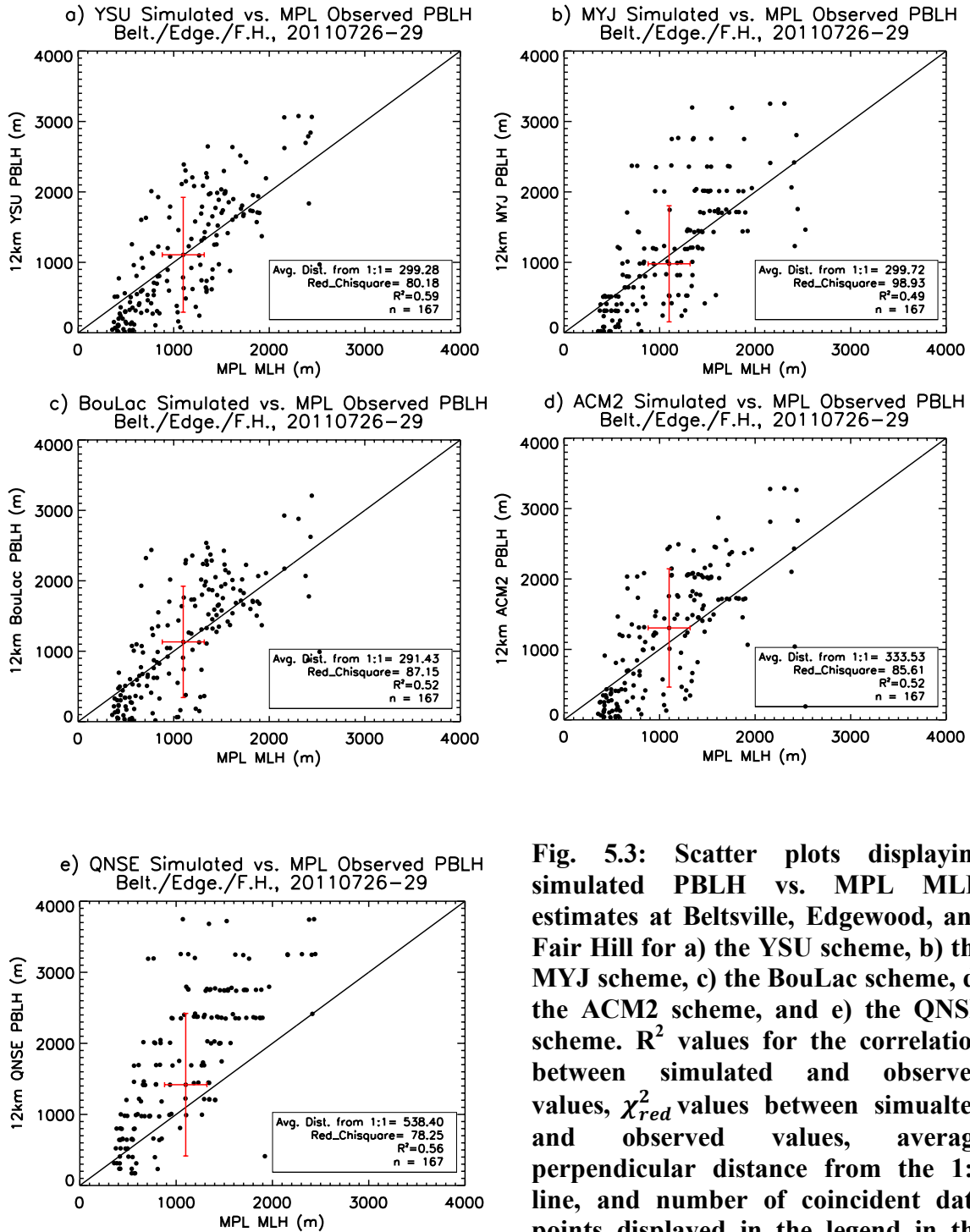
Fig. 5.3: Scatter plots displaying simulated PBLH vs. MPL MLH estimates at Beltsville, Edgewood, and Fair Hill for a) the YSU scheme, b) the MYJ scheme, c) the BouLac scheme, d) the ACM2 scheme, and e) the QNSE scheme. R^2 values for the correlation between simulated and observed values, χ^2_{red} values between simualted and observed values, average perpendicular distance from the 1:1 line, and number of coincident data points displayed in the legend in the lower right. Example uncertainty bars displayed on median data point in red: uncertainty in measurement used for observed PBLH and standard deviation used for simulated PBLH.
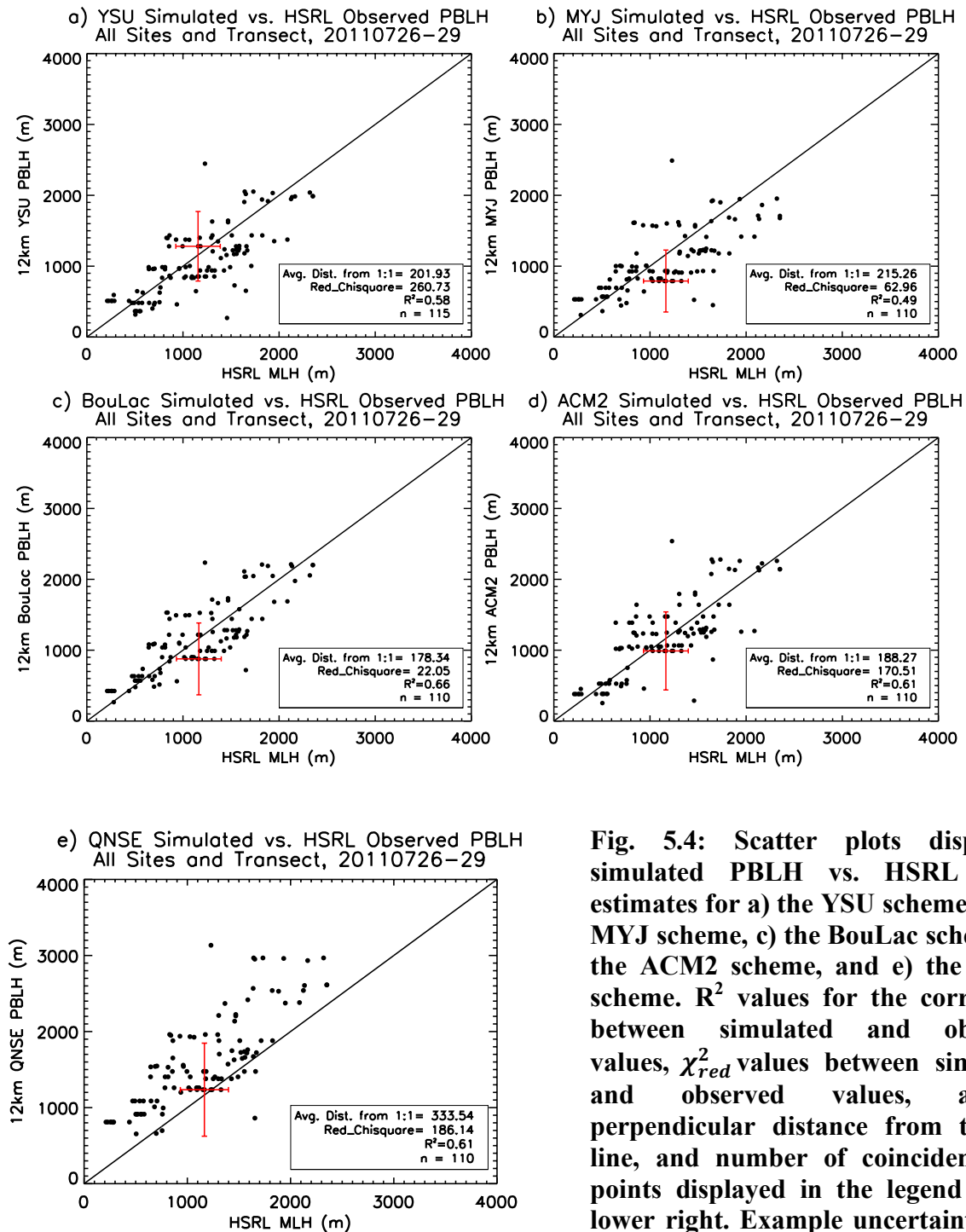
Fig. 5.4: Scatter plots displaying simulated PBLH vs. HSRL MLH estimates for a) the YSU scheme, b) the MYJ scheme, c) the BouLac scheme, d) the ACM2 scheme, and e) the QNSE scheme. R^2 values for the correlation between simulated and observed values, χ^2_{red} values between simualted and observed values, average perpendicular distance from the 1:1 line, and number of coincident data points displayed in the legend in the lower right. Example uncertainty bars displayed on median data point in red: uncertainty in measurement used for observed PBLH and standard deviation used for simulated PBLH.

Mean PBLH Bias – MPL Dataset		
Scheme	Mean Bias (m)	1 Standard Deviation
ACM2	120.5	586.7
BouLac	60.2	545.5
MYJ	63.2	585.3
QNSE	679.0	699.5
YSU	2.9	527.5

Table 5.5: Mean PBL scheme bias values in PBLH relative to the MPL dataset, and the 1 standard deviation value. All values in meters.

Mean PBLH Bias – HSRL Dataset		
Scheme	Mean Bias (m)	1 Standard Deviation
ACM2	26.3	351.3
BouLac	3.2	312.1
MYJ	-92.3	372.9
QNSE	446.6	383.7
YSU	-88.6	345.9

Table 5.6: Mean PBL scheme bias values in PBLH relative to the HSRL dataset, and the 1 standard deviation value. All values in meters.

5.7.2: Comparison of Diurnal Cycles of Observed and Simulated PBL Heights

The average PBLH diurnal behavior within the observational datasets and for each PBL

scheme is displayed in Fig. 5.5a)-c). All available PBLH values were averaged for each hour

between 6 AM and 7 PM EDT for each dataset and PBL scheme tested to produce the average

diurnal cycles. Each scheme captured the general shape of the diurnal cycle relative to each of the three datasets, which is an encouraging result. Consistent with the mean simulated PBLH biases and the scatter plots, the QNSE scheme exhibited a clear, consistent overprediction relative to the P-3B, HSRL, and MPL datasets during all hours examined. The ACM2 scheme also exhibited overprediction from late morning (~10 EDT) through the afternoon (~4 PM EDT) relative to the P-3B and MPL datasets, which corresponds to the development of the CBL. It is notable that the YSU (nonlocal) and MYJ (local) schemes displayed very similar diurnal cycles to each other in terms of PBL depth and the timing of growth and decay of the PBL for each observational dataset examined. The BouLac diurnal cycle is often found between that of the ACM2 scheme and the YSU/MYJ grouping in terms of magnitude and timing for each dataset. Each scheme compared best to the HSRL dataset in both magnitude and diurnal behavior, from approximatley 8 AM EDT until collapse of the PBL initated at approximately 5 pm EDT; the PBL schemes decayed too quickly relative to the MPL and HSRL MLH estimates after this time. Changes were recently implemented into the ACM2 module with WRF v3.7 and newer that were meant to alleviate this error during PBL collapse, after tests in a single column version of WRF; however, these results indicate the problem still exists in the ACM2 implementation in the full, three-dimensional model. The PBL within each scheme also grew too rapidly and too deeply relative to the MPL MLH and P-3B PBLH datasets during the late morning and the early afternoon. However, like the other 4 schemes, the QNSE scheme also displayed the smallest bias in PBLH relative to the HSRL dataset, relative to the comparisons between QNSE and the other two datasets (Fig. 5.5 and Tables 5.4-5.6). Each PBL scheme would also compare more favorably with the P-3B PBLH datasets if not for the low average P-3B PBLH at 3 PM EDT. This low average is due to unusually low PBLH values at Aldino, Beltsville, and Padonia on July

28 and 29. Overall, each scheme well captured the shape of the average diurnal behavior of the P-3B PBLH estimates and the MPL and HSRL MLH estimates. The QNSE scheme exhibited a consistent high bias at all hours analyzed for each dataset, while the remaining four schemes overpredicted the PBLH during the peak depth of the CBL during late morning and afternoon relative to the P-3B and MPL datasets. Each scheme struggled to accurately simulate the development and collapse of the PBL during early morning and late afternoon.
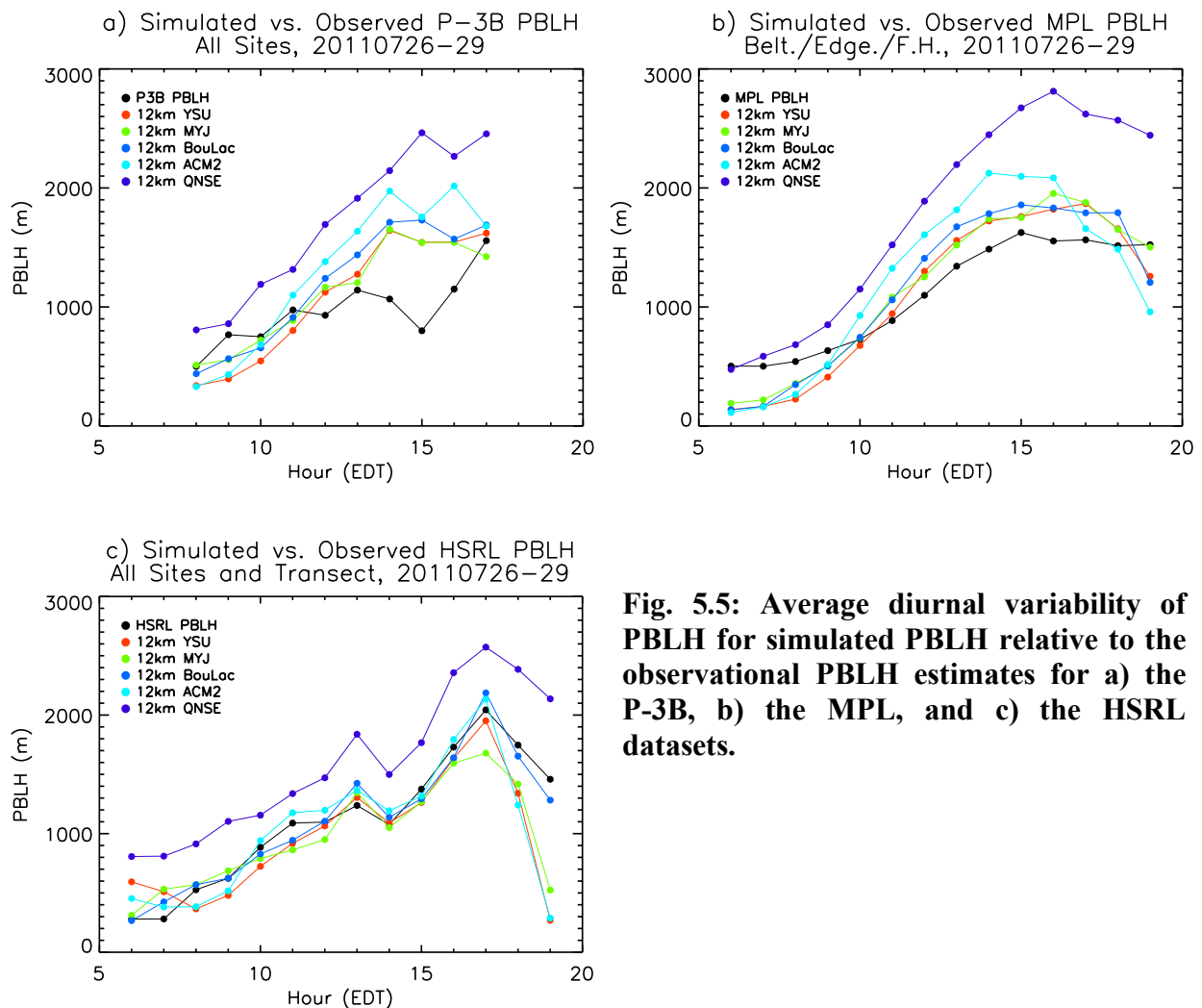
Fig. 5.5: Average diurnal variability of PBLH for simulated PBLH relative to the observational PBLH estimates for a) the P-3B, b) the MPL, and c) the HSRL datasets.

5.8 : Analysis of WRF-Chem Columns and Profiles

5.8.1: P-3B and WRF-Chem Vertical Mixing Ratio Profiles

The median profiles of potential temperature (Fig. 5.6), CO (Fig. 5.7), O_3 (Fig. 5.8), and NO_2 (Fig. 5.9) were computed for each hour between the hours of 8 AM and 5 PM EDT; the example error bars placed at some altitudes represent the 25th and 75th percentile values at that altitude level. The P-3B data were first averaged into 100 m layers before computation of the median profiles, similarly to the column computations, while the CMAQ vertical grid was used with the simulated profiles. Scatter plots of simulated vs. observed values for potential temperature (Fig. 5.10), CO (Fig. 5.11), O_3 (Fig. 5.12), and NO_2 (Fig. 5.13) are also displayed for all available profiles at each hour between 8 AM and 5 PM. The P-3B potential temperature and CO median profiles demonstrate the development of the well mixed, convective boundary layer: by 12 PM EDT, the potential temperature profile within the lowest 1 km of the median profile has become well mixed, the well mixed portion of the profile deepens to approximately 1.2 km by 2 PM, its greatest extent during the simulation period, and remains deeply well mixed throughout the afternoon. The CO profiles are in agreement with the potential temperature profiles, with the lowermost 1 km of the profiles becoming well mixed by 1 PM and remaining well mixed throughout the afternoon. The simulated potential temperature and CO median profiles displayed a well developed PBL by 11 AM; this is particularly evident for the 11 AM ACM2 CO profile, which is more well mixed than the other simualtions and observations. The median P-3B O_3 and NO_2 profiles are consistent with the development of the convective boundary layer indicated by potential temperature and CO. The NO_2 profiles transition from an approximately exponential decay shape, in which maximum mixing ratio values are found near the surface and sharply decay with altitude toward minimum values, to a more well mixed profile resembling that of O_3 or CO

during the afternoon, with a less steep vertical gradient in NO_2 mixing ratios within the PBL but retaining the sharp decrease in mixing ratio values at the top of the PBL and into the free troposphere. The O_3 are fairly well mixed by 10 AM, with some interruption at 12 PM and 1 PM, perhaps due to the mixing up of surface O_3 as the PBL matured, returning to well mixed by 2 PM. It is notable that the ACM2 scheme displayed too much O_3 between approximately 1 km and 2.5 km at 8 AM and 9 AM, which may be indicative that ACM2 mixed too much O_3 and O_3 precursors into this altitude layer on the previous day (which became the residual layer overnight and lasted through early morning).

None of the PBL schemes demonstrated a consistent ability throughout the diurnal cycle to best capture the shapes of the median *in situ* profiles or the magnitudes of the profiles for potential temperature, CO, and O_3. All schemes well captured the magnitudes of the observed potential temperature values thoroughout the diurnal cycle (Fig. 5.10); no scheme outperformed the others. The ACM2 was the scheme to most often overpredict these quantities, particularly during the afternoon after the convective boundary layer (CBL) had developed and throughout large portions of the profiles, including the uppper PBL and the free troposphere. This scheme also tended to exhibit median profiles that were too deeply mixed relative to the P-3B profiles and the other schemes during afternoon. It is notable that all five schemes did not over- or underpredict median potential temperature magnitudes except for 3 PM and 4 PM, as denoted by the overlap of the error bars of the schemes and the P-3B; all schemes overpredicted during these two times, which coincided with the mature phase of the CBL, and likely drove down the correlations in Fig. 5.10. This relative performance did not as consistently carry over to the comparison with the *in situ* CO median profiles. The ACM2 and YSU schemes most adequately captured the shape of the CO profiles at 11 AM, 12 PM and 3 PM, while the BouLac scheme most often compared best to

the shape of the *in situ* profiles after 1 PM. However, most schemes performed well relative to the magnitudes of the observed CO mixing ratios during the morning and middle of the afternoon (until 11 AM and 1-3 PM). The PBL schemes less well captured the CO mixing ratios during the beginning of the evening collapse of the PBL at 4-5 PM, which is a process which parameterizations are known to less well replicate. The correlation of the ACM2 simulated values with observed CO mixing ratio values was significantly lower than those for the other schemes at 1-2 PM (and at 9 AM; Fig. 5.11): ACM2 overpredicted CO mixing ratios, indicative of overly vigorous vertical mixing. It should be noted that all five schemes tended to overpredict the median CO mixing ratio profiles at various times throughout the day, particularly within the upper PBL and lower free tropospheric portions of the profiles; the ACM2 most consistently overpredicted relative to the observations, in agreement with the scatter plots. Somewhat consistently with the comparison to the P-3B CO profiles, the BouLac and QNSE schemes most often performed best of all five schemes relative to the shapes of the O_3 *in situ* median profiles between 11 AM and 5 PM, though the same cannot be said of the comparison of the magnitudes of these simulated and observed values, as the direction and magnitude of the model bias changed throughout the day (frequent poor correlations in Fig. 5.12). The PBL scheme-simulated and P-3B O_3 mixing ratios compared best at 3 PM (Fig. 5.12), during the mature phase of the CBL. As with CO, the ACM2 scheme tended to consistently overpredict relative to the P-3B median O_3 profiles within the upper PBL and lower free troposphere, as denoted by the error bars. Many schemes often missed structure in the P-3B *in situ* O_3 profiles.

The best comparsion among PBL schemes and *in situ* observations was obtained for NO_2, which is a surprising result, given known errors in NO_x emissions and simulated nitrogen chemistry. The BouLac scheme most consistently captured the shapes and magnitudes of the P-

3B median NO$_2$ at most hours of the day (Figs. 5.9 and 5.13), while the MYJ and YSU schemes also compared favorably during the afternoon. The ACM2 median NO$_2$ profiles, as expected, tended to be too deeply mixed, while the QNSE scheme missed the development of the "boot shaped" profile, in which NO$_2$ exhibited a more uniform mixing ratio value within the PBL, with a sharp decrease in magnitude above the PBL. This comparison to NO$_2$ contrasts with those of the other species, and suggests that a scheme with less intense vertical mixing is needed to adequately simulate this short-lived species, as the nonlocal YSU scheme did not consistently compare well thorughout the day to the P-3B NO$_2$ observations, and the ACM2 often overpredicted and mixed the median profile too deeply. The least under- or overprediction relative to the observations was also obtained relative to NO$_2$, though the ACM2 retained its tendency to overpredict during the afternoon hours in the upper PBL and lower free troposphere
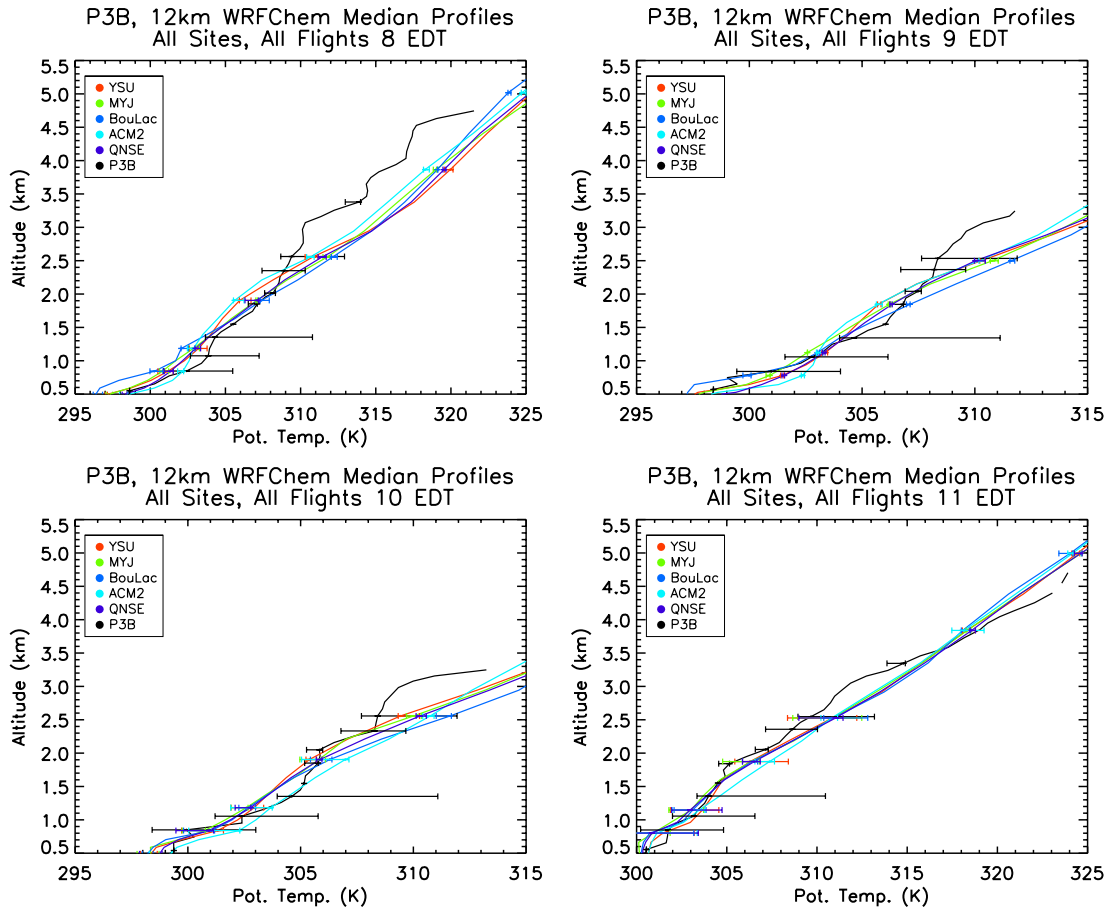
Fig. 5.6 a)-d): Hourly-median plots of potential temperature for the hours between 8 AM and 11 AM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. Example error bars represent the 25th and 75th percentile values at that altitude level. All available profiles included.
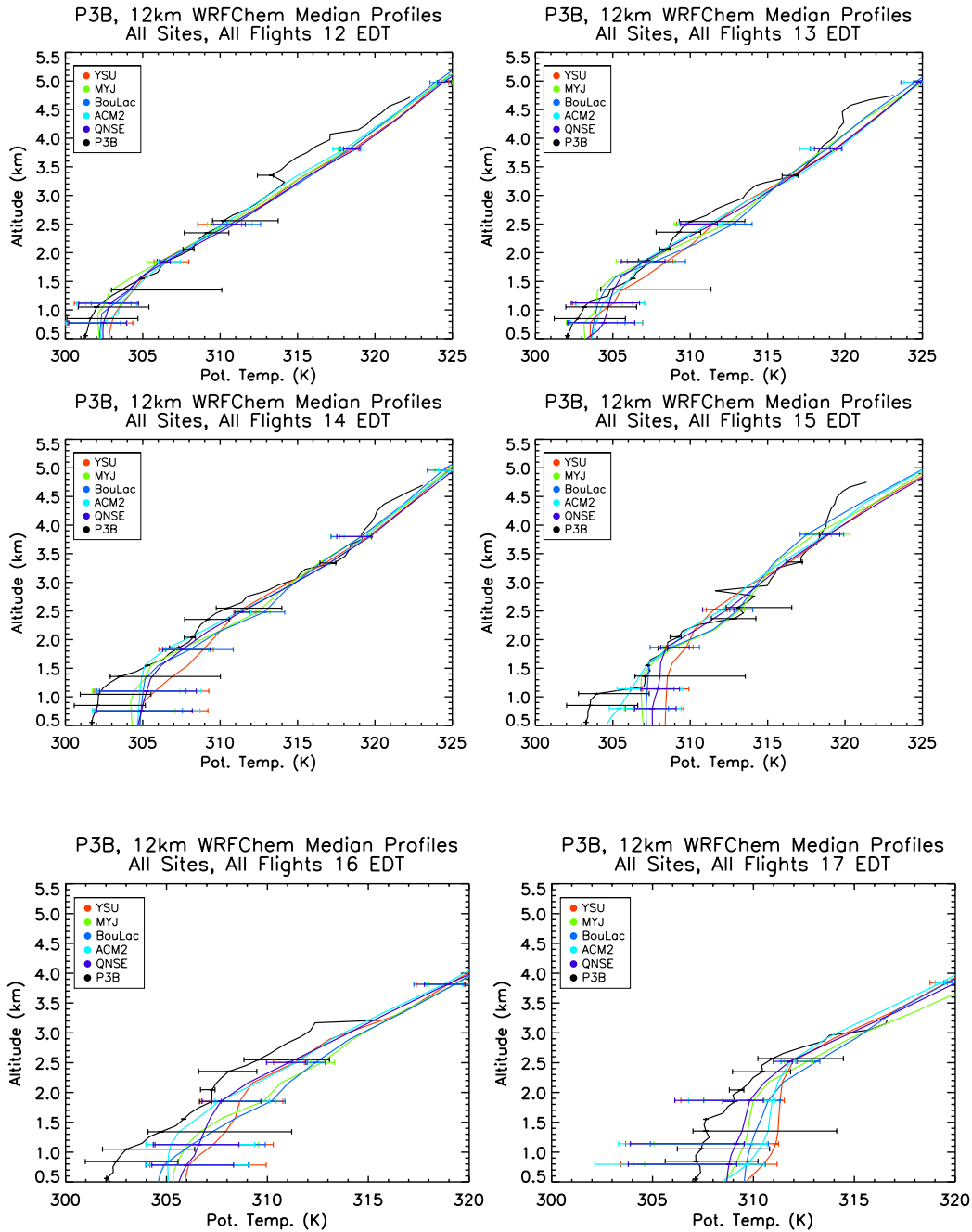
Fig. 5.6 e)-j): Hourly-median plots of potential temperature for the hours between 12 PM and 5 PM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. Example error bars represent the 25th and 75th percentile values at that altitude level. All available profiles included.
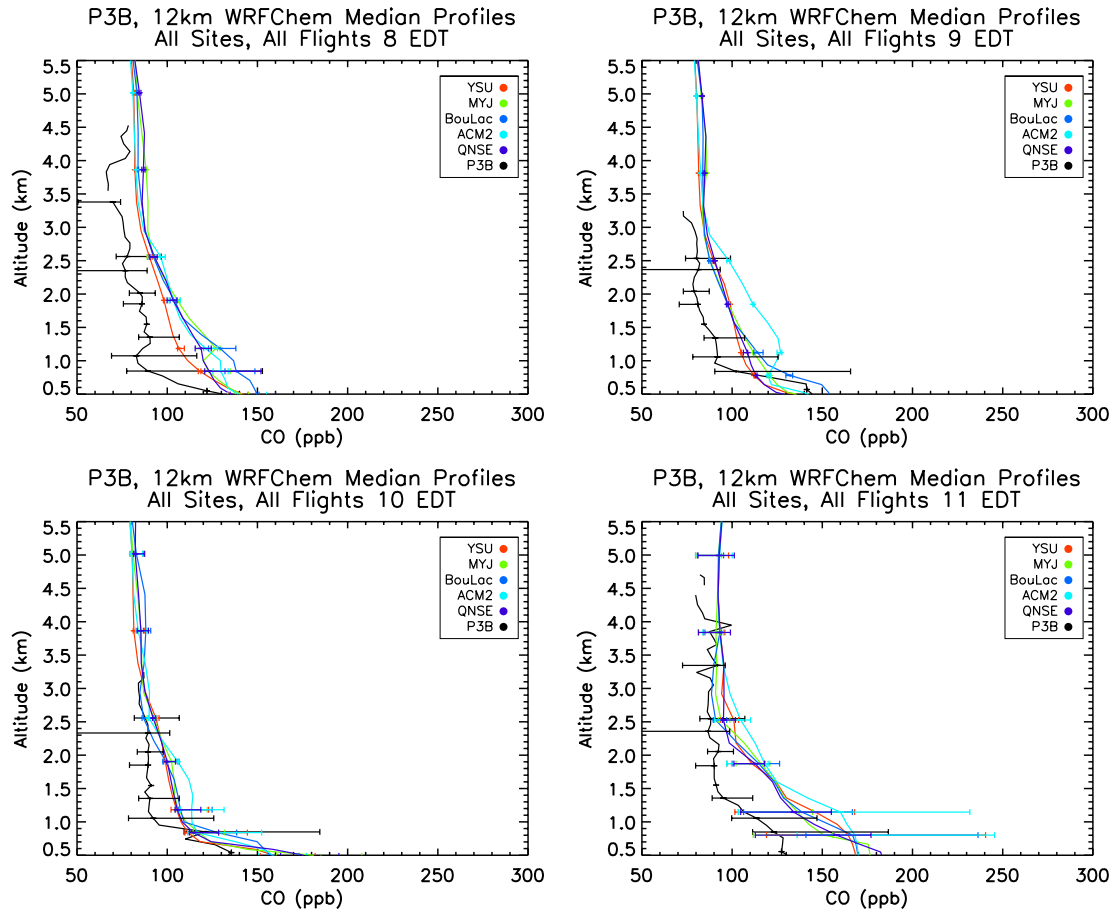
Fig. 5.7 a)-d): Hourly-median plots of CO for the hours between 8 AM and 11 AM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. Error bars represent the 25th and 75th percentile values at that altitude level. All available profiles included.
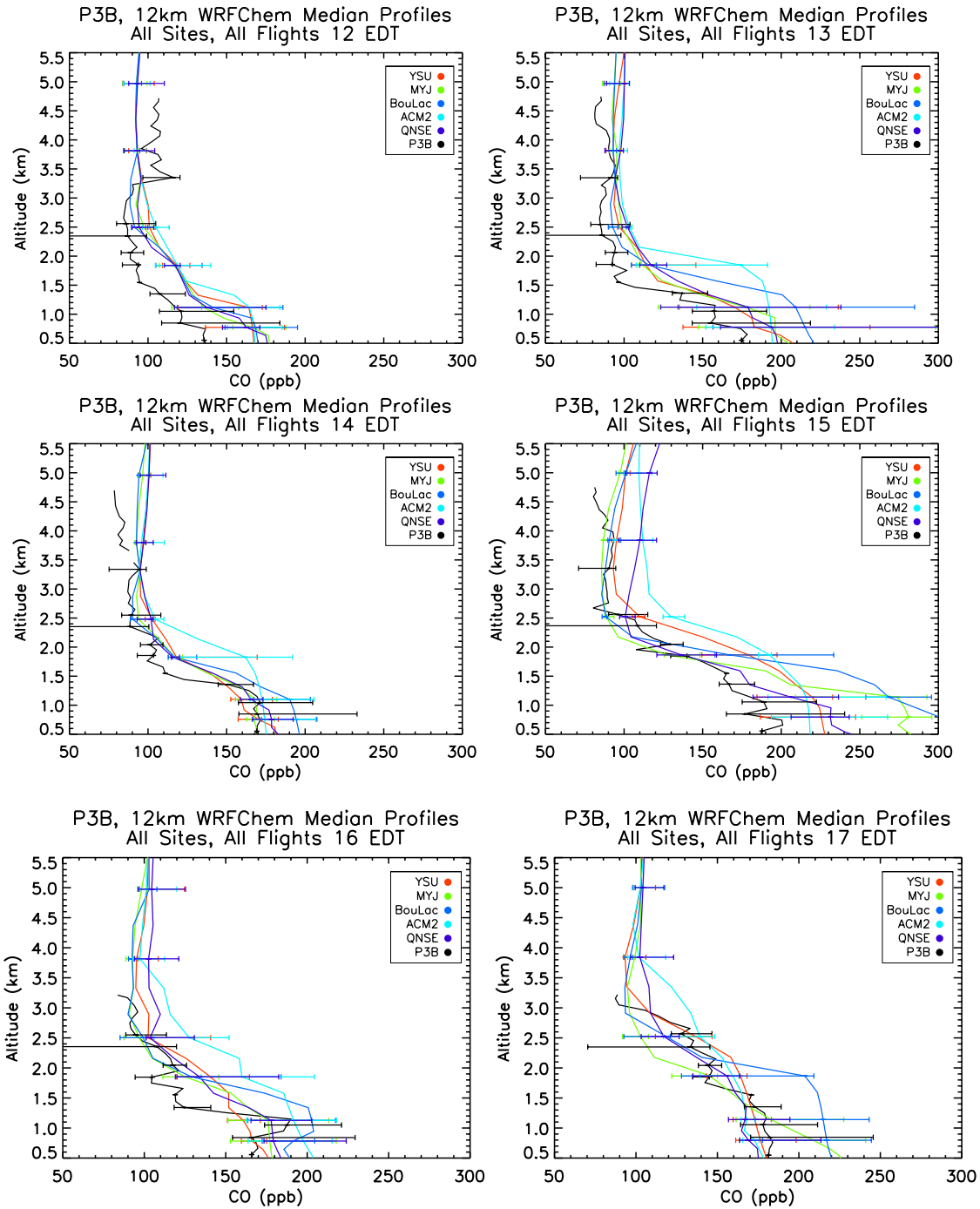
Fig. 5.7 e)-j): Hourly-median plots of CO for the hours between 12 PM and 5 PM EDT from P-3B *in situ* **measurements and WRF-Chem simulation output for each of the five PBL schemes. Error bars represent the 25th and 75th percentile values at that altitude level. All available profiles included.**
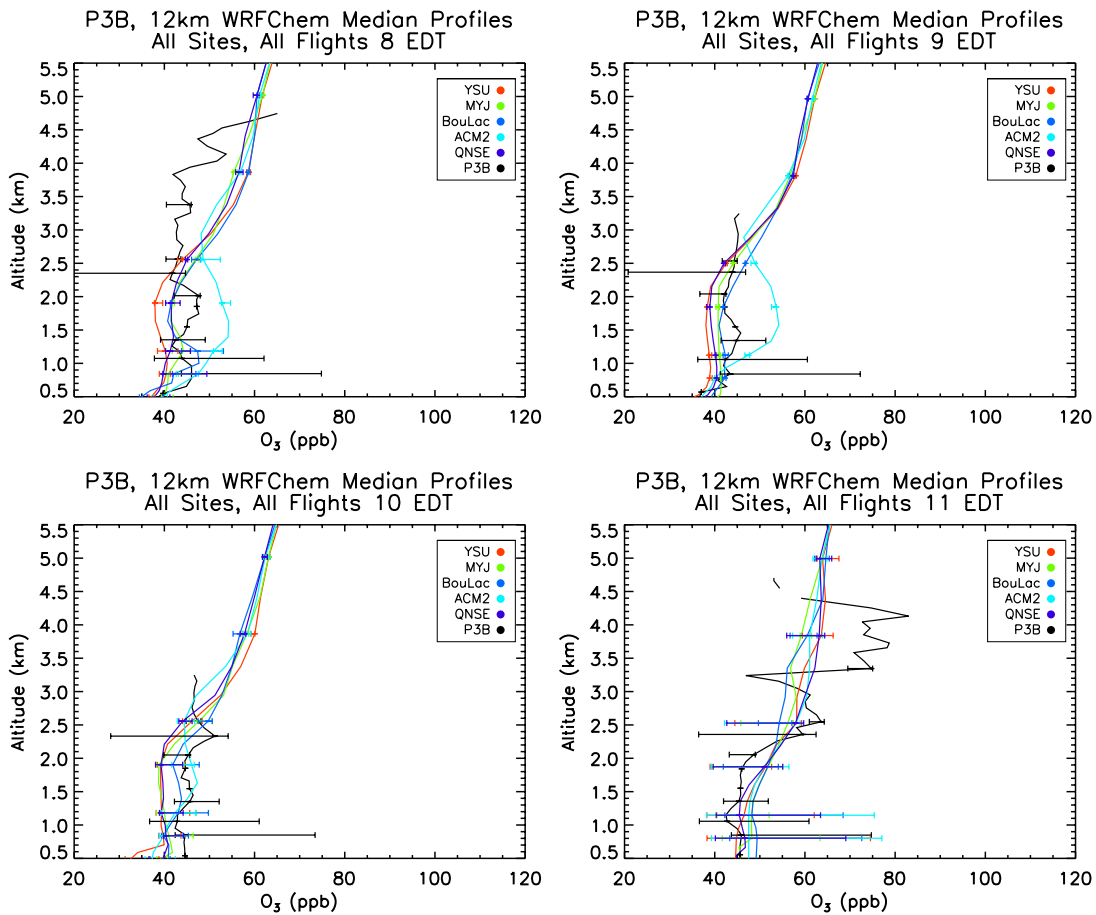
Fig. 5.8 a)-d): Hourly-median plots of O_3 for the hours between 8 AM and 11 AM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. Error bars represent the 25th and 75th percentile values at that altitude level. All available profiles included.
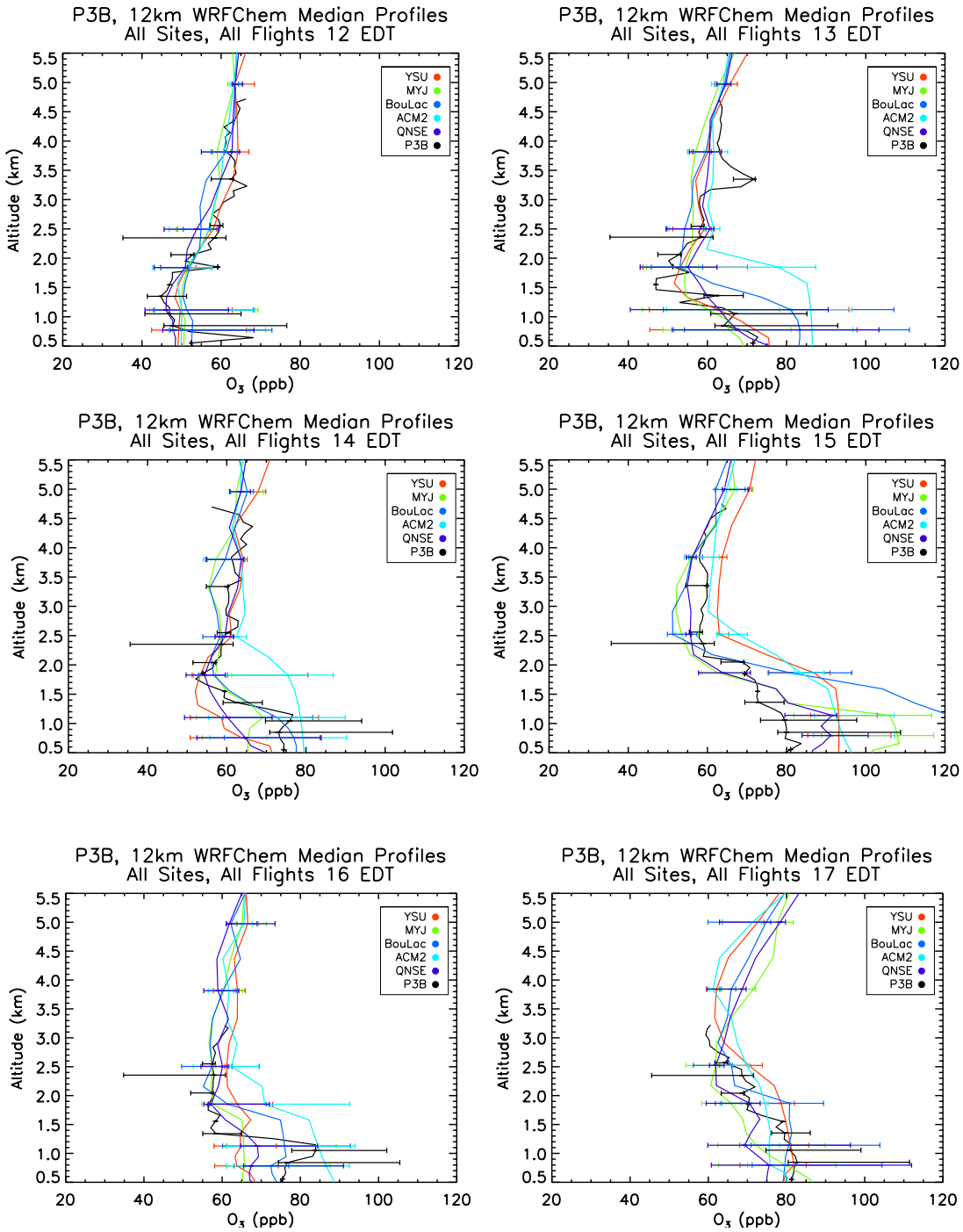
Fig. 5.8 e)-j): Hourly-median plots of O$_3$ for the hours between 12 PM and 5 PM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. Error bars represent the 25th and 75th percentile values at that altitude level. All available profiles included.
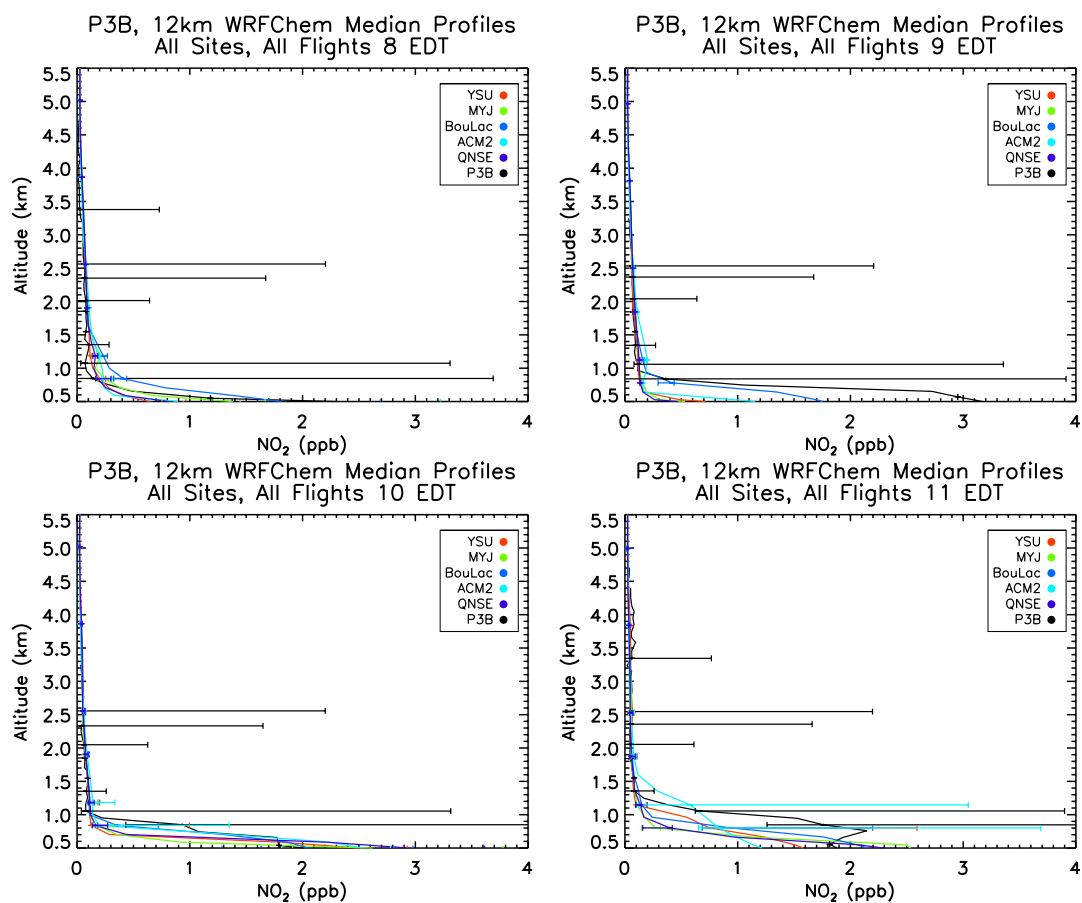
Fig. 5.9 a)-d): Hourly-median plots of NO$_2$ for the hours between 8 AM and 11 AM EDT from P-3B _in situ_ measurements and WRF-Chem simulation output for each of the five PBL schemes. Error bars represent the 25th and 75th percentile values at that altitude level. All available profiles included.
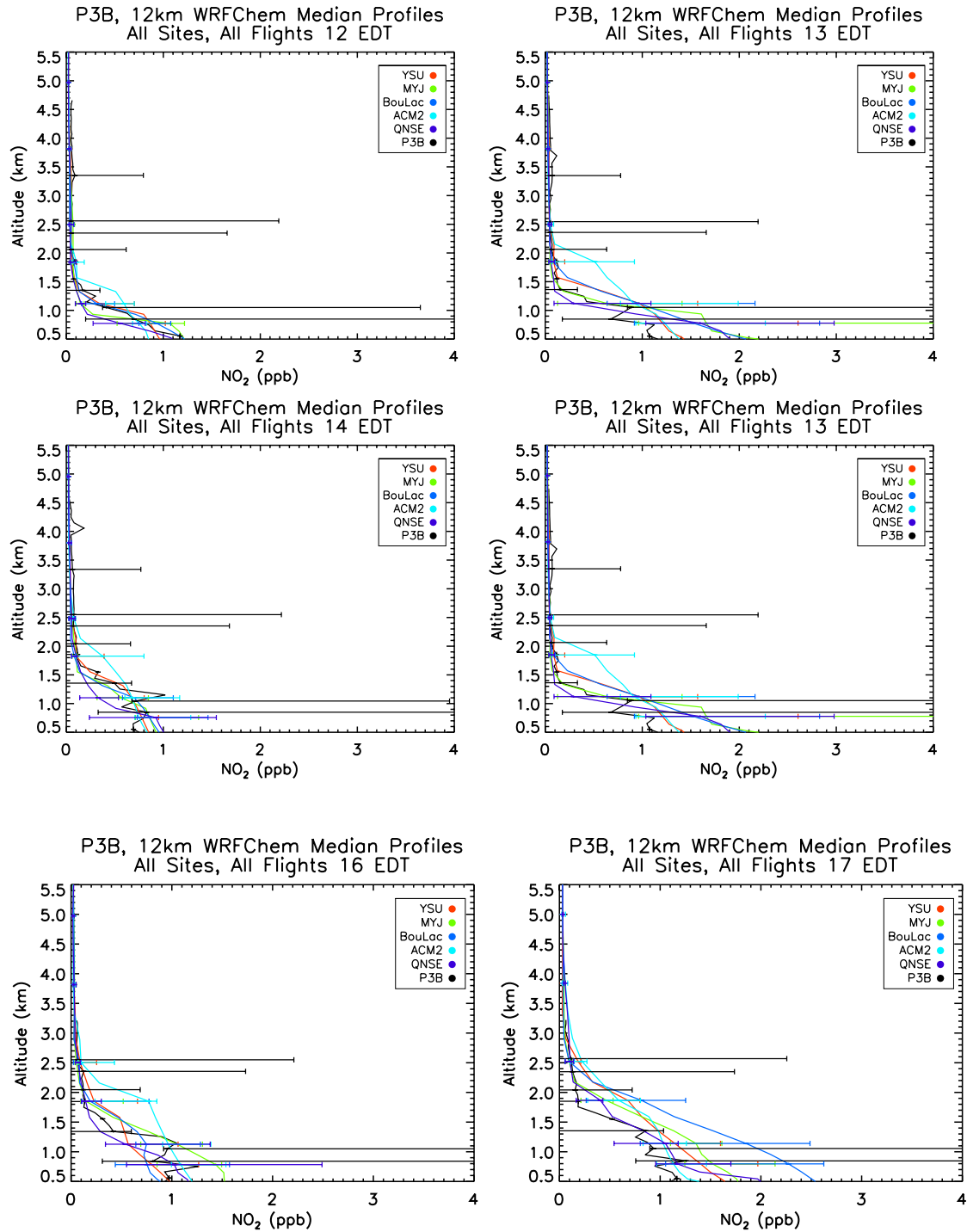
Fig. 5.9 e)-j): Hourly-median plots of NO₂ for the hours between 12 PM and 5 PM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. Error bars represent the 25ᵗʰ and 75ᵗʰ percentile values at that altitude level. All available profiles included.
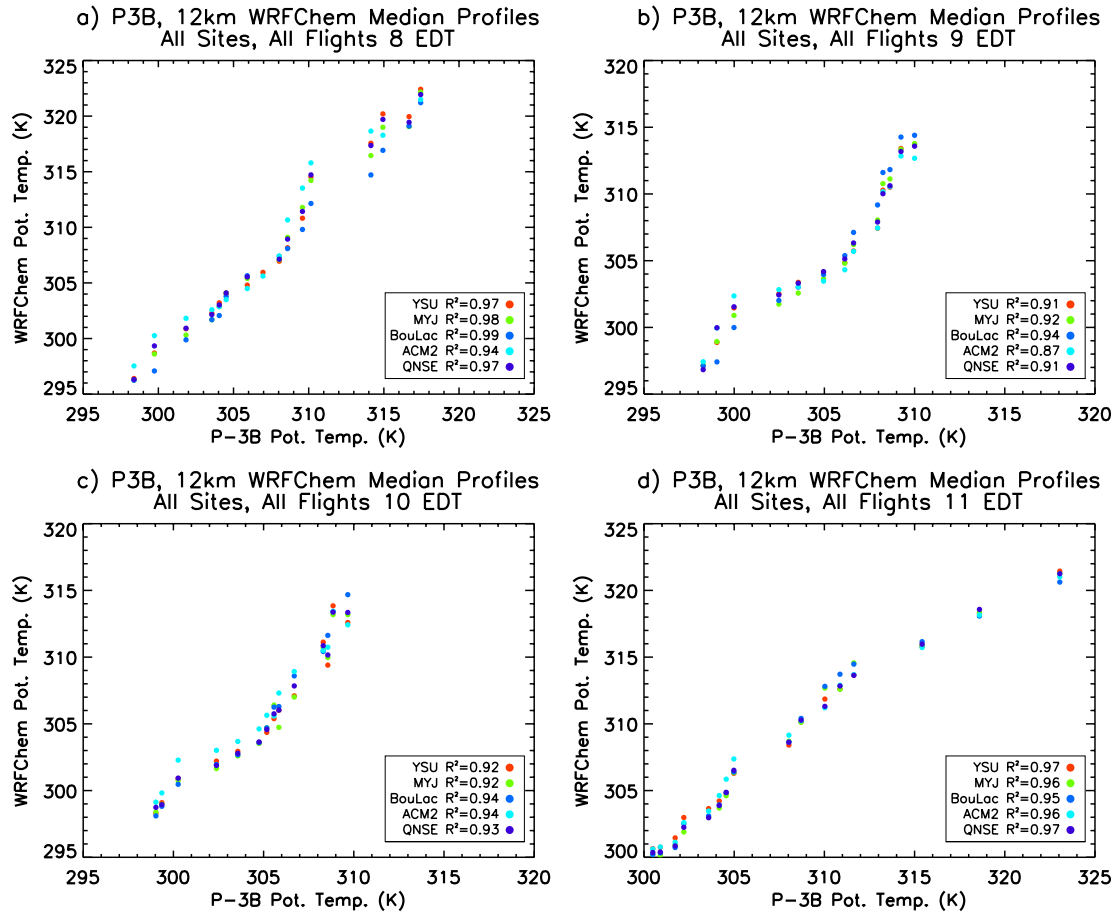
Fig. 5.10 a)-d): Hourly scatter plots of potential temperature for the hours between 8 AM and 11 AM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. All available profiles included.
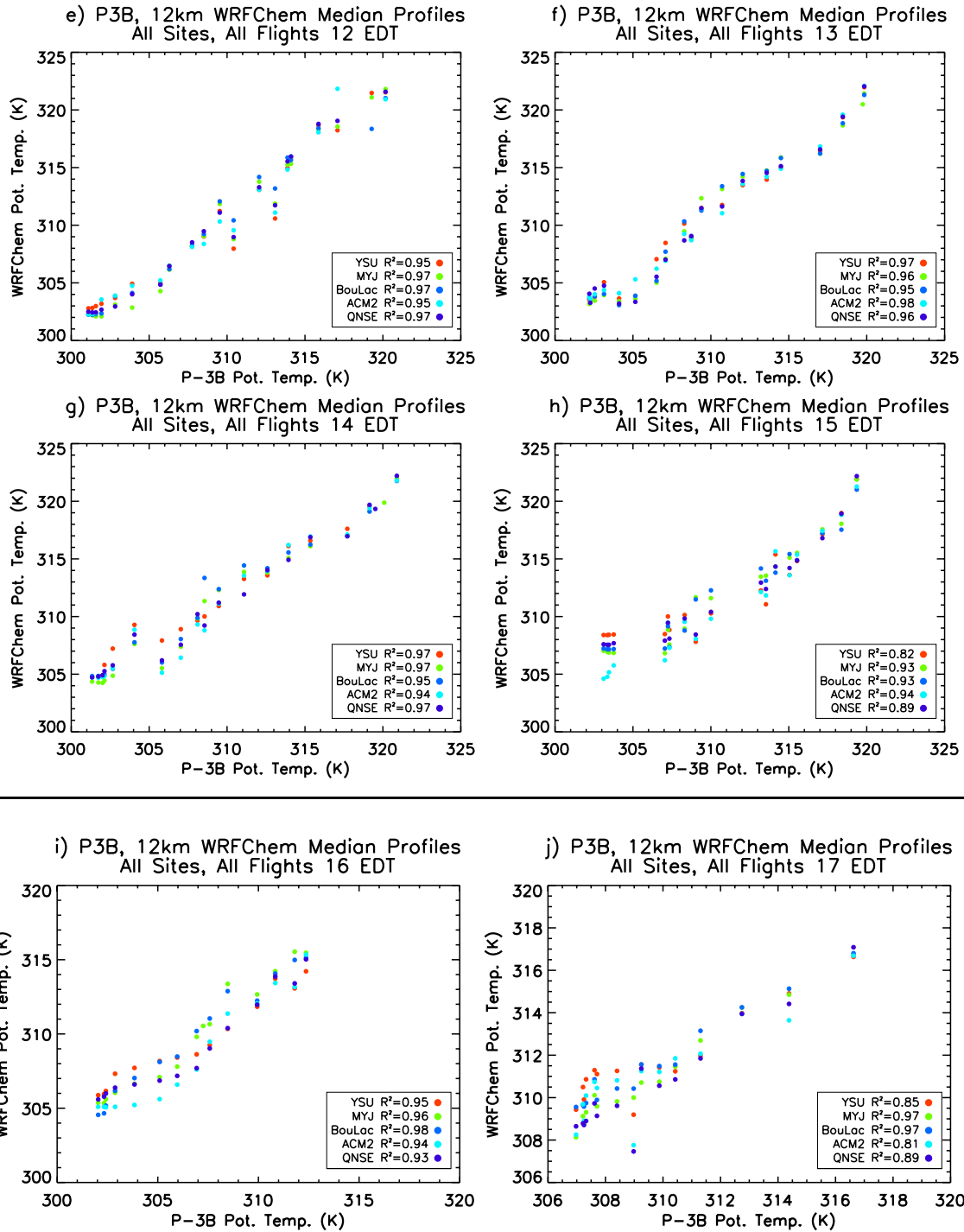
Fig. 5.10 e)-j): Hourly scatter plots of potential temperature for the hours between 12 PM and 5 PM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. All available profiles included.
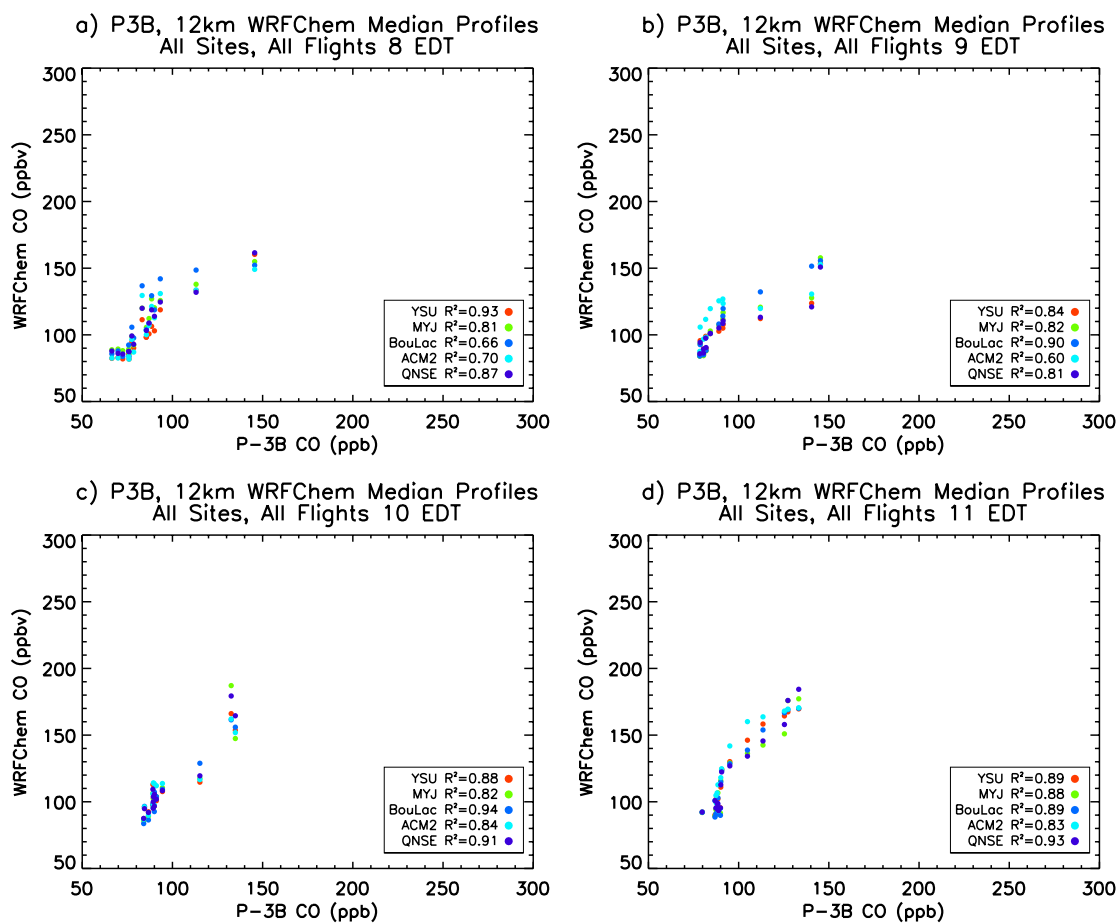
Fig. 5.11 a)-d): Hourly scatter plots of CO for the hours between 8 AM and 11 AM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. All available profiles included.
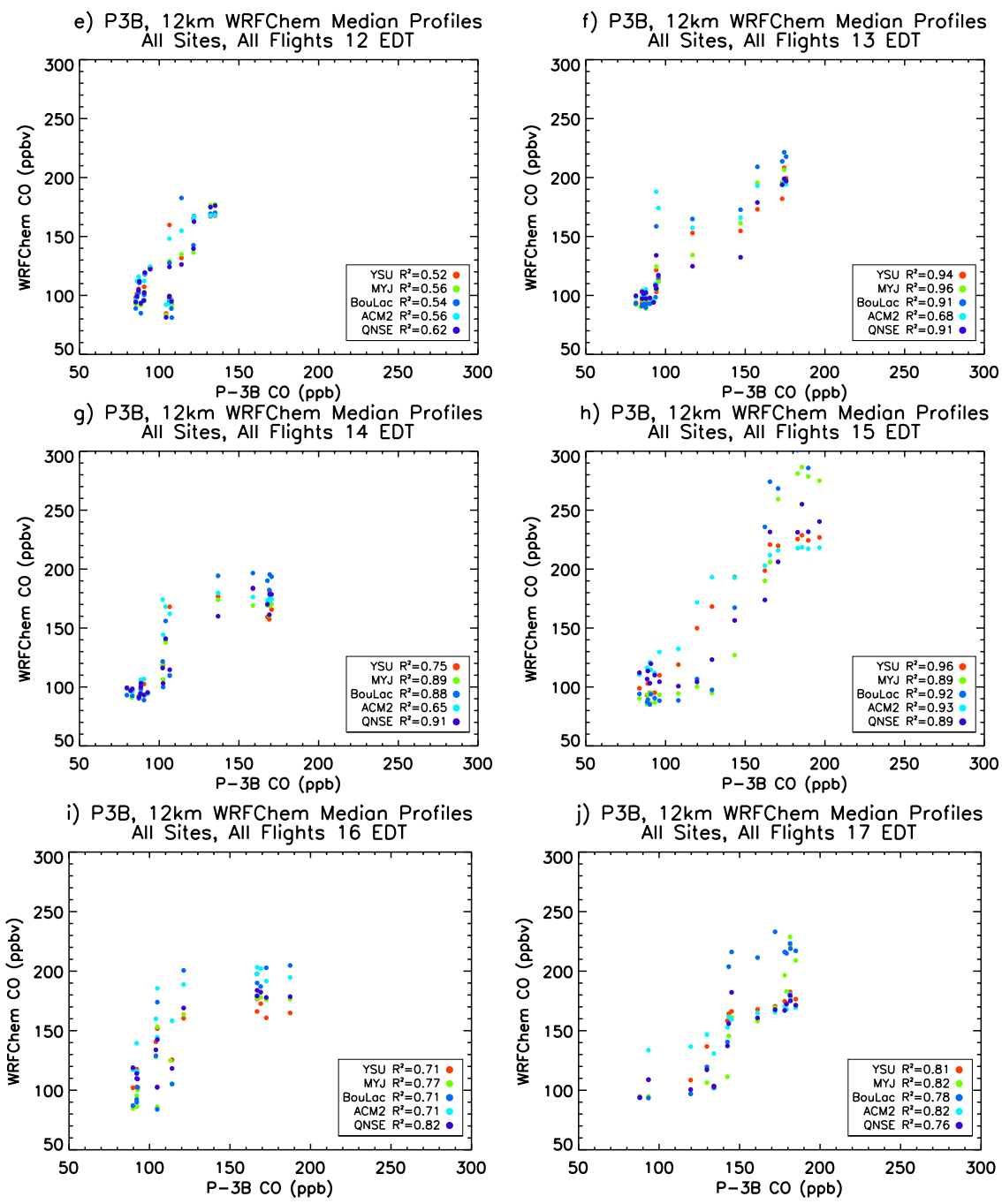
Fig. 5.11 e)-j): Hourly scatter plots of CO for the hours between 12 PM and 5 PM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. All available profiles included.
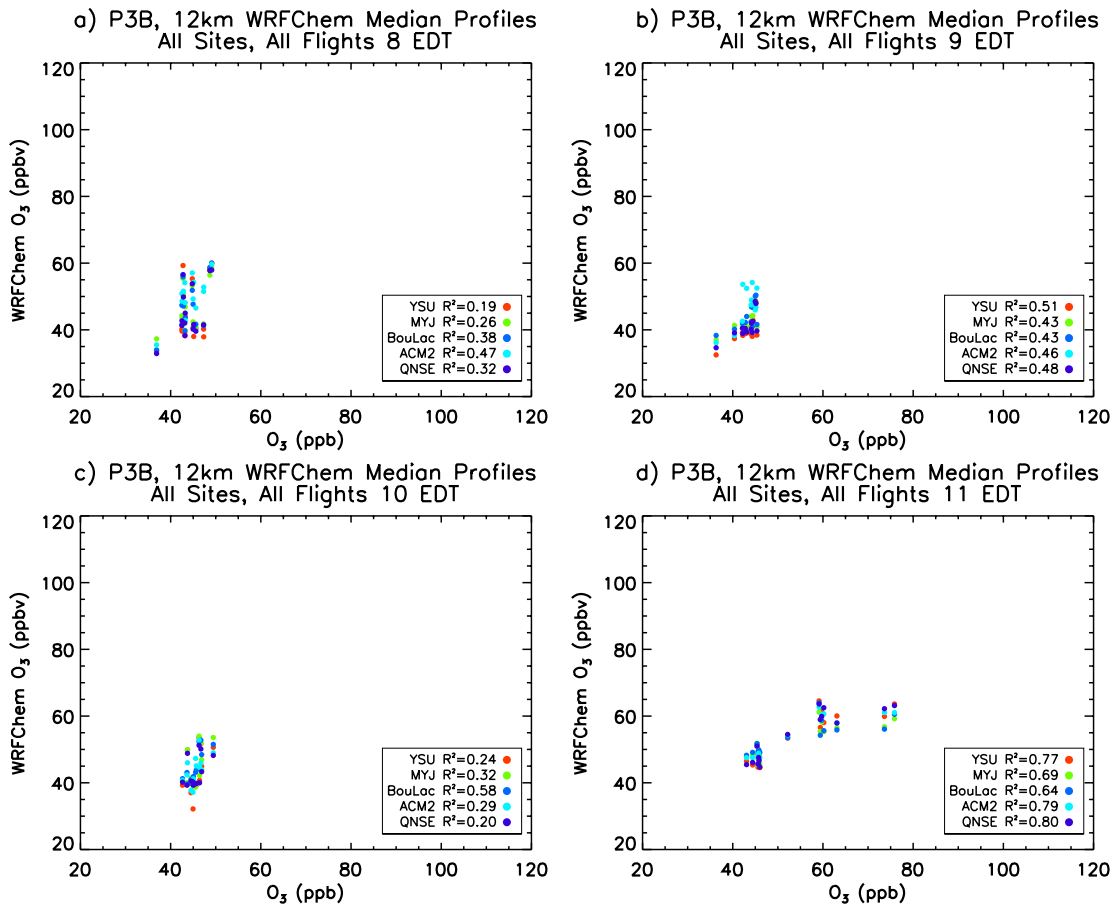
Fig. 5.12 a)-d): Hourly scatter plots of O_3 for the hours between 8 AM and 11 AM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. All available profiles included.
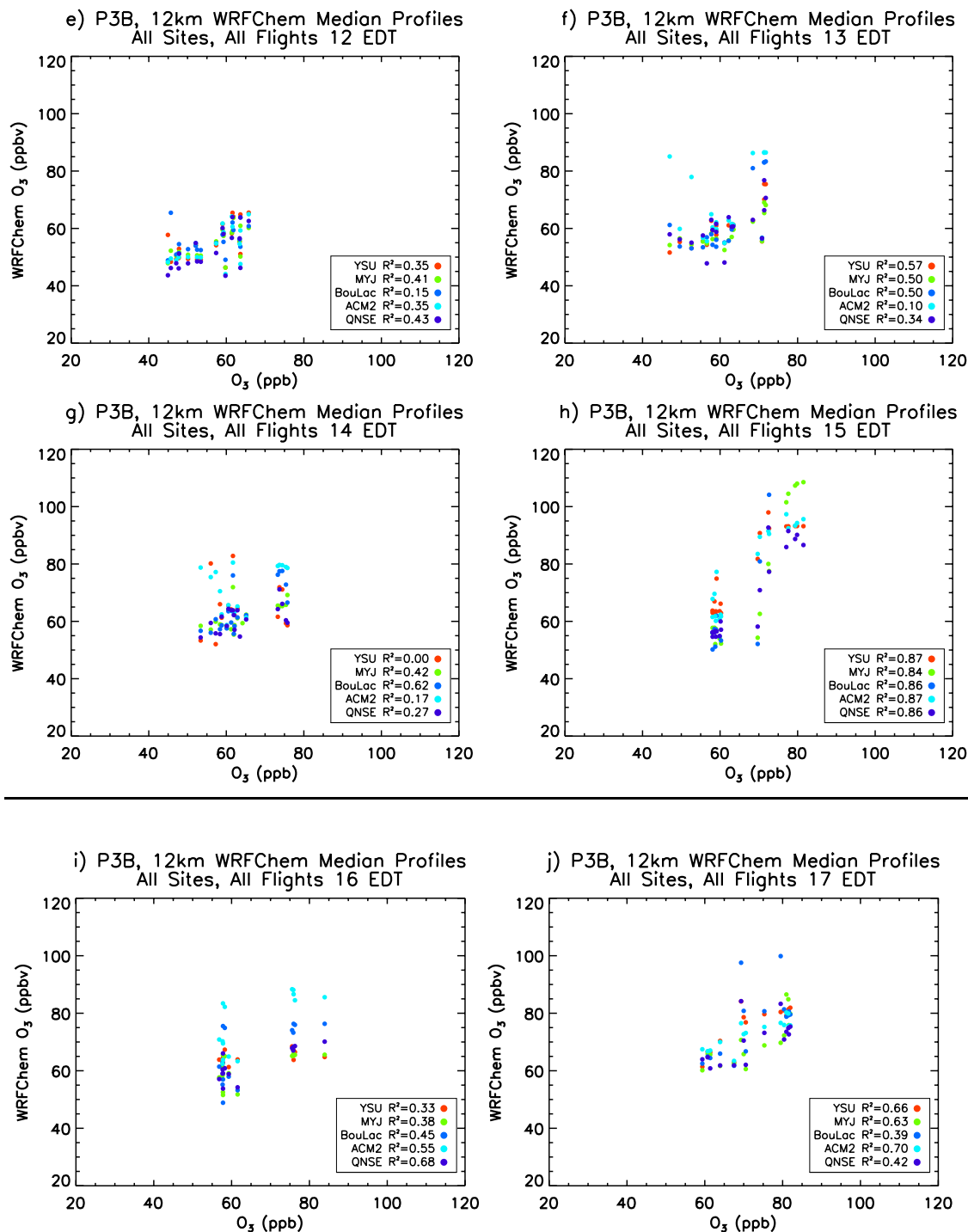
Fig. 5.12 e)-j): Hourly scatter plots of O₃ for the hours between 12 PM and 5 PM EDT from P-3B *in situ* **measurements and WRF-Chem simulation output for each of the five PBL schemes. All available profiles included.**
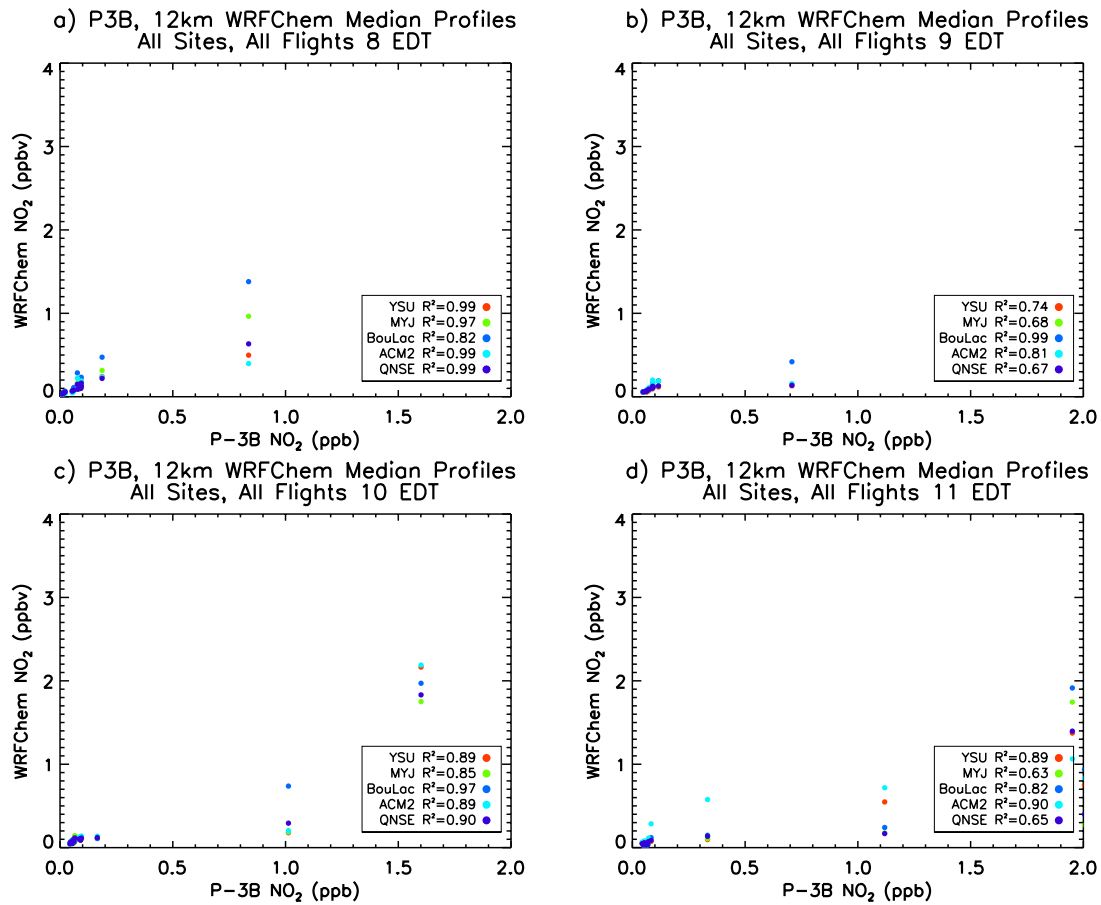
Fig. 5.13 a)-d): Hourly scatter plots of NO₂ for the hours between 8 AM and 11 AM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. All available profiles included.
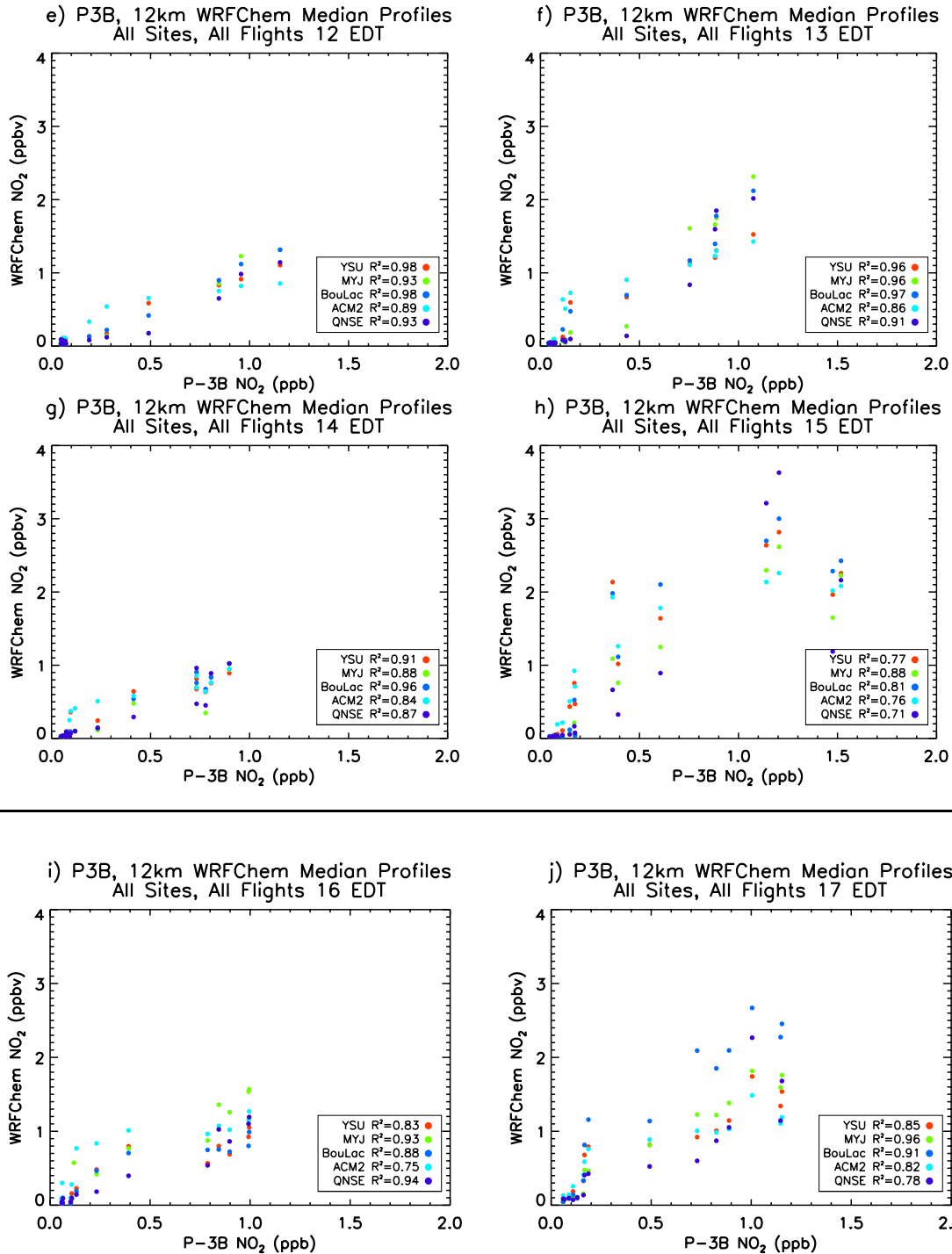
Fig. 5.13 e)-j): Hourly scatter plots of NO$_2$ for the hours between 12 PM and 5 PM EDT from P-3B *in situ* measurements and WRF-Chem simulation output for each of the five PBL schemes. All available profiles included.

5.8.2: P-3B and WRF-Chem Shape Factor Computation

Hourly median O_3 and NO_2 shape factors for the times between 8 am EDT and 5 pm EDT were computed over all six spiral sites for the P-3B observations and WRF-Chem simulated profiles. Simulated profiles most coincident in time and space to the P-3B profiles for each spiral site were sampled, to allow a direct comparison to the P-3B shape factor profiles. The shape factor was defined as the ratio of the partial column within a model vertical layer to the partial tropospheric column over the depth of the P-3B spiral:

$$S(z) = \Omega_z/\Omega_{trop} \tag{5.13}$$

where Ω_z denotes partial column amount at vertical layer z and Ω_{trop} denotes the partial tropospheric column over the full depth of the P-3B spirals. For both observations and model simulations, the partial columns were computed over the depth of each WRF-Chem vertical layer; thus, the vertical distribution was computed over the model vertical grid.

5.8.3: WRF-Chem Partial Tropospheric Column Amounts

To be able to compare the WRF-Chem column-surface correlations to those from the P-3B, partial tropospheric columns were computed from the WRF-Chem output for O_3 and NO_2 over the depths of the P-3B spirals. WRF-Chem profiles coincident to each P-3B profile were sampled, and the WRF-Chem levels below or above the lowest or highest P-3B measurement altitudes were excluded from the column computation. The O_3 or NO_2 simulated partial column amounts were then computed from integration of the simulated lower tropospheric profile.

5.8.4: P-3B and WRF-Chem O_3 and NO_2 Shape Factors

Overall, all five PBL schemes compared well to the median P-3B O_3 shape factors for each hour between 8 AM and 5 PM EDT, as shown in Fig. 5.14. No scheme emerged that produced a consistent, clear "best" comparison to the P-3B diurnal variation in shape factor.. The

values of R^2 for the correlation between observed and simulated shape factor values over all altitudes for each hour and the average orthogonal distance from the 1:1 line, displayed in Fig. 5.15, were similar for among all five schemes at each hour, while the R^2 values were greater than 0.75 for all schemes and all hours, indicating that each scheme well captured the observed shape factor magnitudes throughout the day. However, a small decrease in the values of R^2, accompanied by a small increase in the values of the average distance from the 1:1 line, can be discerned after 12 PM, suggesting that WRF-Chem was somewhat less able to reproduce the observed O_3 shape factors during afternoon, when the PBL had maturely developed. Additionally, all schemes tended to underpredict within the PBL relative to the P-3B at each hour except at 2 PM and 3 PM; each scheme best estimated the PBL magnitudes or slightly overpredicted at these two times. These also corresponded to the mature phase of the CBL, suggesting that local and nonlocal schemes captured the PBL shape factor profile shapes best when the boundary layer is most deeply and turbulently mixed. However, all five schemes tended to overpredict relative to the P-3B in the free troposphere, except between 1 PM and 3PM (mature phase of the CBL) when the schemes tended to underpredict, though no scheme tended to most greatly over- or underpredict relative to the other schemes at any altitude level. As with the comparison of the CMAQ O_3 lower tropospheric shape factors to the P-3B cluster shape factors presented in Chapter 3, these results indicate that WRF-Chem placed too large a portion of the O_3 column relative vertical distribution higher in the vertical relative than seen in the observations. The shape factor comparisons presented here indicate an advantage of a coupled meteorology-chemistry model over the traditional offline model of the WRF/CMAQ model system: each PBL scheme used in WRF-Chem was able to replicate the vertical locations of shape factor local maxima and minima, with few exceptions, such as the erroneous free tropospheric peak in the simulated shape factors

at 11 AM or the overestimated free tropospheric peak at 1 PM. CMAQ, which uses ACM2 exclusively for boundary layer mixing (regardless of the PBL scheme employed in the WRF simulation that is used to drive CMAQ) often placed the peak O_3 shape factor values too high in the vertical relative to the P-3B (Chapter 3). This further suggests that the ACM2 scheme performed better in the online model than in the traditional offline CMAQ model.

The correlations between the P-3B and WRF-Chem O_3 shape factor magnitudes for each PBL scheme over all available shape factor profiles for all hours supports the results of the comparison of the diurnal variation of the observed and simulated shape factors presented above. Values of R^2 were 0.85-0.88 for all five schemes (Fig. 5.16a)-e), a high degree of correlation, indicating that the model adequately represented the shape factor magnitudes. However, the scatter plots of simulated vs. observed shape factor values display a consistent underprediction relative to the P-3B within the lowermost 800 m of the shape factors across all schemes (Fig. 5.16a)-e). Though not as dramatic, a compensating overprediction is apparent for shape factors values above 1200 m altitude, and some overprediction between 800 m and 1200 m altitude. These patterns in shape factor bias with altitude are consistent with biases suggested by the comparison of the shapes of the simulated and observed shape factor profiles. The low bias within the lowermost portions of the shape factors and high bias within the upper portions of the shape factors indicate that these PBL schemes placed a greater portion of the O_3 column relative vertical distribution higher in the vertical than seen in the observations. This is likely due to a combination of errors in the NEI anthropogenic emissions used in these simulations (as detailed in Anderson et al., 2014; Canty et al., 2015; and Goldberg et al, 2016), and overestimated vertical mixing of trace gas species within each PBL scheme. Scatter plots of WRF-Chem vs. P-3B CO shape factors demonstrate a similar pattern of underestimated CO shape factor magnitudes below

800 m, with overestimated magnitudes between 1200 m and 2000 m (Fig. 5.17). The YSU, ACM2, and QNSE also overpredicted the CO shape factor magnitudes above 2000 m. The values of R^2 and average perpendicular distance from the 1:1 line (displayed on each plot in Fig. 5.17) were similar to those for the O_3 shape factor comparisons, displayed in Fig. 5.16, indicating that each PBL scheme overall well captured the magnitudes of the observed CO shape factors. Thus, the underestimation in the lower PBL and overestimation in the upper PBL (and in the lower free troposphere for YSU, ACM2, and QNSE) relative to the observed CO shape factors suggest that vertical mixing is overly vigorous within WRF-Chem, and impacted the relative vertical placements of the O_3 and CO partial column amounts within the model. However, it is emphasized that the median O_3 shape factor comparisons as well as the correlations between simulated and P-3B shape factor values indicate that a regional air quality model such as WRF-Chem may be able to replicate the observed lower tropospheric shape factors, consistently with the results of Chapter 3.

Each scheme likely well replicated the shape factors throughout the day because, firstly, O_3 is a well mixed trace gas (chemical lifetime longer than the timescale of turbulent mixing). Further, the shape factor computation is based directly on partial column amounts, and indirectly on the *in situ* mixing ratio profile. The shape factor is a measure of the relative vertical distribution of pollutant mass rather than an absolute distribution, as the *in situ* profiles are. Errors in mixing ratio above the PBL do not produce as much error in the shape factor as errors in mixing ratio within the PBL, due to lower air density in the free troposphere. With the moderately adequate comparison of the simulated and observed *in situ* O_3, differences in density lead to a better ability of the model to capture O_3 column amounts and partial column vertical distribution than the ability to capture the *in situ* mixing ratio profiles.
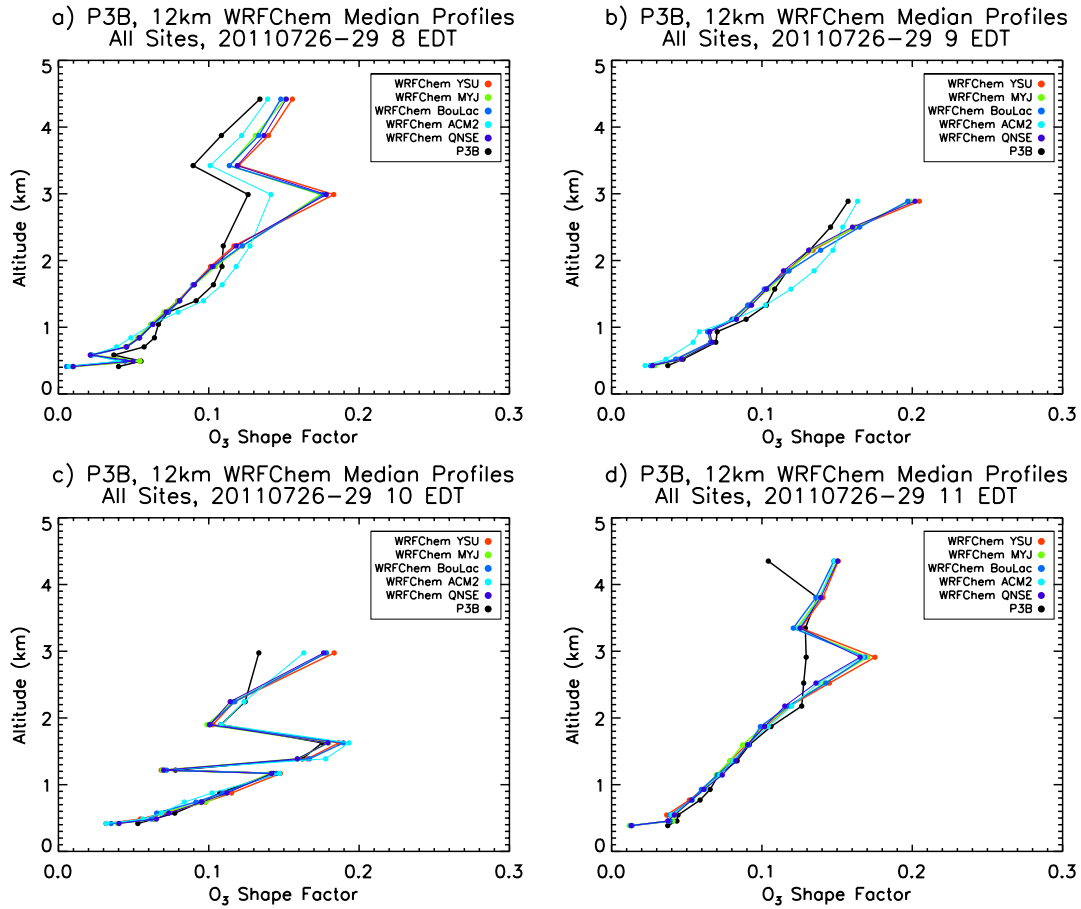
Fig. 5.14: a)-d) Hourly-median plots of O$_3$ shape factors for the hours between 8 AM and 11 AM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 8 AM and 11 AM. All available profiles included.
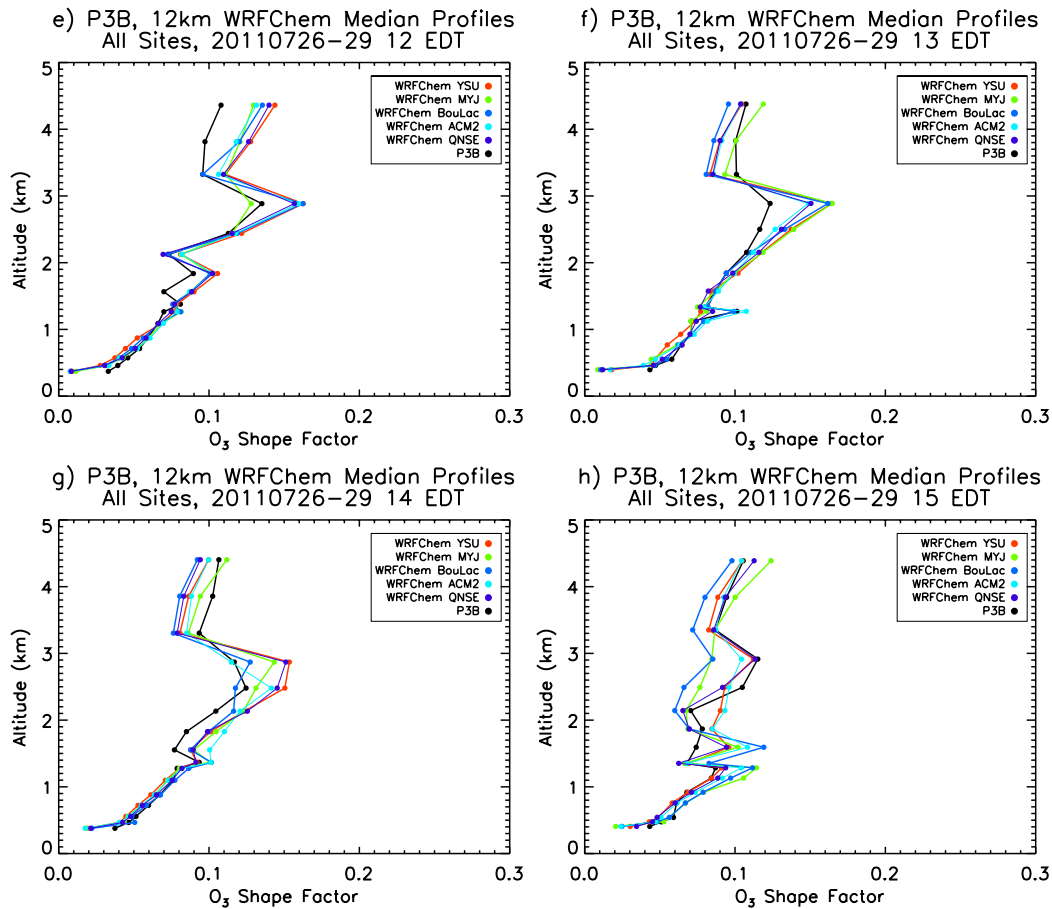
Fig. 5.14: e)-h) Hourly-median plots of O₃ shape factors for the hours between 12 PM and 3 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 12 PM and 3 PM. All available profiles included.

Fig. 5.14: i)-j) Hourly-median plots of O$_3$ shape factors for the hours between 4 PM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 4 PM and 5 PM. All available profiles included.

Fig. 5.15: a)-d) Scatter plots of O$_3$ shape factor magnitudes for the hours between 8 AM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 8 AM and 11 AM. All available profiles included. R^2 for the correlation between simulated and observed shape factor values and average orthogonal distance from the 1:1 line displayed in bottom right.

213

Fig. 5.15: e)-h) Scatter plots of O$_3$ shape factor magnitudes for the hours between 8 AM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 12 PM and 3 PM. All available profiles included. R^2 for the correlation between simulated and observed shape factor values and average orthogonal distance from the 1:1 line displayed in bottom right.

Fig. 5.15: i)-j) Scatter plots of O₃ shape factor magnitudes for the hours between 8 AM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 4 PM and 5 PM. All available profiles included. R² for the correlation between simulated and observed shape factor values and average orthogonal distance from the 1:1 line displayed in bottom right.
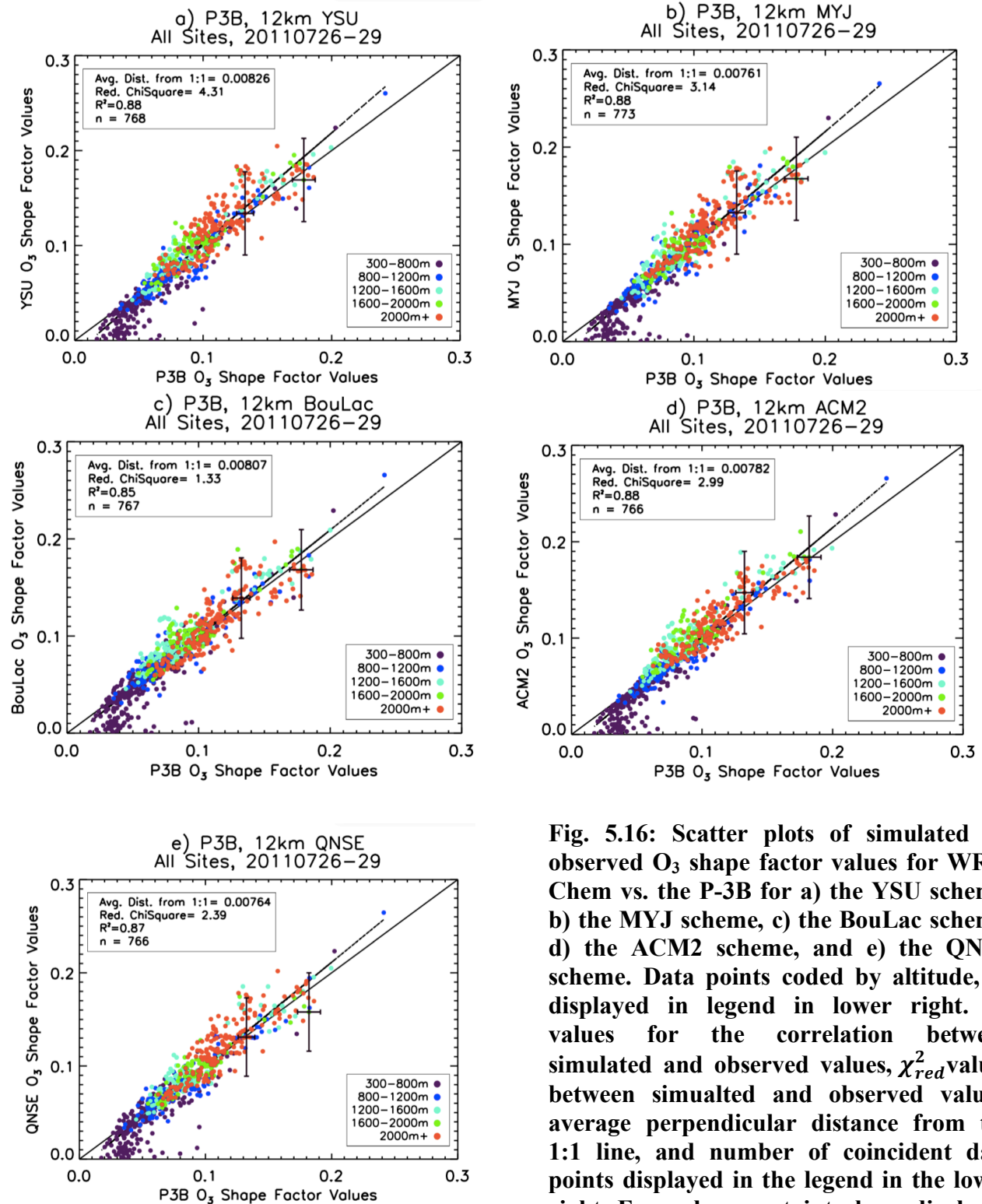
Fig. 5.16: Scatter plots of simulated vs. observed O₃ shape factor values for WRF-Chem vs. the P-3B for a) the YSU scheme, b) the MYJ scheme, c) the BouLac scheme, d) the ACM2 scheme, and e) the QNSE scheme. Data points coded by altitude, as displayed in legend in lower right. R^2 values for the correlation between simulated and observed values, χ^2_{red} values between simualted and observed values, average perpendicular distance from the 1:1 line, and number of coincident data points displayed in the legend in the lower right. Example uncertainty bars displayed on median data point in red: uncertainty in measurement used for observed PBLH and standard deviation used for simulated PBLH.

a) P3B, 12km YSU
All Sites, 20110726-29

Avg. Dist. from 1:1= 0.00804
Red. ChiSquare= 6.97
R^2=0.78
n = 763

300-800m
800-1200m
1200-1600m
1600-2000m
2000m+

b) P3B, 12km MYJ
All Sites, 20110726-29

Avg. Dist. from 1:1= 0.00822
Red. ChiSquare= 2.76
R^2=0.74
n = 767

300-800m
800-1200m
1200-1600m
1600-2000m
2000m+

c) P3B, 12km BouLac
All Sites, 20110726-29

Avg. Dist. from 1:1= 0.00803
Red. ChiSquare= 2.50
R^2=0.73
n = 762

300-800m
800-1200m
1200-1600m
1600-2000m
2000m+

d) P3B, 12km ACM2
All Sites, 20110726-29

Avg. Dist. from 1:1= 0.00892
Red. ChiSquare= 7.79
R^2=0.74
n = 760

300-800m
800-1200m
1200-1600m
1600-2000m
2000m+

e) P3B, 12km QNSE
All Sites, 20110726-29

Avg. Dist. from 1:1= 0.00819
Red. ChiSquare= 1.88
R^2=0.74
n = 760

300-800m
800-1200m
1200-1600m
1600-2000m
2000m+
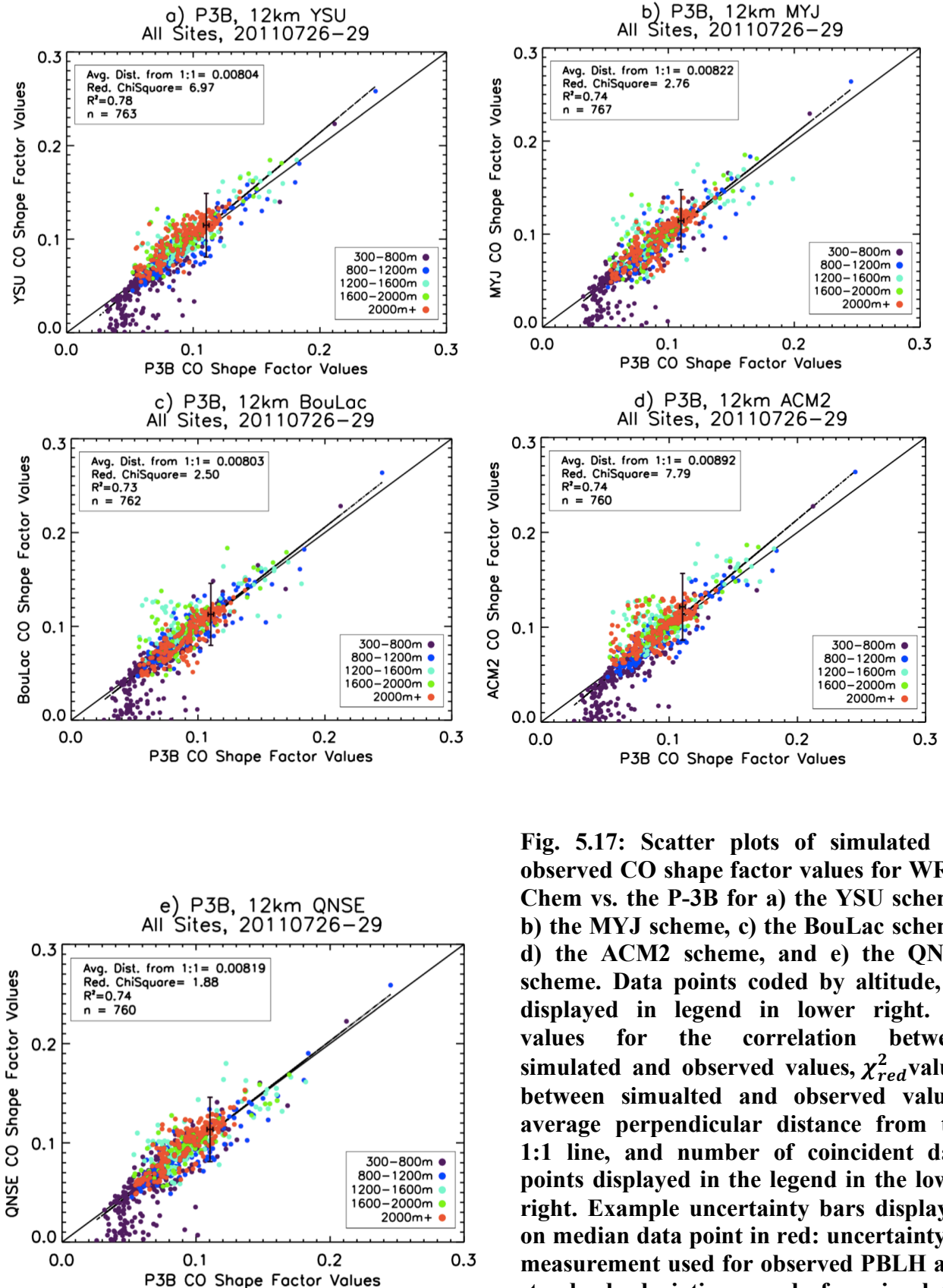
Fig. 5.17: Scatter plots of simulated vs. observed CO shape factor values for WRF-Chem vs. the P-3B for a) the YSU scheme, b) the MYJ scheme, c) the BouLac scheme, d) the ACM2 scheme, and e) the QNSE scheme. Data points coded by altitude, as displayed in legend in lower right. R^2 values for the correlation between simulated and observed values, χ^2_{red} values between simualted and observed values, average perpendicular distance from the 1:1 line, and number of coincident data points displayed in the legend in the lower right. Example uncertainty bars displayed on median data point in red: uncertainty in measurement used for observed PBLH and standard deviation used for simulated PBLH.

217

The five PBL schemes struggled to replicate the hourly median NO_2 P-3B shape factor profiles (Fig. 5.18). The ACM2 scheme tended to too deeply mix the NO_2 shape factor in the vertical relative to the P-3B and the other schemes during the afternoon, between 11 AM and 5 PM, leading to overestimations. No scheme consistently compared well to the observations throughout the day, as seen in the hourly shape factor scatter plots (Fig. 5.19). However, despite YSU overestimating shape factors above the PBL, this scheme and the BouL.ac scheme well captured the shape of the shape factor profiles during the early afternoon (12-2 PM). During late afternoon, it is notable that the ACM2 produced the best comparison at 4 PM, as it was the only scheme to replicate the sharp decrease in NO_2 shape factor values at approximately 1.2 km, while the QNSE scheme well captured the shape of the shape factor profiles at 3 PM and 5 PM. The MYJ scheme typically did not compare as well as the other schemes to the P-3B. The shape factors indicate model biases varying with time of day and altitude for NO_2. All schemes tended to overpredict in the upper free tropospheric portions of the shape factors except between 12 PM and 4 PM, while the five schemes tended to overpredict in the upper PBL/lower free troposphere at most hours of the day relative to the P-3B. No clear patterns of model bias emerged for the lower PBL, though bias is certainly evident in the comparison plots.

The correlations between simulated and observed NO_2 shape factor values over all available profiles over all hours are consistent with the mediocre comparison of the simulated and observed shape factor profiles. As with O_3, each PBL scheme produced a similar correlation with the observations (R^2 values between 0.32 and 0.42; Fig. 5.20a)-e), indicating that no scheme clearly outperformed the others. This is also a weak to moderate degree of correlation. The scatter

plots (Fig. 5.20a)-e) of simulated and observed shape factor values also display much more scatter than did the O_3 scatter plots for each PBL scheme, with average perpendicular distance values of approximately 0.02 for NO_2 instead of values of approximately 0.008 for O_3, further indicating that each scheme struggled to capture the shape factor magnitudes at most vertical levels. The greater scatter also prevented patterns of model bias from emerging that were as clear as those for O_3, though patterns are present. Each scheme tended to overpredict above 1200 m altitude; the ACM2, BouLac, and YSU schemes tended to most noticeably overpredict at these altitudes. It is also notable that the ACM2 scheme most significantly overpredicted above 1200 m, relative to the other schemes. All schemes displayed a tendency to underpredict between 800 m and 1200 m. The amount of scatter in the comparison of simulated and observed shape factor values increases below 1200 m, and is greatest in the lowermost 800 m of the shape factors, such that no clear tendencies to under- or overpredict in the lowermost shape factor profiles emerged; all schemes struggled most in the lowermost portion of the profiles, and often wildly missed the observed shape factor magnitudes. Part of this error in NO_2 shape factor magnitudes may be due to errors in representation of NO_2 and other NO_y species within the chemical mechanism, as well as errors in the anthropogenic emissions (Anderson et al., 2014; Canty et al., 2015; and Goldberg et al, 2016). However, consistent with the simulated O_3 and CO shape factors, the tendency to overpredict the magnitudes in the upper portions of the shape factors indicates that each scheme placed a greater portion of the NO_2 column relative vertical distribution higher in the vertical than seen in the observations. This suggests that overestimated vertical mixing of trace gases within each scheme also played a role in the errors in the simulated NO_2 partial column vertical distribution. The overestimation of NO_2 shape factor values above 1200 m in the ACM2 scheme particularly

suggests the impact of overly vigorous simulated vertical mixing on the NO_2 column amounts, as well as overestimated mixing within this particular scheme.

These results are not encouraging that any one scheme may best represent NO_2 shape factor profiles, and are at odds with the comparison of the simulated and observed *in situ* NO_2 volume mixing ratio profiles. This may be for one of the same reasons that WRF-Chem well replicated the O_3 shape factors: vertical mixing. The short-lived NO_2 is not as well mixed horizontally, as much of the NO_2 burden in the PBL is located near emissions sources, or as well mixed vertically as O_3 or CO, as its chemical lifetime is of a similar order of magnitude as the time to turbulently mix the PBL (Zhang et al., 2016), and vertical mixing does not exert a primary influence on NO_2 mixing ratio profile shape variability (Chapter 3), impairing the ability of the shape factor computation to well replicate the observed NO_2 partial column relative vertical distribution and leading to a better comparison to the *in situ* NO_2 mixing ratio profiles. Additionally, the PBL schemes tended to overpredict the NO_2 shape factors within the upper PBL/lower free troposphere; this is the region of the atmosphere to which satellite instruments are more sensitive for NO_2 relative to near the surface, and requires accurate simulation of shape factors (see Chapter 3). Thus, these errors in the WRF-Chem simulated shape factors would lead to greater retrieval error than similar shape factor errors located near the surface.
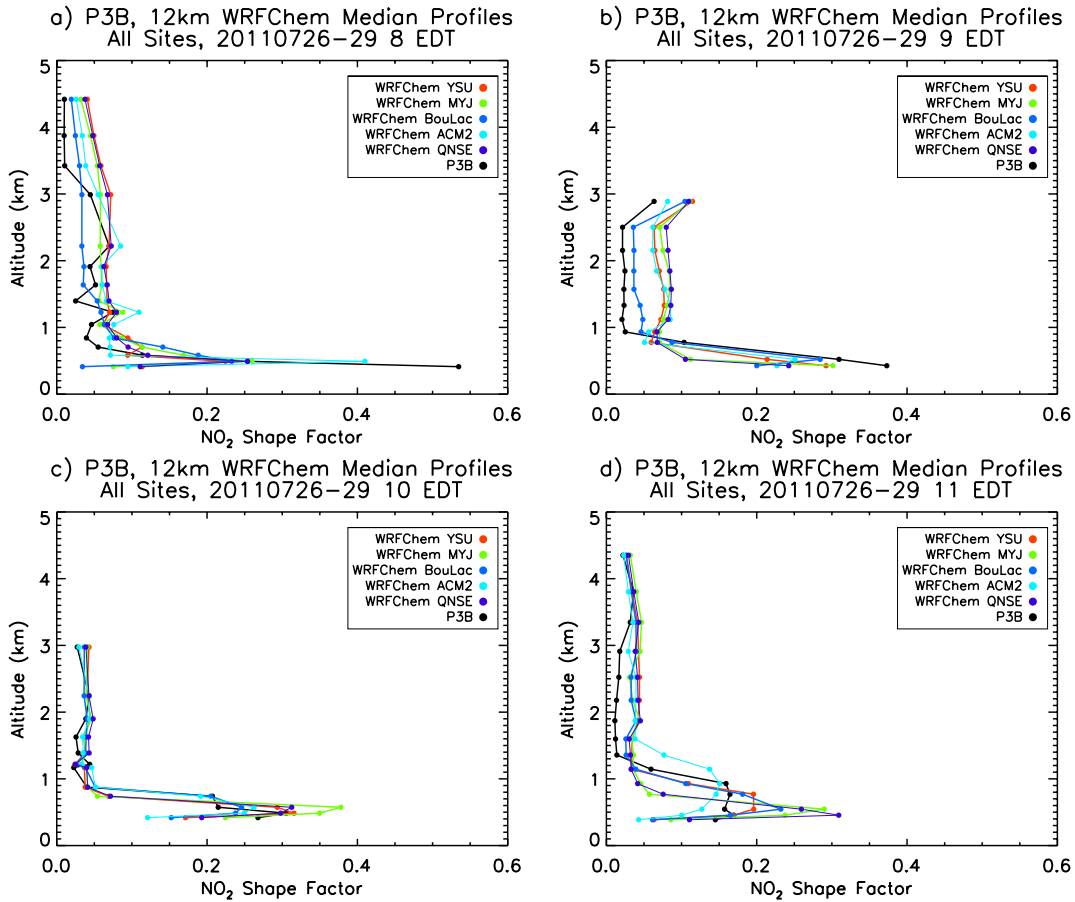
Fig. 5.18 a)-d): Hourly-median plots of NO₂ shape factors for the hours between 8 AM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 8 AM and 11 AM. All available profiles included.
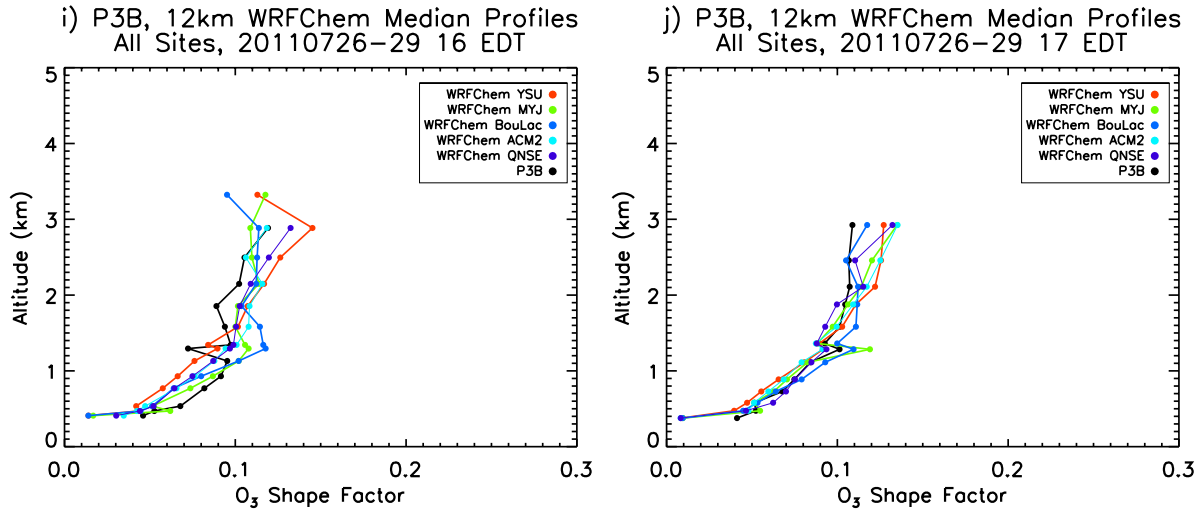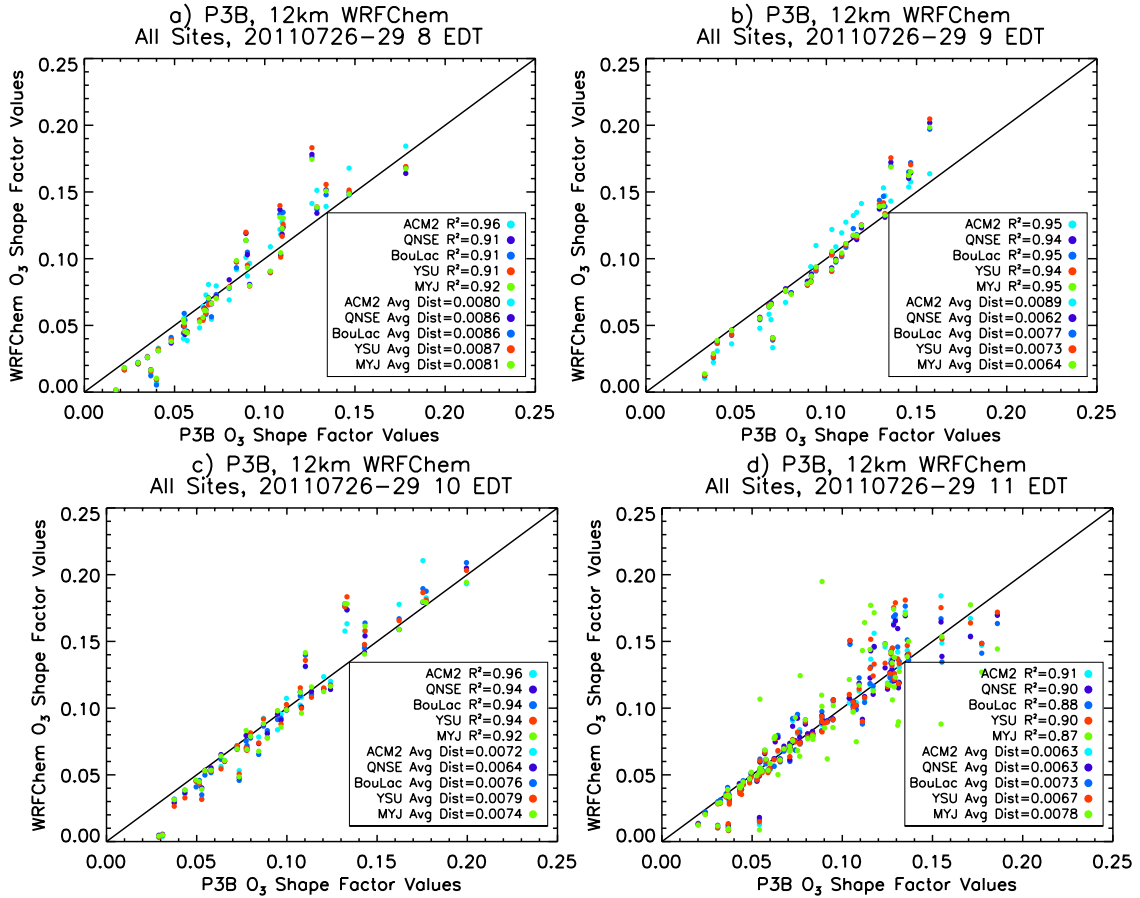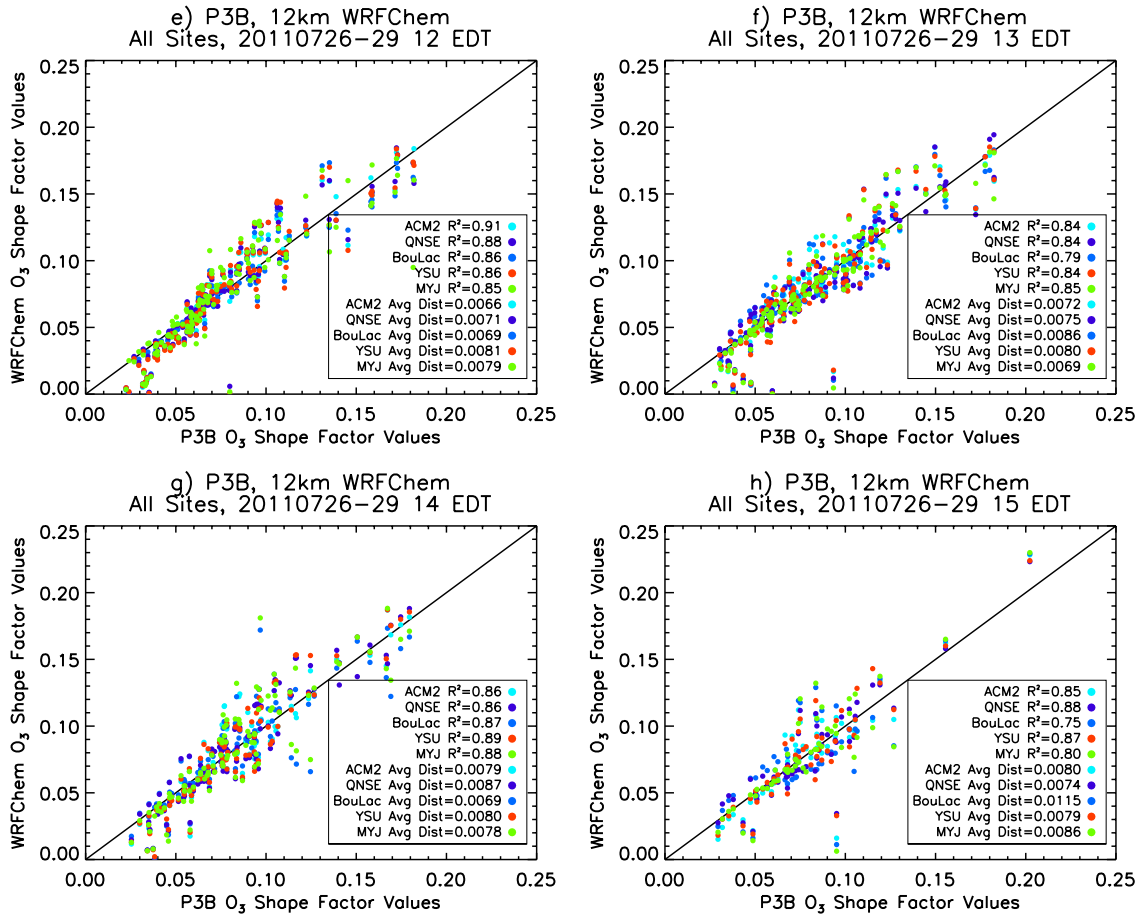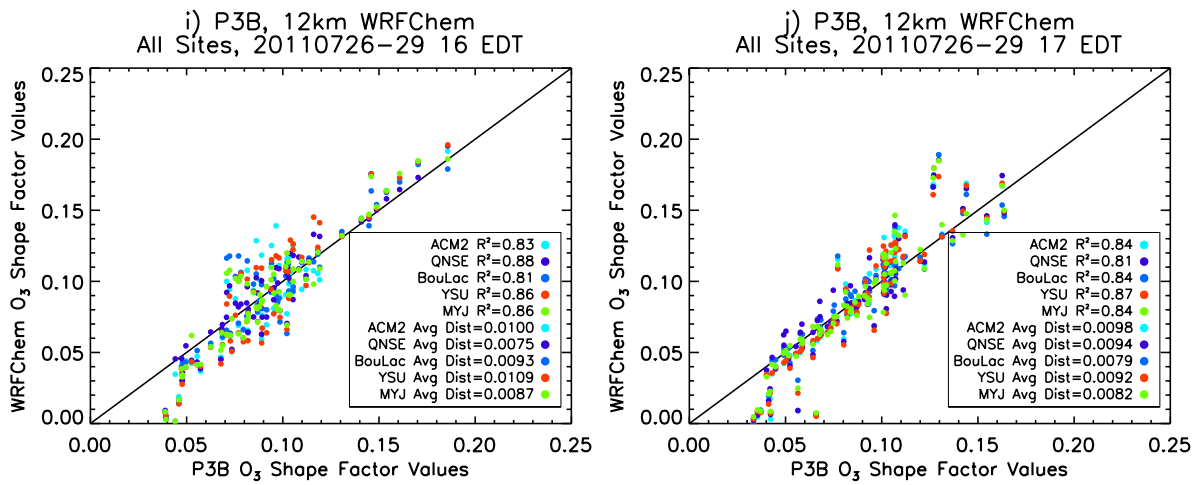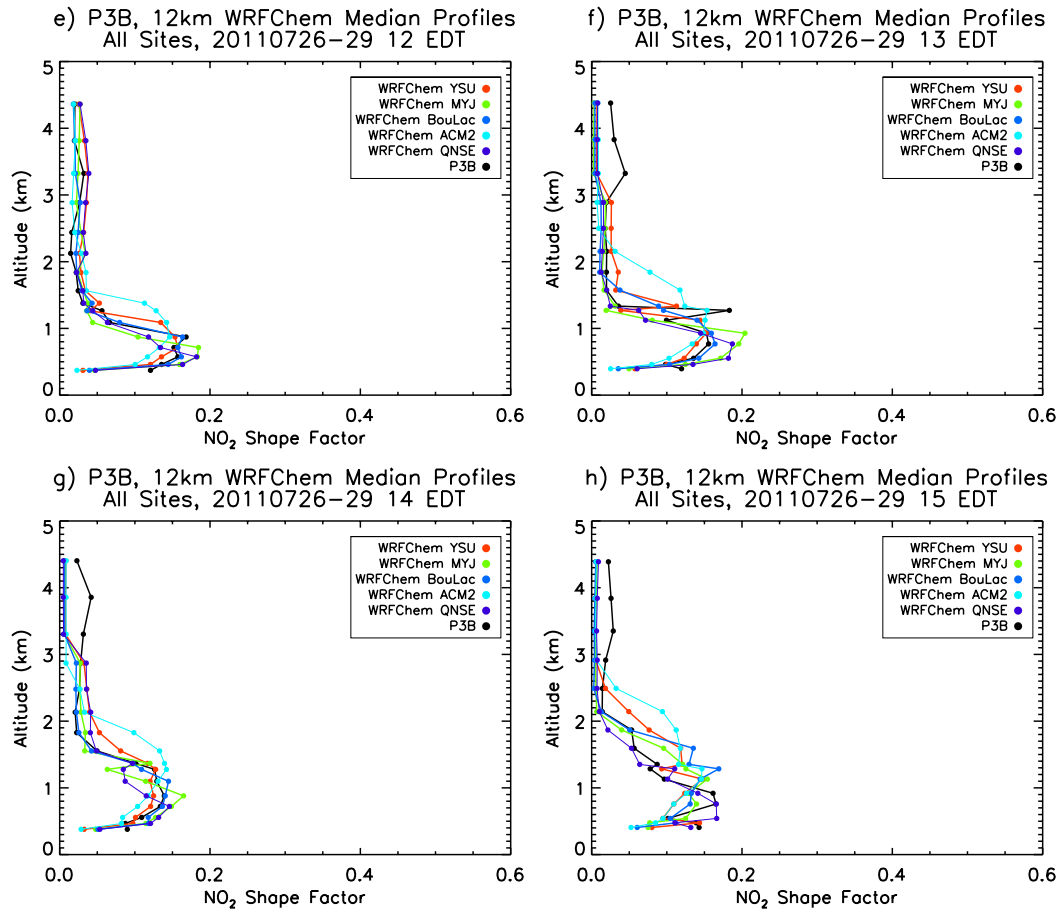
Fig. 5.18 e)-h): Hourly-median plots of NO₂ shape factors for the hours between 8 AM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 12 PM and 3 PM. All available profiles included.
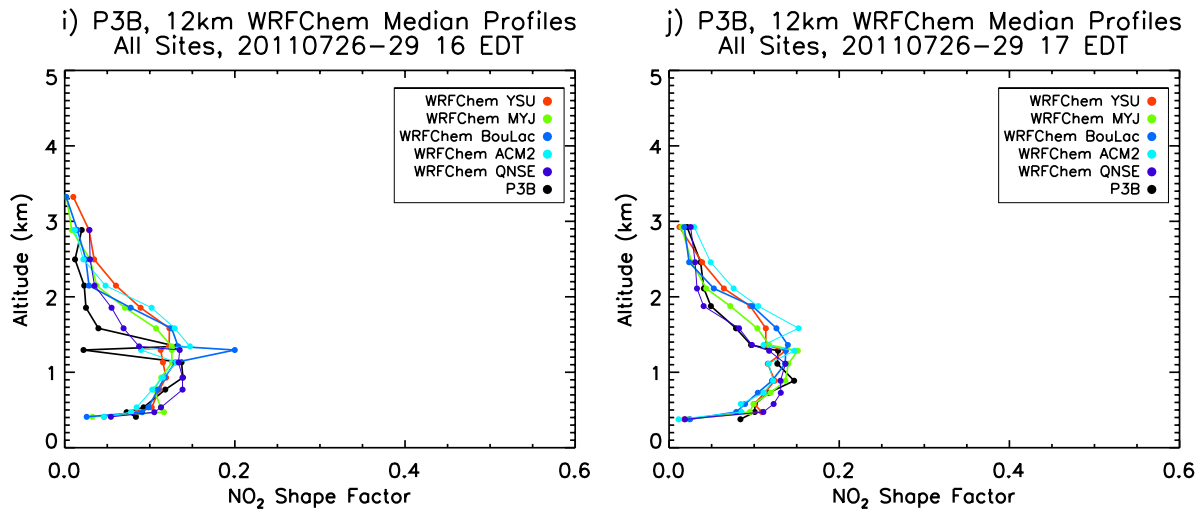
Fig. 5.18 i)-j): Hourly-median plots of NO$_2$ shape factors for the hours between 8 AM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 4 PM and 5 PM. All available profiles included.
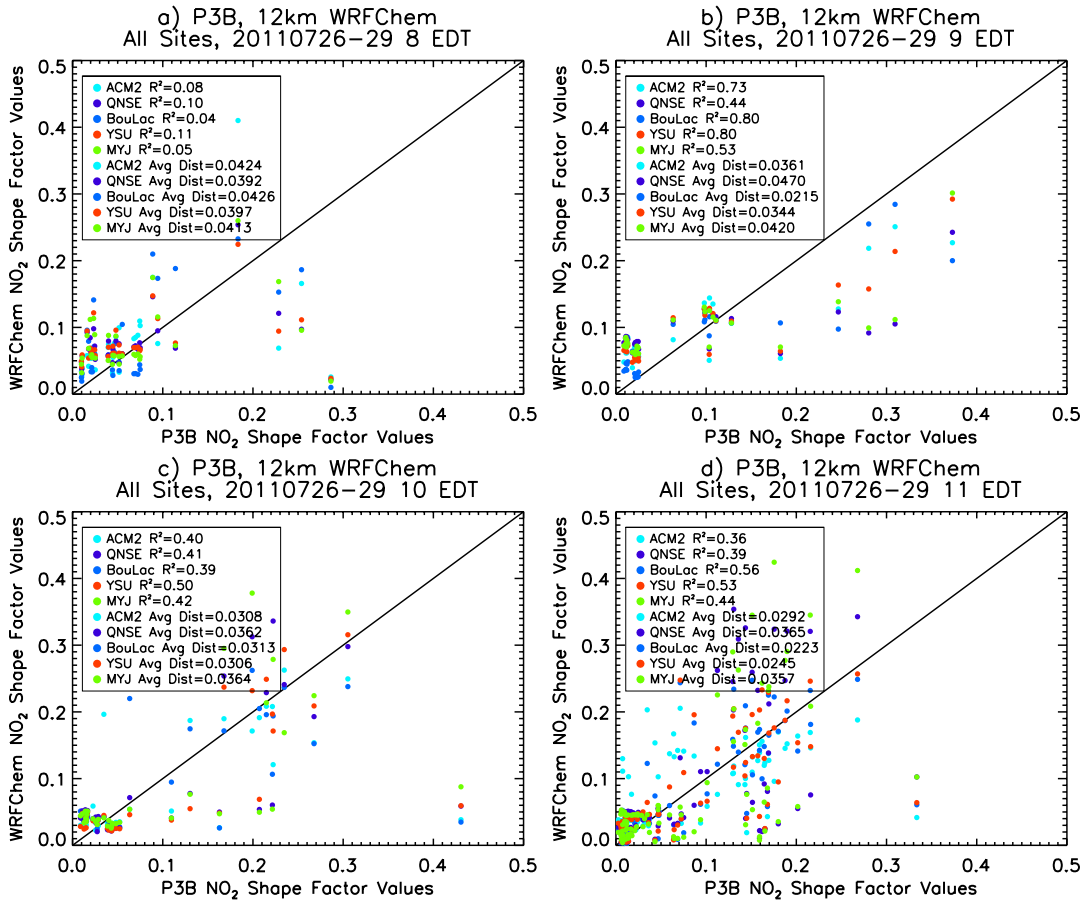
Fig. 5.19: a)-d) Scatter plots of NO₂ shape factor magnitudes for the hours between 8 AM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 8 AM and 11 AM. All available profiles included. R^2 for the correlation between simulated and observed shape factor values and average orthogonal distance from the 1:1 line displayed in bottom right.
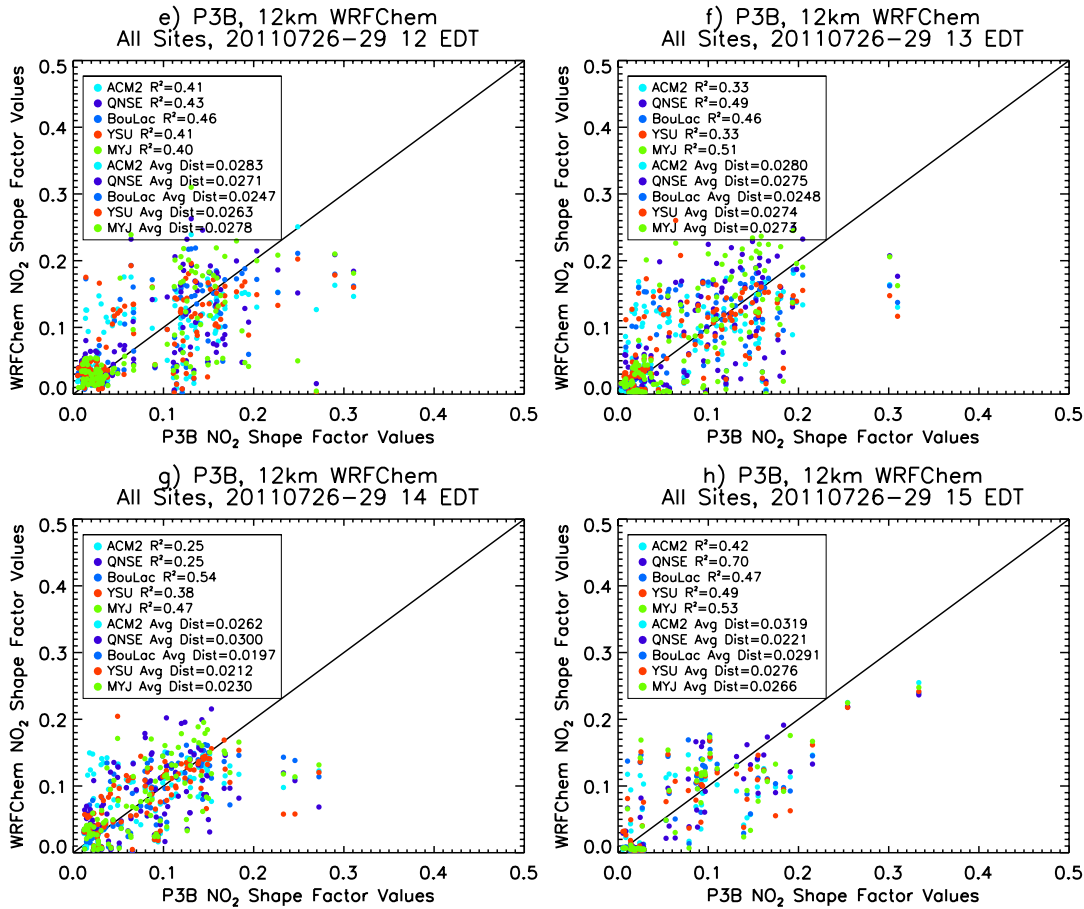
Fig. 5.19: e)-h) Scatter plots of NO₂ shape factor magnitudes for the hours between 8 AM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 12 PM and 3 PM. All available profiles included. R^2 for the correlation between simulated and observed shape factor values and average orthogonal distance from the 1:1 line displayed in bottom right.
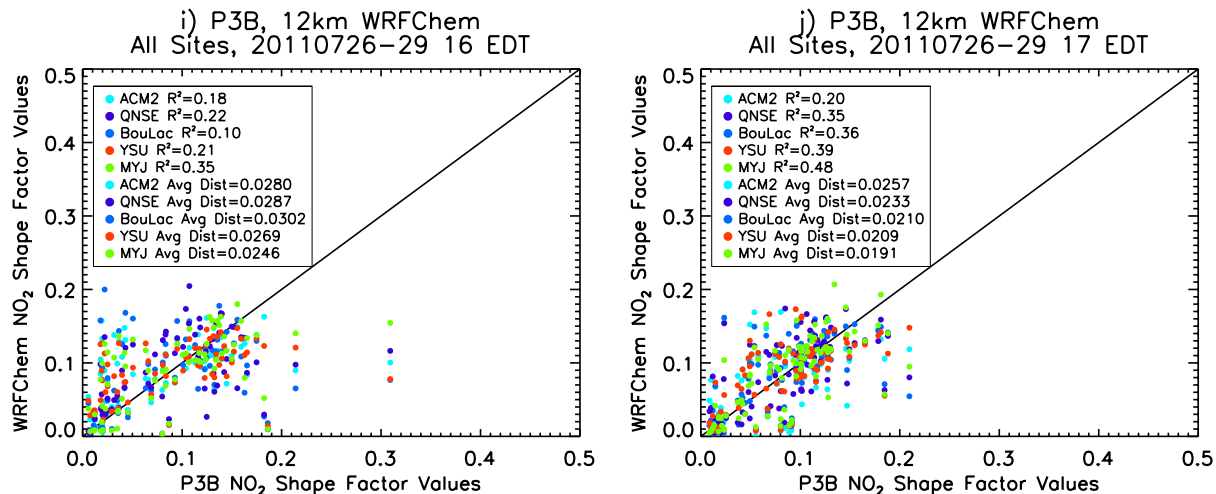
Fig. 5.19: i)-j) Scatter plots of NO₂ shape factor magnitudes for the hours between 8 AM and 5 PM EDT from the P-3B observations and WRF-Chem simulation output for each of the five PBL schemes. Hours between 4 PM and 5 PM. All available profiles included. R² for the correlation between simulated and observed shape factor values and average orthogonal distance from the 1:1 line displayed in bottom right.
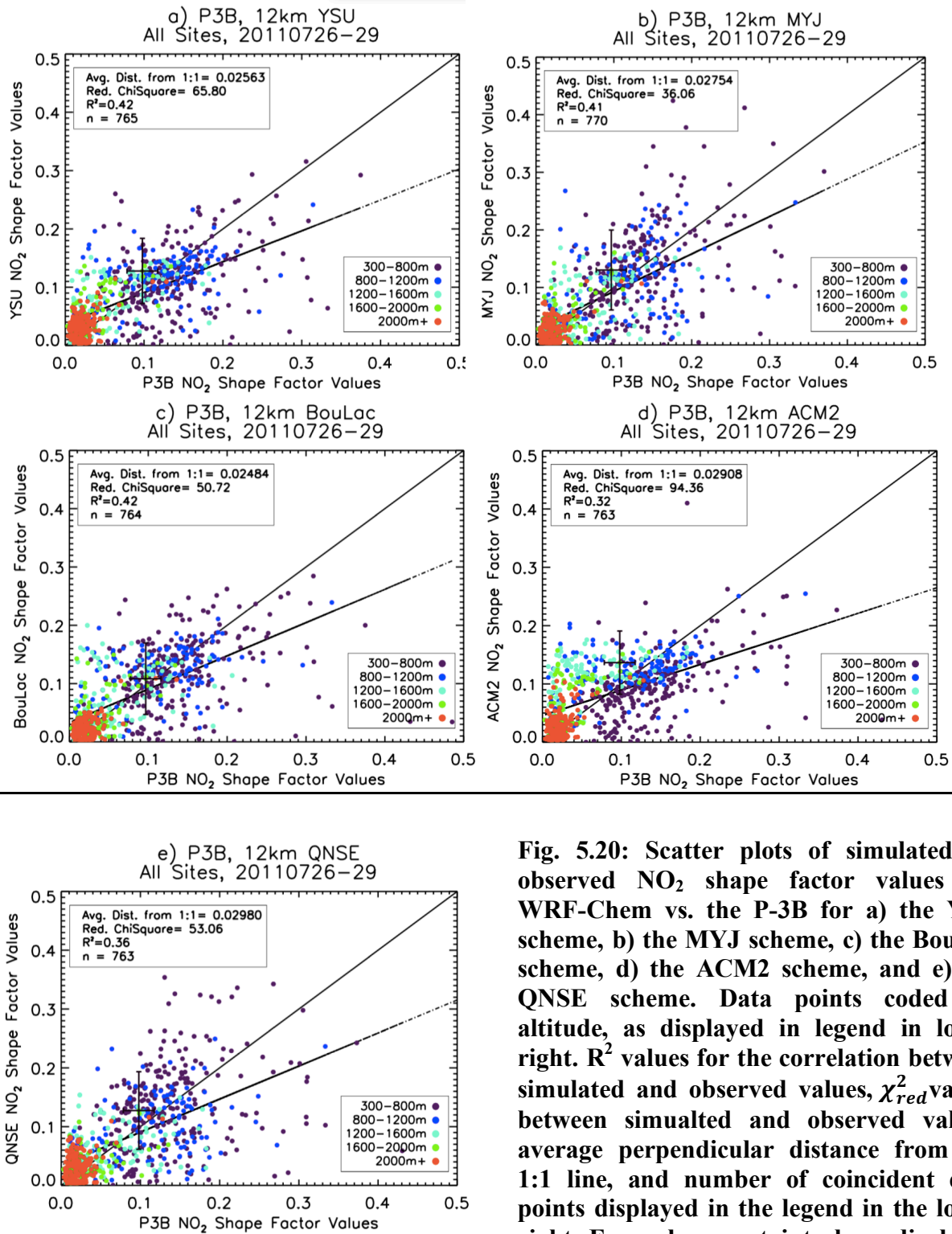
Fig. 5.20: Scatter plots of simulated vs. observed NO$_2$ shape factor values for WRF-Chem vs. the P-3B for a) the YSU scheme, b) the MYJ scheme, c) the BouLac scheme, d) the ACM2 scheme, and e) the QNSE scheme. Data points coded by altitude, as displayed in legend in lower right. R^2 values for the correlation between simulated and observed values, χ^2_{red} values between simualted and observed values, average perpendicular distance from the 1:1 line, and number of coincident data points displayed in the legend in the lower right. Example uncertainty bars displayed on median data point in red: uncertainty in measurement used for observed PBLH and standard deviation used for simulated PBLH.

5.8.5: P-3B and WRF-Chem O_3 and NO_2 Column-Surface Correlations

A simple linear regression analysis, following the procedure outlined in Chapter 2, was performed between the O_3 or NO_2 column and surface values for the P-3B observations and WRF-Chem output for each PBL scheme, to investigate the degree of correlation between column and surface data during the July 26-29 period. Correlations are the focus here, to investigate if column and surface data were related during this time period, and the relative strengths of this relationship among the observations and PBL schemes, a preliminary step before a true regression model between column and surface data can be built. Values of R^2 for these correlation analyses are provided in Tables 5.7-5.12. Results are presented only for the Edgewood, Essex, Fair Hill, and Padonia spiral sites, as photolytic or corrected molybdenum converter NO_2 *in situ* surface observations were available for only these four sites. Representative scatter plots of the observed O_3 and NO_2 correlations are presented in Fig. 5.21 for Edgewood. Consistent with the correlation results presented in Chapter 2, all O_3 correlations were statistically significant and large for P-3B column_air and column_ground, with R^2 values between 0.84 and 0.93 (Table 5.7). The correlations for O_3 column_air and column_ground were also not statistically different from each other, indicative that O_3 is well mixed horizontally and vertically, as expected. No NO_2 column_air correlations were statistically significant for this period, indicating no connection between column and surface quantities (Table 5.7); column_ground is not presented as the comparison between the P-3B and Cessna NO_2 column and profile data presented in Chapter 2 demonstrated that column_air better represented the true lower tropospheric NO_2 column. These results are also consistent with those of Chapter 2, which demonstrated that NO_2 exhibited at best

a low correlation between P-3B column amounts and surface mixing ratios when all available data were used in the correlation analysis.

Each of the five PBL schemes produced a high degree of correlation between the simulated O_3 column and surface amounts at most sites, with most R^2 values between 0.70 and 0.96 (Tables 5.8-5.12). Representative scatter plots for the simulated O_3 and NO_2 simulated correlations are presented for the ACM2 at Edgewood in Fig. 5.22, and correlations were > 0.80 for all schemes at the site. However, only a moderate degree of correlation (R^2 = 0.47) was obtained between O_3 column and surface amounts at Essex for the MYJ scheme, and Fair Hill presented no significant correlations for any PBL scheme except BouLac. Additionally, the WRF-Chem O_3 correlations were not statistically significantly different from those for the P-3B column_air and column_ground at any site except Fair Hill; the P-3B correlations were significantly larger than those for all five PBL schemes at this site. These comparisons suggest that, while all five schemes performed well relative to the observed correlations at most sites during this pollution event, the BouLac scheme most accurately replicated the observed O_3 column-surface relationship because it produced a significant correlation at all sites including Fair Hill. The good comparison between the simulated O_3 column-surface connection and the observed column-surface connection is consistent with the good comparison between simulated and observed O_3 shape factors.

The PBL scheme correlations for the simulated NO_2 column and surface amounts are much more varied than for O_3 (Tables 5.7-5.12). Each PBL scheme produced a significant correlation between the simulated NO_2 columns and surface mixing ratios for at least one site, and, when significant, the R^2 values for these correlations were at least 0.38. Interestingly, significant correlations were obtained at all four sites with the BouLac scheme, with R^2 values

between 0.48 and 0.71. The ACM2 and QNSE schemes each produced only one significant correlation, at the Essex (R^2 = 0.38) and Fair Hill (R^2 = 0.42) sites, respectively, while the MYJ and YSU schemes both produced significant correlations at Edgewood and Essex (R^2 values between 0.49 and 0.57). Furthermore, the BouLac correlations at each site were statistically significantly larger than those for the P-3B, supporting that this scheme greatly overestimated the NO_2 column-surface relationship. The YSU Essex and QNSE Fair Hill correlations were also significantly larger than those for the P-3B at these sites, indicating that, as they each produce a significant correlation, these schemes also overestimated the column-surface relationship. However, the ACM2 Essex correlation was not significantly different from that for the P-3B, meaning that the simulated correlation is not, in fact, statistically significant and that this scheme produced no significant correlations. The ACM2 scheme was thus the PBL scheme that most accurately simulated the NO_2 column-surface relationship during this period, followed closely by the YSU and QNSE schemes. As with O_3, the correlation results are consistent with the shape factor comparisons for NO_2. The poor observed correlation also reflects the degree of O_3 production during this pollution event: high temperatures and abundant sunlight quickly converted NO_2 to O_3, depleting the NO_2 column and surface and further interfering with the column-surface connection. Further, these results highlight the relative impact of overestimated vertical mixing in the model on the simulated O_3 and NO_2 column-surface correlations. Overly vigorous model boundary layer mixing did not impair the O_3 correlations relative to the observations for any PBL scheme as much as it did for NO_2, given the longer lifetime of O_3 within the PBL; all schemes except the ACM2 scheme mixed NO_2 too well relative to the observations to produce greater correlations than observed.

July 26-29 Period Correlations – P3-B			
Site	**Column_Air O_3 R^2**	**Column_Ground O_3 R^2**	**Column_Air NO_2 R^2**
Edgewood	0.86	0.88	N.S.
Essex	0.84	0.84	N.S.
Fair Hill	0.93	0.93	N.S.
Padonia	0.84	0.88	N.S.

Table 5.7: Values of R^2 for the correlations between P-3B column_air and column_ground O_3 and column_air NO_2 and surface mixing ratios at the four spiral sites. N.S. denotes a correlation that was not statistically significant at a confidence level of 95%.

July 26-29 Period Correlations – ACM2		
Site	**O_3 R^2**	**NO_2 R^2**
Edgewood	0.80	N.S.
Essex	0.73	0.38 (N.S.)
Fair Hill	N.S.	N.S.
Padonia	0.87	N.S.

Table 5.8: Values of R^2 for the correlations between simulated column O_3 and column NO_2 and surface mixing ratios at the four spiral sites for the ACM2 scheme. N.S. denotes a correlation that was not statistically significant at a confidence level of 95%.

July 26-29 Period Correlations – BouLac		
Site	$O_3 R^2$	$NO_2 R^2$
Edgewood	0.87	0.71
Essex	0.75	0.51
Fair Hill	0.41	0.69
Padonia	0.95	0.48

Table 5.9: Values of R^2 for the correlations between simulated column O_3 and column NO_2 and surface mixing ratios at the four spiral sites for the BouLac scheme. N.S. denotes a correlation that was not statistically significant at a confidence level of 95%.

July 26-29 Period Correlations – MYJ		
Site	$O_3 R^2$	$NO_2 R^2$
Edgewood	0.81	0.50
Essex	0.47	0.49
Fair Hill	N.S.	N.S.
Padonia	0.86	N.S.

Table 5.10: Values of R^2 for the correlations between simulated column O_3 and column NO_2 and surface mixing ratios at the four spiral sites for the MYJ scheme. N.S. denotes a correlation that was not statistically significant at a confidence level of 95%.

July 26-29 Period Correlations – QNSE		
Site	O_3 R^2	NO_2 R^2
Edgewood	0.82	N.S.
Essex	0.73	N.S.
Fair Hill	N.S.	0.42
Padonia	0.88	N.S.

Table 5.11: Values of R^2 for the correlations between simulated column O_3 and column NO_2 and surface mixing ratios at the four spiral sites for the QNSE scheme. N.S. denotes a correlation that was not statistically significant at a confidence level of 95%.

July 26-29 Period Correlations – YSU		
Site	O_3 R^2	NO_2 R^2
Edgewood	0.82	0.57
Essex	0.72	0.56
Fair Hill	N.S.	N.S.
Padonia	0.96	N.S.

Table 5.12: Values of R^2 for the correlations between simulated column O_3 and column NO_2 and surface mixing ratios at the four spiral sites for the YSU scheme. N.S. denotes a correlation that was not statistically significant at a confidence level of 95%.
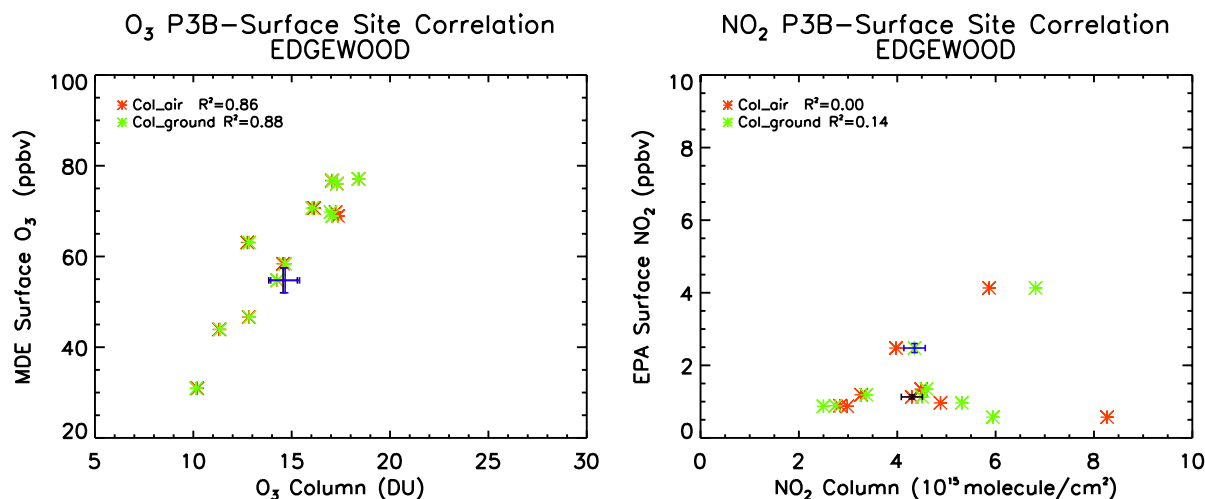
Fig. 5.21: Representative scatter plots of column vs. surface O₃ (left) and NO₂ (right) correlations for the P-3B spirals at Edgewood over July 26-29, 2011. R^2 values for column_air and column_ground displayed in upper left.
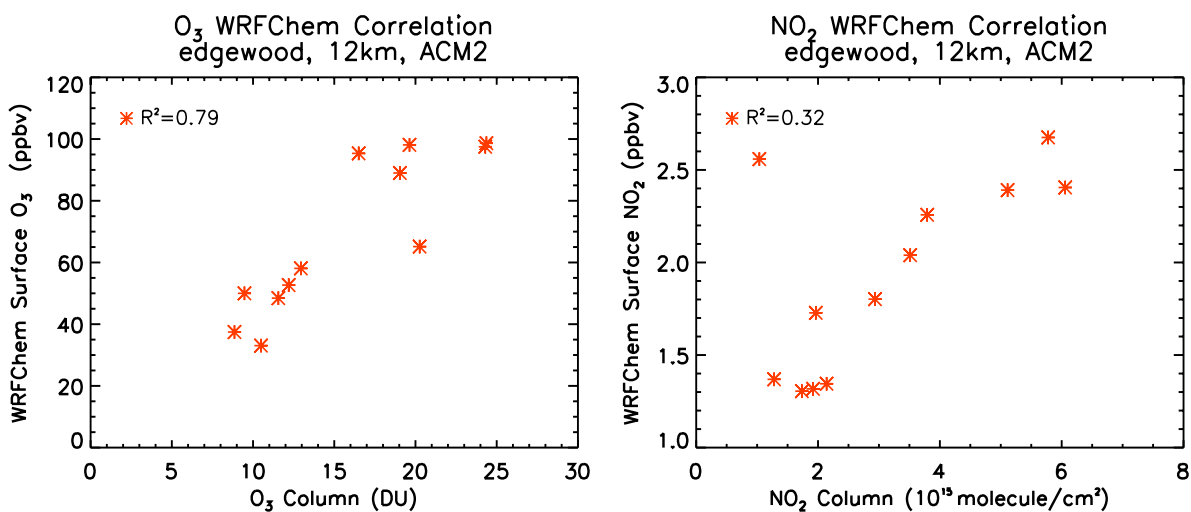
Fig. 5.22: Representative scatter plots of column vs. surface O₃ (left) and NO₂ (right) correlations for the times of P-3B spirals at Edgewoond from WRF-Chem simulations with the ACM2 scheme. R^2 values for column_air and column_ground displayed in upper left.

5.9: Conclusions

Five PBL schemes were tested in the WRF-Chem v3.7.1 model, including the ACM2, BouLac, MYJ, QNSE, and YSU schemes. The QNSE scheme consistently overpredicted the PBL

height relative to all three observational datasets (the P-3B, MPL, and HSRL PBLH estimates). The ACM2 scheme compared best to the P-3B observational estimates, while the ACM2 and BouLac schemes compared well to the MPL and HSRL estimates, in terms of replicating the observed PBL height. However, though ACM2 and BouLac compared best to the observations relative to the other schemes, the ACM2 scheme also produced a consistent high bias in PBLH, indicating an overestimation of vertical mixing. With the exception of PBL growth that was too rapid in the early morning and PBL collapse that was too fast during late afternoon, each of the five schemes replicated the shapes of the average diurnal cycles of the P-3B, MPL, and HSRL datasets. The YSU and MYJ schemes behaved similarly to each other in terms of the average PBLH diurnal variation and simulated PBLH values. These results suggest that the ACM2 scheme best captures PBLH of the five schemes tested, though overestimated vertical mixing remains an issue.

All schemes also better captured the NO_2 hourly median *in situ* mixing ratio profiles than the *in situ* potential temperature, CO, and O_3 profiles. The BouLac scheme best captured the NO_2 profiles, followed by the YSU and MYJ schemes, at most hours of the day. The ACM2 scheme tended to mix the profiles for potential temperature, CO, O_3, and NO_2 too deeply; however, it was often able to capture the shapes of the potential temperature profiles. Though no scheme emerged that produced a consistently good comparison to the CO and O_3 *in situ* mixing ratio profiles, the BouLac scheme most often compared best to these median profiles. These results, coupled with the overprediction of PBLH by ACM2 and QSNE, suggest that a scheme with less intense vertical mixing (such as the BouLac scheme, or ACM2 modified to provide less intense vertical mixing) is needed to capture *in situ* PBLH observations and profile observations for both short-lived and well mixed species.

Each PBL scheme well captured the shapes of the P-3B O_3 hourly median shape factors profiles, and presented high correlation between simulated and observed shape factor values. Additionally, WRF-Chem was able to replicate most of observed local minima and maxima in the O_3 shape factors at the correct altitudes, presenting a distinct advantage of a regional online meteorology-chemistry model over offline air quality models such as CMAQ. However, WRF-Chem demonstrated a consistent low bias in shape factor values within the lowermost 800 m of the shape factor profiles, with compensating overpredictions at higher levels, indicating that overestimated vertical mixing caused the model to place too large a proportion of the O_3 column burden too high in the vertical relative to the observations. This further indicates the need for improvements in vertical mixing within nonlocal PBL schemes. Each scheme struggled to reproduce the P-3B NO_2 shape factors, though the ACM2 BouLac, and YSU most well compared to the observations during early afternoon, despite the ACM2 and YSU schemes mixing their respective shape factors too deeply. Each scheme tended to overpredict the shape factor values above 2000 m, though different schemes over- and underpredicted at different altitudes below 2000 m, with mainly random errors within the lowermost shape factor profiles. The BouLac scheme most accurately simulated the observed O_3 column-surface correlations, while the ACM2 scheme most accurately simulated the NO_2 column-surface correlations.

Overall, no one PBL scheme was able to accurately simulate all observed quantities or relationships. However, as the BouLac and ACM2 schemes frequently produced good comparisons to the P-3B for different types of analyses, this suggests that a nonlocal scheme with vertical mixing adjusted to be somewhat more local, may be able to more accurately simulate profiles and column quantities, leading to an improved column-surface connection within the model relative to the observed relationship. Finally, despite deficiencies that would lead to

retrieval errors, it bears repeating that these results are encouraging that a regional, coupled meteorology-chemistry model may reasonably specify *a priori* O_3 profile shapes for remote sensing retrievals of O_3 columns, particularly as PBL schemes continue to be improved.

Chapter 6: Conclusions and Future Research Directions

Satellite column observations of trace gases have great potential for diagnosis of surface or near-surface air quality conditions, particularly over regions lacking a sufficiently dense surface air quality monitoring network. However, biases and uncertainties within satellite instrumentation and retrieval algorithms currently limit our ability to relate column abundances to surface mixing ratios. As the launch dates for planned geostationary air quality satellites such as TEMPO draw nearer, this need to understand the linkage between column abundance and surface mixing ratio becomes more urgent. The NASA DISCOVER-AQ mission was designed to provide sufficient observations of key meteorological and chemical species over four different regions within the U.S. exactly for this purpose: to provide information relevant to improving our ability to relate column and surface observations for key trace gases and aerosols.

The work presented in this dissertation has been conducted in support of the goals of DISCOVER-AQ: to better understand how well column quantities represent surface air quality for the EPA criteria pollutants O_3 and NO_2 during the four DISCOVER-AQ deployments. Three peer-reviewed journal articles have also been produced from this dissertation, including Flynn et al. (2014) published in *Atmospheric Environment*, Flynn et al. (2016) submitted to *Atmospheric Environment*, and Flynn et al. (in prep.) to be submitted to *Atmospheric Chemistry and Physics*. This work is focused on characterizing the degree of correlation between column and surface abundances for O_3 and NO_2, understanding the variability of *in situ* profiles and column quantities and how that variability relates to surface quantities, and investigation of the role of boundary layer mixing in controlling the column-surface relationship through use of the WRF-Chem model:

Initial Investigation of Column-Surface Relationship during the Maryland Campaign

In the first part, we analyzed the degree of correlation between O_3 and NO_2 observational column and surface data sets from the Maryland campaign of the DISCOVER-AQ project, as well as investigated the impact of boundary layer mixing and compared the observed column-surface correlations to those simulated by the CMAQ regional air quality model. This work was the initial investigation into the column-surface relationship, begun immediately after the Maryland deployment, and served as the springboard for all subsequent analysis of this and the other three campaigns. A simple linear regression analysis was applied to the P-3B column_air and column_ground lower tropospheric column amounts and coincident surface mixing ratio data, to the Pandora full tropospheric column amounts and coincident surface data, and to the simulated lower tropospheric column amounts and surface mixing ratios within CMAQ for the six spiral sites of this campaign. P-3B column_ground and column_air O_3 demonstrated the greatest correlation between column and surface quantities of all datasets, and exhibited correlations that were not statistically different from each other, indicating that O_3 is well mixed horizontally and vertically; NO_2 typically exhibited poor correlation for most datasets. The simple linear regression analyses were repeated for each additional correlation analysis, first by adding inverse PBLH as a second predictor variable and second by normalizing the O_3 or NO_2 column amount by PBLH before performing the regressions, to investigate the impact of boundary layer mixing on the strength of the correlations. These results indicate that PBL height added meaningful information to the column-surface relationship for NO_2, because it is a short-lived species that is not as well mixed as O_3. The CMAQ model well replicated the observed P-3B O_3 correlations during the Maryland deployment, but overestimated the NO_2 correlations, as the model likely underestimated the NO_2 mixing ratio vertical gradient below the lowest P-3B measurement altitude. The CMAQ correlations were statistically significantly larger than those

of the observations at most spiral sites during the afternoon, when the CBL was maturely developed, while PBLH did not add useful information to the CMAQ correlations for either trace gas, indicating that these species are too well mixed within CMAQ. Lastly, The DISCOVER-AQ measurements suggest that O_3 observations from future satellite instruments can be meaningful for surface air quality analysis if they have sufficient sensitivity to the lowest 2-3 km of the troposphere.

<u>Investigation of the Variability of *In Situ* Profiles and Column Abundances for All Campaigns</u>

Column abundances depend on the *in situ* mixing ratio profiles for trace gases, as the mixing ratio profile partially determines where the greatest burden of a pollutant is located in the vertical. The location(s) in the vertical of the greatest O_3 or NO_2 burden will also determine how well that column amount relates to the surface, through boundary layer mixing processes. To further understanding of the relationship between column and surface quantities, an agglomerative hierarchical clustering algorithm was applied to the P-3B *in situ* O_3 or NO_2 profiles during each of the four campaigns, to determine if typical profile shapes emerged, factors that influenced these profile clusters, and how well the column and surface data associated with each cluster correlated. All *in situ* O_3 profile clusters produced by the agglomerative hierarchical cluster analysis were significant for the California, Texas, and Colorado deployments, with Texas producing the greatest number of distinct clusters (five). In contrast, the Maryland deployment produced only one cluster significantly different from the remaining five clusters, indicating that O_3 displayed the least profile variability during this deployment. Not surprisingly, atmospheric stability, as indicated by the lapse rate and potential temperature profiles associated with each cluster, emerged as important influences on the shapes of the profile clusters during each campaign. The column-surface correlations associated with no

Texas or Colorado O_3 profile cluster were significant, while most cluster correlations for the Maryland campaign were significant, suggesting that O_3 column observations may be most representative of surface concentrations under the conditions of deep, convective boundary layers, less intense wind shear, and few geography-meteorology interactions (with the exception of the bay breeze) associated with the Maryland deployment relative to the other campaigns. Consistent with the findings of the simple linear regression analyses performed for the Maryland campaign spiral sites discussed previously, the profile cluster results emphasize the important role of vertical mixing in the O_3 column-surface relationship. Median shape factors were also computed for each profile cluster for each campaign for both the P-3B and the regional CMAQ and global GMI model output. Both models moderately well captured the P-3B lower tropospheric shape factors for the Maryland and California campaigns, with better performance relative to the observations for the Maryland deployment.

Unlike O_3, NO_2 displayed relatively uniform profile behavior for all four campaigns, as denoted by the lack of many significantly distinct profiles clusters, while no important influences beyond NO_2 photochemical loss during daytime (with sufficient sunlight and warm temperatures) emerged that regulated the NO_2 profile variability. Almost no correlations associated with the NO_2 profile clusters were significant, indicating that accurate representation of the lower tropospheric NO_2 profile from a model in a satellite retrieval does not guarantee the ability to connect column and surface. CMAQ produced more realistic NO_2 shape factors than did GMI, though both models often struggled to capture the observed shape factors. The CMAQ NO_2 shape factors were typically overpredicted in the upper PBL/lower free troposphere, an altitude region to which OMI-like instruments are more sensitive than to the near-surface. However, both models performed best for the Maryland campaign; the Texas and Colorado

campaigns also experienced deep boundary layers, but also many more cloudy days than did the Maryland campaign. These results demonstrate that models may be best able to capture O_3 and NO_2 profiles under the conditions of convective boundary layers and O_3 photochemical formation associated with the Maryland campaign. It must be stressed that both the hierarchical profile cluster results and the linear regression analyses presented thus far emphasize that satellites observations may best be able to capture surface conditions for O_3 and NO_2 for the conditions associated with the Maryland campaign, and that a regional air quality model may adequately prescribe the O_3 and NO_2 profiles for use in satellite retrievals. However, the CMAQ shape factor errors for NO_2 could lead to retrieval errors.

However, these results leave some important questions unanswered. First, satellite retrievals often employ monthly-mean simulated *a priori* profiles in the shape factor computation that goes into the air mass factor. Would use of these typical profile shapes that emerged from the clustering analysis within satellite retrievals over each region of the U.S. improve retrievals of vertical column abundances over those regions? How much variability is contained within each profile cluster (i.e.; how greatly does each individual profile within a cluster differ from the cluster median profile)? What could such a variability analysis tell us about the representativeness of these profile clusters of the true *in situ* lower tropospheric profile for the conditions associated with each cluster, and would additional influences on the profile shapes and thus column-surface relationship emerge? This work is left for future endeavors.

As geostationary air quality satellites will observe column abundances during all daylight hours, it is important to understand how the O_3 and NO_2 column amounts vary throughout the day, and if that diurnal variability relates to the diurnal cycle of surface mixing ratio. This may reveal indications of when during the day satellite observations may be most representative of

surface concentrations. The campaign-average diurnal timeseries was computed for O_3 and NO_2 for each spiral site of each campaign. The O_3 full troposphere and lower tropospheric column observational datasets demonstrated little diurnal variation, in contrast to the surface O_3 diurnal cycle. This is indicative that boundary layer mixing plays a crucial role in controlling the O_3 column diurnal cycle: surface concentrations are mixed up into the PBL as the PBL grows, while O_3 remaining within the residual layer is mixed down into the growing PBL, dampening the variability of the column amount relative to the surface. The NO_2 full and lower tropospheric column datasets, however, displayed a consistent diurnal variability at most spiral sites during all campaigns, though the column diurnal variability was smaller in amplitude and offset in time relative to the surface diurnal variation. Additionally, NO_2 photochemical loss (as indicated by the diurnal cycle of the photolysis frequency, $j(NO_2)$) emerged as an important influence on the NO_2 column variability. Neither set of results for O_3 and NO_2 suggest a time of day when satellite column observations may be most representative of surface concentrations. Both the CMAQ and GMI models replicated the shapes of the O_3 and NO_2 column diurnal cycles. Vertical mixing yet again emerged as an important theme in the column-surface connection, as the model results further highlight the importance of vertical mixing. However, questions also again remain unanswered and left for future endeavors. How do the diurnal timeseries on individual days at these spiral sites compare to the campaign average, and what could that reveal about the variability of the column diurnal variation? As a sanity check, how do the column diurnal timeseries compare to that of PBLH, and does the change in column with time for each hour (d(column)/dt) exhibit a similar diurnal cycle as that for the surface mixing ratios (d(surface)/dt) for NO_2? What is it about the differences in the remote sensing retrievals that lead to differences

in the column diurnal variations? Lastly, would the diurnal cycles of the O_3 and NO_2 columns normalized by PBLH better resemble the surface diurnal cycles?

Investigation of the Boundary Layer Mixing in the WRF-Chem Model

The last section of this dissertation focused on evaluation of the ability of the regional, coupled chemistry-meteorology WRF-Chem v3.7.1 model to effectively simulate the interplay between boundary layer mixing and O_3 and NO_2 vertical profiles, and the associated impacts on the column-surface correlations for these trace gases, for the July 26-29, 2011, O_3 pollution episode. An online meteorology-chemistry model was chosen to avoid the middle-man time averaging of the WRF output required to run CMAQ offline, while the Maryland deployment was chosen because it presented the greatest connection between O_3 column and surface amounts. Five PBL schemes were evaluated, including two nonlocal schemes (ACM2 and YSU) and three local schemes (MYJ, BouLac, and QNSE). The QNSE consistently greatly overpredicted PBLH relative to the P-3B, MPL, and HSRL observational PBLH datasets. The ACM2 scheme also produced a high bias in PBLH relative to each observational dataset, though not as greatly as with QNSE, indicating an overestimation of vertical mixing. Additionally, these schemes presented mixed results in the comparisons to the observed *in situ* hourly median potential temperature, CO, O_3, and NO_2 mixing ratio profiles: no one scheme emerged that compared well to the P-3B for these four species, and typically no one scheme emerged that captured the diurnal variation of even one of these species well throughout the day between 8 AM and 5 PM EDT. These results suggest that a scheme with less intense vertical mixing (such as the BouLac scheme) is needed to capture *in situ* PBLH observations and profile observations for both short-lived and well mixed species. These results emphasize the need for improvements

to vertical mixing within PBL schemes to more accurately capture the observed mixing and interplay between mixing and pollutant behavior.

However, each PBL scheme well captured the shapes of the P-3B O_3 hourly median shape factor profiles, and presented high correlation between simulated and observed shape factor values, while each scheme struggled to capture the P-3B NO_2 hourly median shape factor profiles. The ACM2 scheme also tended to overestimate the shape factor in the upper PBL/lower free troposphere, altitude levels where satellite instruments are more sensitive than near the surface; hence, satellite retrieval error would be much greater when there is error in the assumed shape factor at these levels than for error at the surface. However, the ACM2 scheme most accurately simulated the NO_2 column-surface correlations, exhibiting no significant correlation at each spiral site just as the P-3B did. Despite deficiencies in the PBL schemes, it should be emphasized that each of the five schemes was most able to replicate most local maxima and minima in the observed O_3 median shape factor profiles of the five schemes, bolstering the conclusions of the comparison of CMAQ to the P-3B in the profile clustering results that a regional air quality model may adequately prescribe *a priori* profiles for use in satellite retrievals. Additionally, because WRF-Chem replicated much of the observed local shape factor maxima and minima, these results indicate an advantage to use of a coupled model rather than offline meteorology/chemistry model system.

However, one important question remains unanswered in this section: how can the ACM2 scheme be improved to bring vertical mixing, and associated impacts on the column-surface connection, more in line with the observations? The ACM2 scheme is of particular importance as it is the boundary layer mixing scheme used in CMAQ, the EPA regulatory model.

Experiments are currently underway to elucidate which parameters within the scheme, if any, can be modified to improve the representation of vertical mixing within this scheme.

Appendix A: Scatter Plots and Extended Analysis for Column vs. Surface for Full Data Set Correlations at Each Site for the Maryland Campaign Analysis from Chapter 2

Scatter plots of the correlation between surface and column data are presented in this appendix. These plots are presented for each data set at each site for O_3 (Figures A1-A3) and NO_2 (Figures A4-A6).

The sites of maximum and minimum correlation between column and surface in each data set are summarized in Table A1. Fair Hill presented the largest correlation for the P-3B NO_2 data likely because it is the site farthest from large NO_2 sources; NO_2 is most vertically and horizontally well mixed at Fair Hill. It is interesting to note that Beltsville presented the largest correlation for the P-3B O_3 data and Essex presented the lowest correlation for the P-3B and Pandora data, as Beltsville is one of the sites least impacted by the bay breeze while Essex is often impacted. The bay breeze most often impacted the Edgewood site, causing the O_3 column-surface relationship to be complex here (Loughner et al., 2013; Stauffer et al., 2012). The P-3B and Pandora results highlight the considerable variability in the column-surface relationship that exists among the surface sites for both trace gases. Because no OMI correlations were significant for either trace gas, no sites of maximum or minimum correlation can be identified

	Max. NO_2 Correlation	Min. NO_2 Correlation	Max. O_3 Correlation	Min. O_3 Correlation
P-3B Col_air	Fair Hill	Beltsville	Beltsville	Essex
P-3B Col_ground	Fair Hill	Beltsville	Beltsville	Essex
Pandora	Essex	Aldino	Fair Hill	Essex
CMAQ (Loughner et	Fair Hill	Beltsville	Padonia	Edgewood
CMAQ (NOAA)	Fair Hill	Edgewood	Beltsville	Essex

Table A1: Summary of the sites with the maximum and minimum degree of correlation between column amounts and surface mixing ratios for each trace gas and analysis.
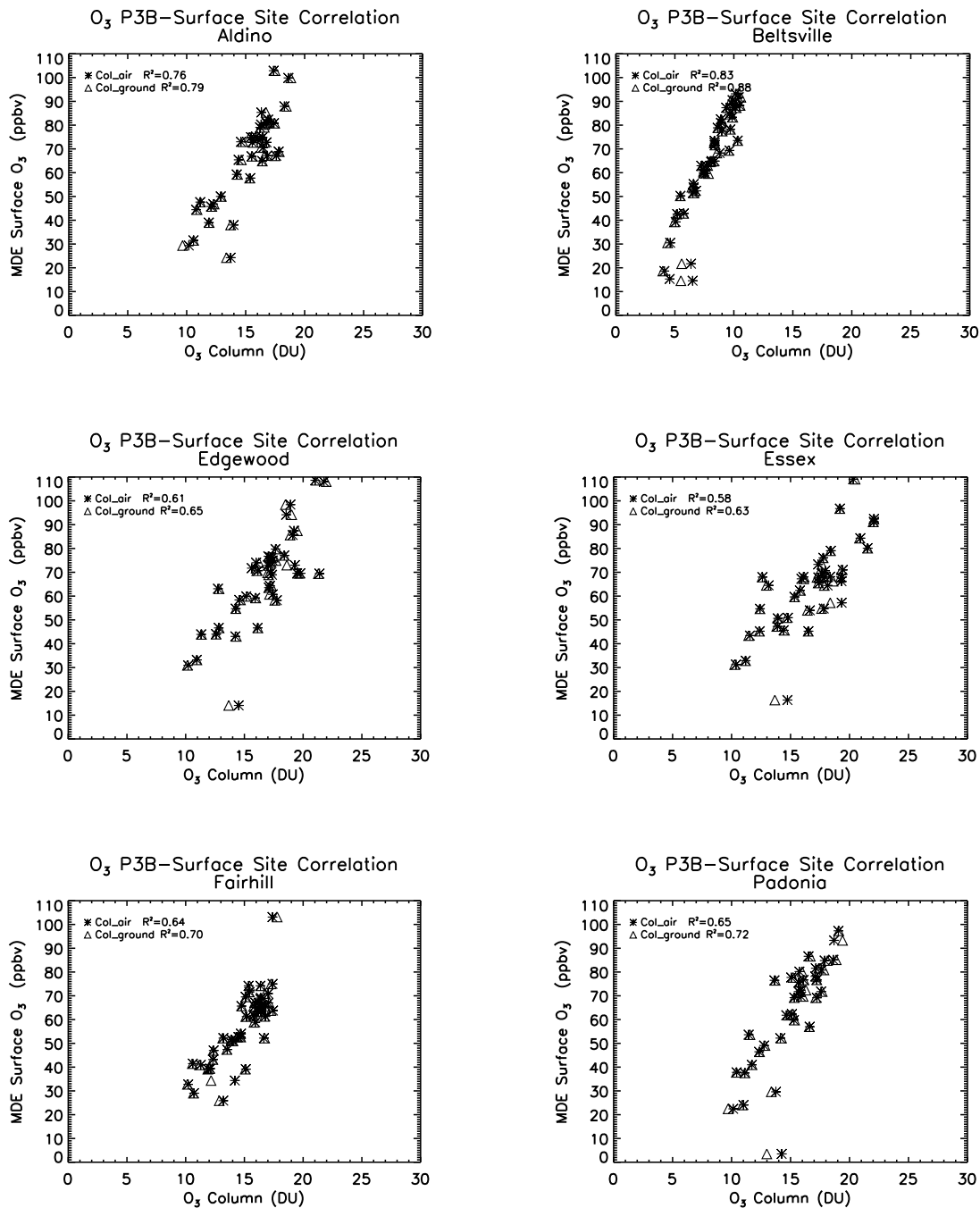
Figure A1: Scatter plots of O3 column vs. surface O3 mixing ratio at each of the 6 surface sites for P-3B column_air and column_ground. Full data set correlation. R^2 values are displayed in the upper left of each plot.
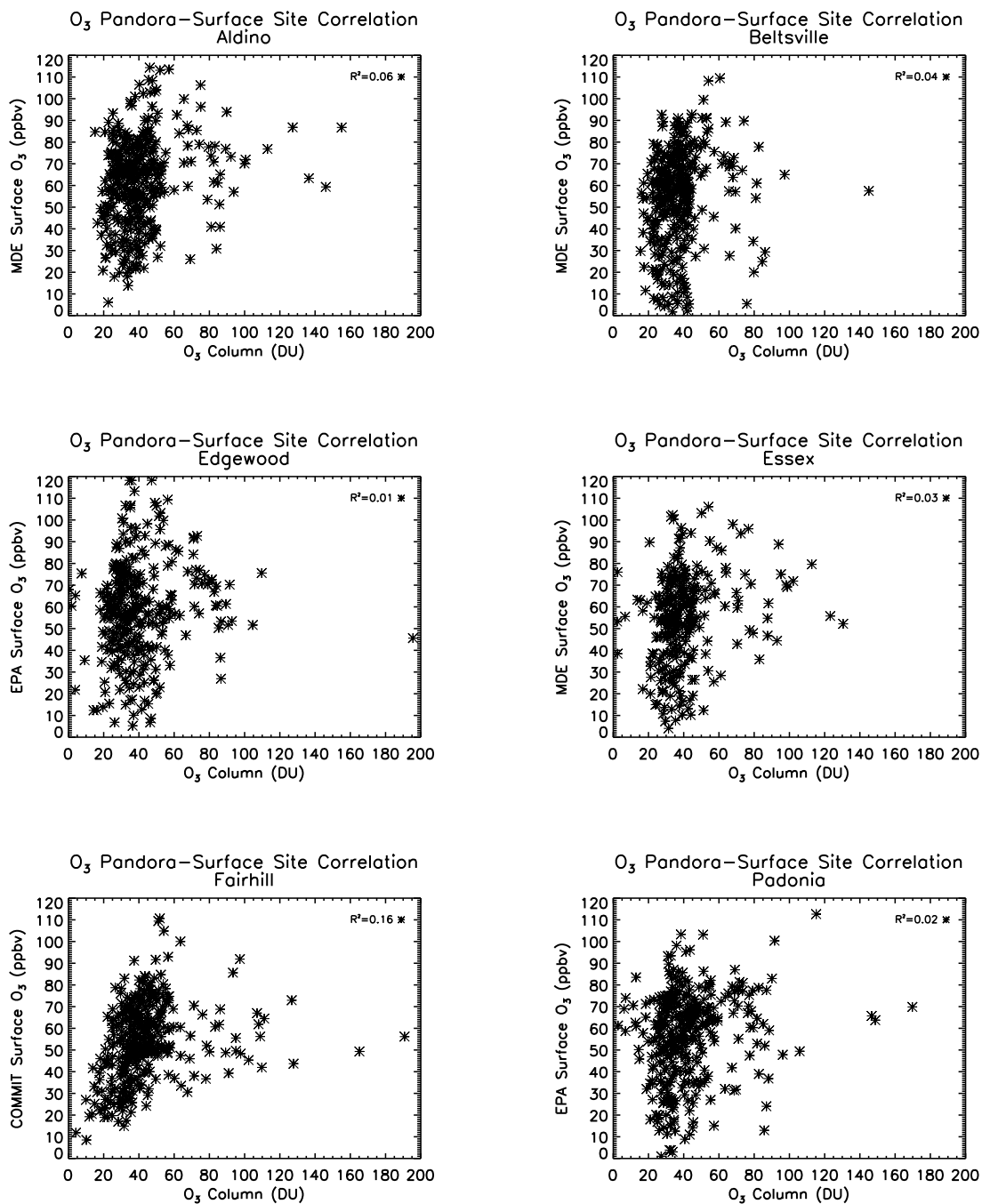
Figure A2: Scatter plots of O3 column vs. surface O3 mixing ratio at each of the 6 surface sites for Pandora. Full data set correlation. R^2 values are displayed in the upper left of each plot.
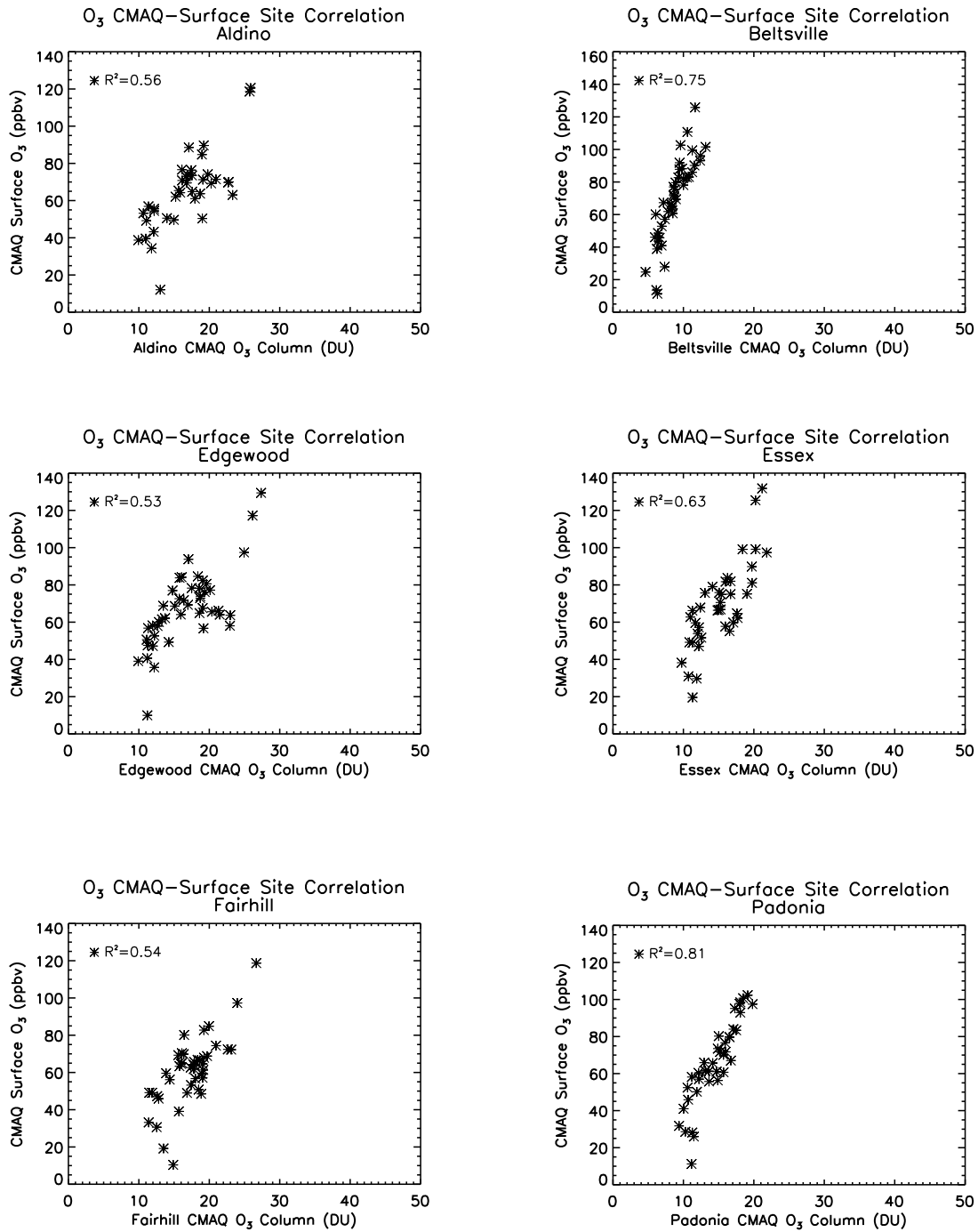
Figure A3: Scatter plots of O3 column vs. surface O3 mixing ratio at each of the 6 surface sites for CMAQ (Lougher et al., 2013). Full data set correlation. R^2 values are displayed in the upper left of each plot.
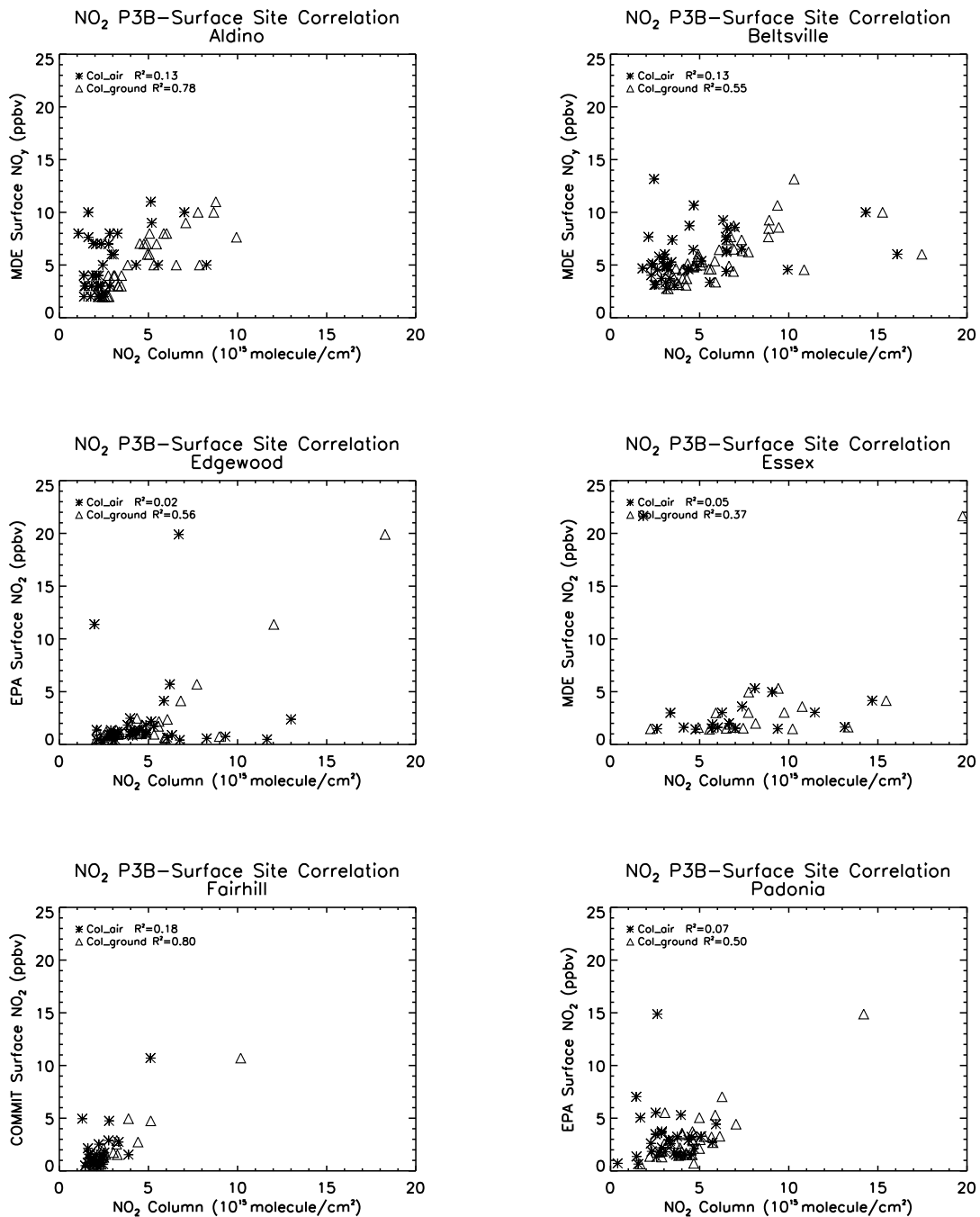
Figure A4: Scatter plots of NO2 column vs. surface NO2 mixing ratio at each of the 6 surface sites for P-3B column_air and column_ground. Full data set correlation. R^2 values are displayed in the upper left of each plot. Surface observations for Aldino and Beltsville are NO$_y$.
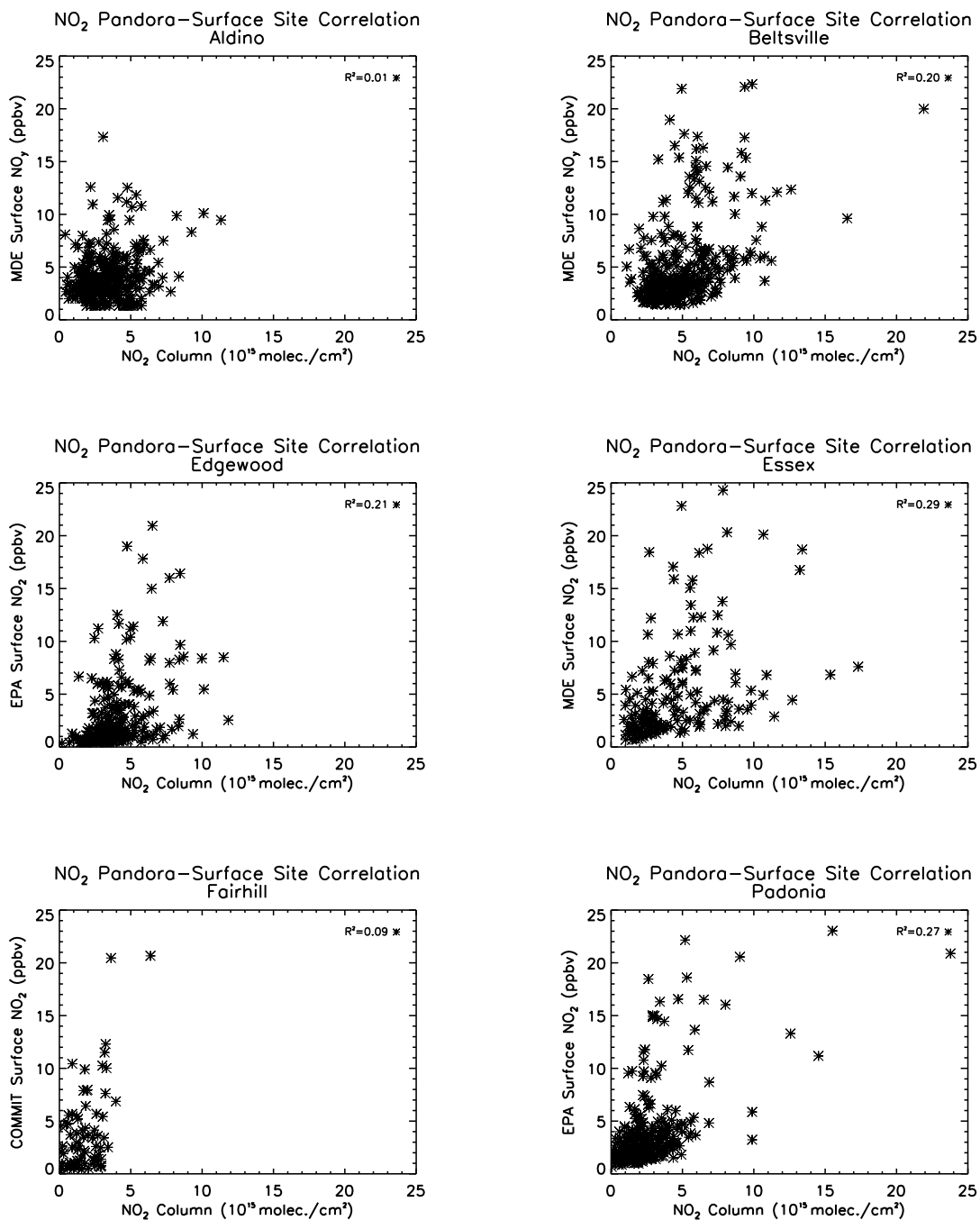
Figure A5: Scatter plots of NO2 column vs. surface NO2 mixing ratio at each of the 6 surface sites for Pandora. Full data set correlation. R^2 values are displayed in the upper left of each plot.
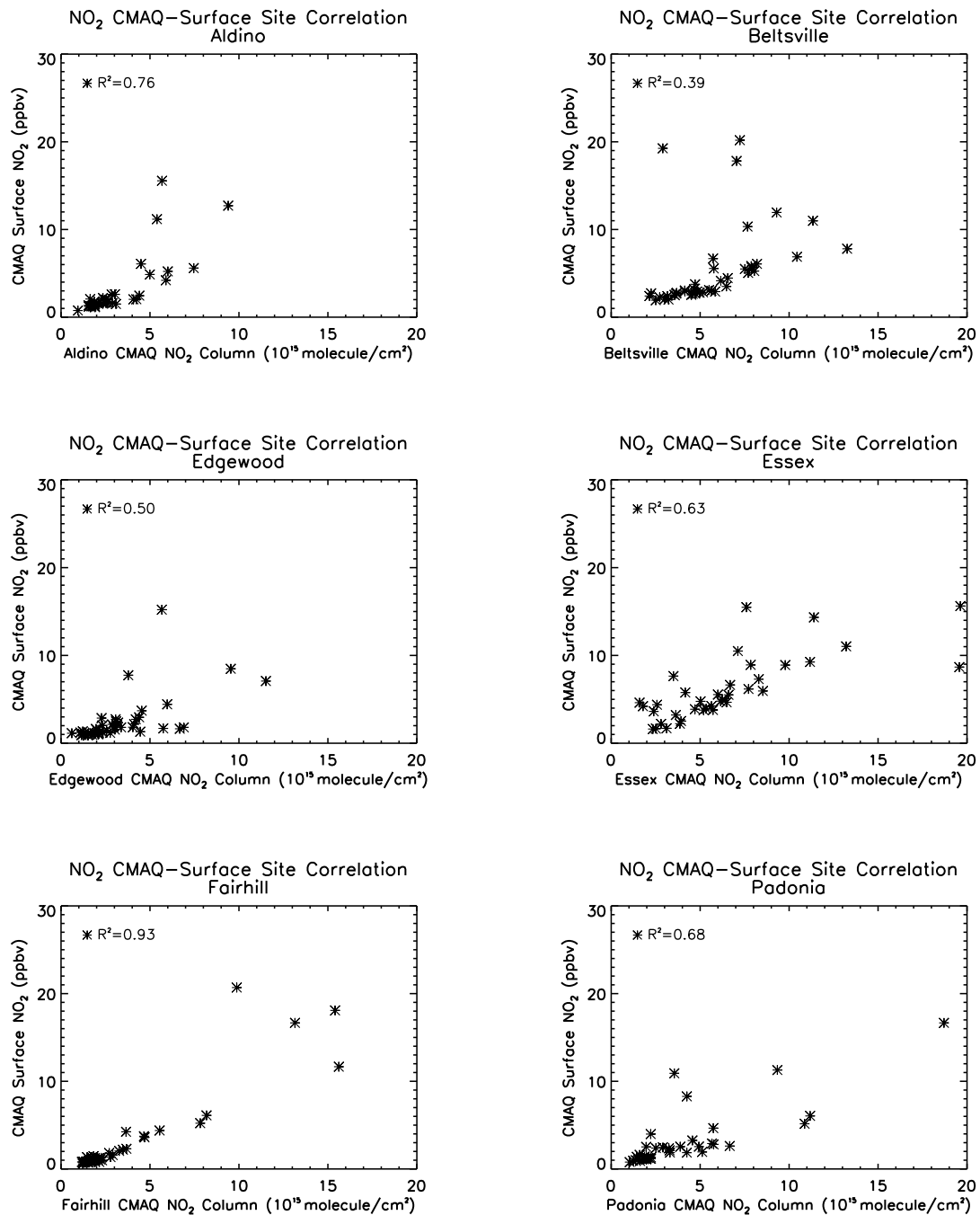
Figure A6: Scatter plots of NO2 column vs. surface NO2 mixing ratio at each of the 6 surface sites for CMAQ (Loughner et al., 2013). Full data set correlation. R^2 values are displayed in the upper left of each plot.

Appendix B: Values of R^2 for the Correlations Separated into Data Groups for the Maryland Campaign Analysis from Chapter 2

Table B1: Values of R^2 for P-3B columns vs. surface for Time of Day Analysis

	Correlation Coefficient (Column_ground) Morning	Correlation Coefficient (Column_ground) Afternoon	Correlation Coefficient (Column_air) Morning	Correlation Coefficient (Column_air) Afternoon
Aldino NO$_2$	0.86	0.84	NS	0.36
Beltsville NO$_2$	0.29	0.77	NS	0.33
Edgewood NO$_2$	0.86	0.54	NS	NS
Essex NO$_2$	0.92	NS	NS	NS
Fair Hill NO$_2$	0.94	0.43	0.92	NS
Padonia NO$_2$	0.56	0.43	NS	0.41
Aldino O$_3$	0.70	0.72	0.64	0.69
Beltsville O$_3$	0.81	0.84	0.70	0.82
Edgewood O$_3$	0.57	0.53	0.47	0.48
Essex O$_3$	0.69	0.55	0.58	0.51
Fair Hill O$_3$	0.62	0.58	0.51	0.53
Padonia O$_3$	0.55	0.81	0.43	0.81

Table B2: Values of R^2 for Pandora columns vs. surface for Time of Day Analysis

	Correlation Coefficient (O_3) Morning	Correlation Coefficient (O_3) Afternoon	Correlation Coefficient (NO_2) Morning	Correlation Coefficient (NO_2) Afternoon
Aldino	NS	0.02	0.06	NS
Beltsville	NS	0.03	0.23	0.33
Edgewood	NS	NS	0.34	0.32
Essex	NS	NS	0.52	0.34
Fair Hill	0.08	NS	0.23	0.10
Padonia	NS	0.03	0.30	0.23

Table B3: Values of R^2 for P-3B columns vs. surface for PBL Separation Analysis

	Correlation Coefficient (Column_ground) Low PBL Group	Correlation Coefficient (Column_ground) High PBL Group	Correlation Coefficient (Column_air) Low PBL Group	Correlation Coefficient (Column_air) High PBL Group
Aldino NO_2	0.81	0.85	NS	NS
Beltsville NO_2	NS	0.70	NS	0.30
Edgewood NO_2	0.71	NS	NS	NS
Essex NO_2	NS	0.3	NS	NS
Fair Hill NO_2	0.78	NS	0.26	NS
Padonia NO_2	0.67	0.23	NS	0.28
Aldino O_3	0.73	0.88	0.70	0.86
Beltsville O_3	0.83	0.89	0.75	0.89
Edgewood O_3	0.68	0.55	0.63	0.52
Essex O_3	0.52	0.69	0.47	0.69
Fair Hill O_3	0.68	0.83	0.633	0.73
Padonia O_3	0.65	0.82	0.54	0. 81

Table B4: Values of R^2 for Pandora columns vs. surface for PBL Separation Analysis

	Correlation Coefficient (O$_3$) Low PBL	Correlation Coefficient (O$_3$) High PBL	Correlation Coefficient (NO$_2$) Low PBL	Correlation Coefficient (NO$_2$) High PBL
Aldino	0.03	NS	NS	NS
Beltsville	NS	NS	0.30	0.29
Edgewood	NS	NS	0.21	0.38
Essex	0.03	NS	0.25	0.23
Fair Hill	0.14	NS	0.26	NS
Padonia	NS	NS	0.27	0.31

Table B5: Values of R^2 for CMAQ columns vs. surface for Time of Day Analysis (Loughner et al. Runs)

	Correlation Coefficient (O$_3$) Morning	Correlation Coefficient (O$_3$) Afternoon	Correlation Coefficient (NO$_2$) Morning	Correlation Coefficient (NO$_2$) Afternoon
Aldino	NS	0.52	0.78	0.72
Beltsville	0.59	0.65	NS	0.61
Edgewood	NS	0.44	NS	0.57
Essex	NS	0.55	NS	0.59
Fair Hill	0.37	0.55	0.92	0.89
Padonia	0.63	0.83	0.71	0.66

Table B6: Values of R^2 for CMAQ columns vs. surface for PBL Separation Analysis (Loughner et al. Runs)

	Correlation Coefficient (O$_3$) Low PBL	Correlation Coefficient (O$_3$) High PBL	Correlation Coefficient (NO$_2$) Low PBL	Correlation Coefficient (NO$_2$) High PBL
Aldino	0.60	0.80	0.70	0.78
Beltsville	0.61	0.73	NS	0.52
Edgewood	0.54	0.83	0.44	0.52
Essex	0.70	0.80	0.68	0.85
Fair Hill	0.66	0.65	0.93	0.72
Padonia	0.74	0.84	0.71	0.83

Table B7: Values of R^2 for CMAQ columns vs. surface for Time of Day Analysis (NOAA Runs)

	Correlation Coefficient (O_3) Morning	Correlation Coefficient (O_3) Afternoon	Correlation Coefficient (NO_2) Morning	Correlation Coefficient (NO_2) Afternoon
Aldino	0.60	0.80	0.70	0.78
Beltsville	0.61	0.73	NS	0.52
Edgewood	0.54	0.83	0.44	0.52
Essex	0.70	0.80	0.68	0.85
Fair Hill	0.66	0.65	0.93	0.72
Padonia	0.74	0.84	0.71	0.83

Table B8: Values of R^2 for CMAQ columns vs. surface for PBL Separation Analysis (NOAA Runs)

	Correlation Coefficient (O_3) Low PBL	Correlation Coefficient (O_3) High PBL	Correlation Coefficient (NO_2) Low PBL	Correlation Coefficient (NO_2) High PBL
Aldino	0.83	0.90	0.34	0.79
Beltsville	0.87	0.85	0.94	0.81
Edgewood	0.79	0.87	0.68	0.66
Essex	0.64	0.80	0.95	0.88
Fair Hill	0.75	0.93	0.61	0.97
Padonia	0.69	0.82	0.25	0.91

Appendix C: Quadratic Least Squares Regression Analysis for the P-3B and Pandora for the Maryland Campaign Analysis from Chapter 2

A quadratic least squares regression was performed for P-3B and Pandora NO_2. This yielded an equation of the form $y = \beta_1*x_1 + \beta_2*x_1^2 +$ intercept, where y is the surface mixing ratio, x_1 is the column abundance, β_1 is the first regression coefficient associated with the column, and β_2 is the second regression coefficient associated with the square of the column. A polynomial fit was applied only to the NO_2 data because scatter plots of NO_2 surface data vs. columns often exhibited non-linear behavior.

A non-linear regression model was applied to only the NO_2 data because the scatter plots of NO_2 surface data versus columns often exhibited some non-linear behavior. Due to the shallowness of the layer containing the large NO_2 values in the lower portion of the boundary layer, the contribution to column content may not be that great, leading to a non-linear relationship between column and surface values. The values of R^2 for column_ground were larger than for the simple linear regression at most sites; many column_air regressions were not significant at a confidence level of 95% (Table C1). However, the average percentage errors did not demonstrate a consistent improvement over those for the simple linear regressions; four out of six sites demonstrated an increase in average error. The percentage of estimations within a ±10% or ±50% error did not demonstrate consistent improvement (Table C2). The quadratic fit regressions for the Pandora NO_2 data also displayed slightly improved R^2 values at half of the sites, though the average percentage errors also did not (Tables C1, C2). However, the standard deviations associated with the average percentage errors decreased relative to those for the simple linear regressions at some sites for the P-3B and Pandora NO_2 analyses, indicating some reduction in the variability of the NO_2 data. Histograms of the regression residuals for P-3B column_air revealed that residuals were more normally distributed for this analysis than for the

259

simple linear regression analysis at most sites, but plots of the residuals against the second predictor (log-transformed square of the column) displayed some linear structure for column_ground at some sites. Autocorrelation of the residuals remained apparent in the lag-1 residual plots for Pandora at each site, but the histograms of the Pandora residuals also displayed a more normal distribution compared to the simple linear regressions (Figure C1). More investigation is needed to determine the correct model for column and surface NO_2, but a quadratic fit improved the appropriateness of a least squares regression to fit the P-3B and Pandora NO_2 data.

	P-3B NO$_2$ Col_Air R^2	P-3B NO$_2$ Col_Ground R^2	P-3B NO$_2$ Col_Air F-ratio	P-3B NO$_2$ Col_Ground F-ratio	Pandora NO$_2$ R^2	Pandora NO$_2$ F-ratio
Aldino	0.15	0.81	2.93 (0.067)	70.2 (<0.001)	0.03	5.03 (0.007)
Beltsville	0.13	0.63	3.02 (0.059)	33.80 (<0.001)	0.24	53.25 (<0.001)
Edgewood	0.02	0.60	0.38 (0.688)	24.35 (<0.001)	0.22	32.00 (<0.001)
Essex	0.38	0.50	4.57 (0.027)	7.59 (0.014)	0.29	34.30 (<0.001)
Fair Hill	0.40	0.80	6.88 (0.005)	43.10 (<0.001)	0.14	5.98 (0.004)
Padonia	0.13	0.51	22.6 (0.305)	25.51 (<0.001)	0.31	58.7 (<0.001)
Aggregate	0.01	0.503	1.19 (0.276)	72.30 (<0.001)	0.16	136.36 (<0.001)

Table C1: Summary of the R^2 statistic and F-ratio (p-value) for the P-3B and Pandora NO$_2$ quadratic fit regressions.

	P-3B NO	% of Cases w/in	% of Cases w/in	Pandora NO Mean Error	% of Cases w/in	% of Cases w/in
Aldino	12. (± 58.2) %	18.3%	73.2%	27.1 (± 88.1) %	22.6%	80.4%
Beltsville	2.9 (± 18.2) %	51.2%	98.8%	19.3 (± 56.2) %	24.2%	76.1%
Edgewood	- (± 512.1 %)	2.6%	22.4%	- (± 2304.1) %	5.2%	28.6%
Essex	30. (± 79.6%)	25.0%	66.7%	57.7 (± 370.7) %	10.6%	62.4%
Fair Hill	- (± 172.1) %	10.4%	37.5%	- (± 2334.6) %	2.7%	41.9%
Padonia	6.5 (± 93.1) %	21.2%	66.7%	107. (± 839.7) %	16.5%	63.8%
Aggregate	-3.3 (± 1185.9) %	7.3%	55.2%	-12.8 (± 1772.9) %	15.5%	64.8%

Table C2: Summary of percentage errors of P-3B and Pandora quadratic fit regression for all sites and the aggregate relative to observed surface values. P-3B Col_air and col_ground are analyzed together for each site and the aggregate.
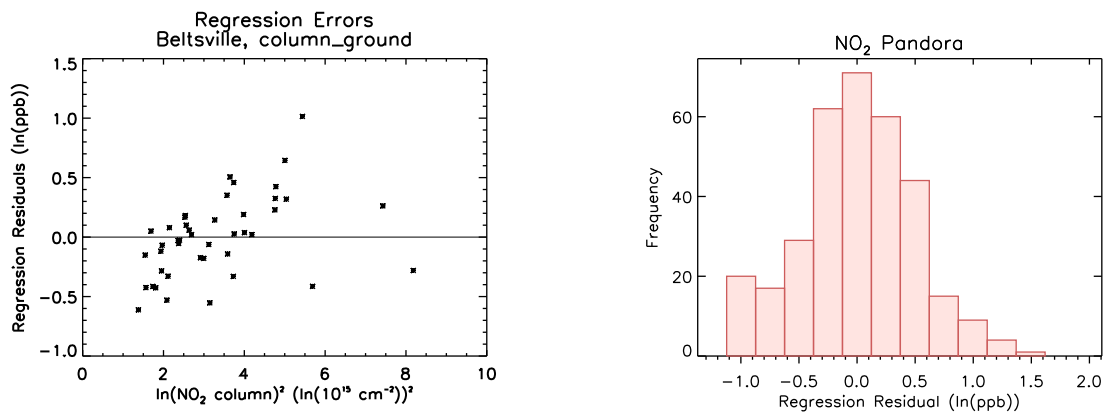
Figure C1: Example scatter plots for NO$_2$ quadratic fit residuals. Plot of residuals vs. square of the NO2 column for the Beltsville column_ground analysis and histogram of residuals for the Pandora Aldino NO2 analysis for the quadratic fit regression analysis.

References

Anderson, D. C., et al. (2014), Measured and modeled CO and NOy in DISCOVER-AQ: An evaluation of emissions and chemistry over the eastern US, *Atmos. Environ.,* 96,78–87.

Allen, D., Pickering, K., Duncan, B., Damon, M. (2010). Impact of lightning NO emissions on North American photochemistry as determined using the Global Modeling Initiative (GMI) model. *Journal of Geophysical Research Atmospheres*, *115*(D22), 10.1029/2010JD014062.

Barnard, J. C., Fast, J. D., Paredes-Miranda, G., Arnott, W. P., & Laskin, A. (2010). Technical note: evaluation of the WRF-chem" aerosol chemical to aerosol optical properties" module using data from the milagro campaign. *Atmospheric Chemistry and Physics*, *10*(15), 7325-7340.

Beirle, S., Platt, U., Wenig, M., Wagner, T. (2003). Weekly cycle of NO2 by GOME measurements : a signature of anthropogenic sources. *Atmospheric Chemistry and Physics*, *3*(2), 2225–2232.

Bhartia, P. K. (2002). OMI Algorithm Theoretical Basis Document, Volume II, OMI Ozone Products. NASA-OMI, Washington, DC, ATBD-OMI-02, Version 2.0, II(August).

Bloomer, B. J., J.W. Stehr, C. A. Piety, R. J. Salawitch, and R. R. Dickerson (2009), Observed relationships of ozone air pollution with temperature and emissions, Geophys. Res. Lett., 36, L09803, doi:10.1029/ 2009GL037308.

Boersma, K. F., Jacob, D. J., Eskes, H. J., Pinder, R. W., Wang, J., & van der A, R. J. (2008). Intercomparison of SCIAMACHY and OMI tropospheric NO2 columns: Observing the diurnal evolution of chemistry and emissions from space. Journal of Geophysical Research: Atmospheres, 113(16). http://doi.org/10.1029/2007JD008816

Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., & Eskes, H. J. (2009). Validation of urban NO2 concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities. Atmospheric Chemistry and Physics, 9(12), 3867–3879. http://doi.org/10.5194/acp-9-3867-2009

Bougeault, P., & Lacarrere, P. (1989). Parameterization of Orography-Induced Turbulence in a Mesobeta--Scale Model. Monthly Weather Review. http://doi.org/10.1175/1520-0493(1989)117<1872:POOITI>2.0.CO;2

Brooks, I. M. (2003). Finding boundary layer top: Application of a wavelet covariance transform to lidar backscatter profiles. Journal of Atmospheric and Oceanic Technology, 20(8), 1092–1105. http://doi.org/10.1175/1520-0426(2003)020<1092:FBLTAO>2.0.CO;2

Bucsela, E. J., Krotkov, N. A., Celarier, E. A., Lamsal, L. N., Swartz, W. H., Bhartia, P. K., Boersma, K.F., Veefkind, J.P., Gleason, J. F., & Pickering, K. E. (2013). A new stratospheric and tropospheric NO2 retrieval algorithm for nadir-viewing satellite instruments: Applications to OMI. Atmospheric Measurement Techniques, 6(10), 2607–2626. http://doi.org/10.5194/amt-6-2607-2013.

Canty, T. P., L. Hembeck, T. P. Vinciguerra, D. C. Anderson, D. L. Goldberg, S. F. Carpenter, D. J. Allen, C. P. Loughner, R. J. Salawitch, and R. R. Dickerson (2015), Ozone andNOx chemistry in the eastern US: Evaluation of CMAQ/CB05 with satellite (OMI) data, Atmos. Chem. Phys., 15, 10,965–10,982.

Castellanos, P., Marufu, L. T., Doddridgem, B. G., Taubman, B. F., Schwab, J. J., Hains, J. C., Ehrman, S. H., and Dickerson, R. R.: Ozone, oxides of nitrogen, and carbon monoxide dur-30 ing pollution events over the eastern United States: an evaluation of emissions and vertical mixing, J. Geophys. Res., 116, D16307, doi:10.1029/2010JD014540, 2011.

Chance, K. V. (2002). OMI Algorithm Theoretical Basis Document Volume IV, OMI Trace Gas Algorithms. *NASA-OMI, Washington, DC, ATBD-OMI-04, Version 2.0*, *IV*(August), 1–50. Retrieved from http://ozoneaq.gsfc.nasa.gov/media/docs/ATBD-OMI-04.pdf

Chance, K., Liu, X., Suleiman, R. M., Flittner, D. E., Al-Saadi, J., & Janz, S. J. (2013, September). Tropospheric emissions: monitoring of pollution (TEMPO). In *SPIE Optical Engineering+ Applications* (pp. 88660D-88660D). International Society for Optics and Photonics.

Chatfield, R. B., & Esswein, R. F. (2012). Estimation of surface O3 from lower-troposphere partial-column information: Vertical correlations and covariances in ozonesonde profiles. *Atmospheric Environment*, *61*, 103–113. http://doi.org/10.1016/j.atmosenv.2012.06.033

Diab, R. D., Raghunandan, A., Thompson, A. M., & Thouret, V. (2003). Classification of tropospheric ozone profiles over Johannesburg based on MOZAIC aircraft data. *Atmospheric Chemistry and Physics Discussions*, *3*(1), 705-732.

Diab, R. D., Thompson, A. M., Mari, K., Ramsay, L., & Coetzee, G. J. R. (2004). Tropospheric ozone climatology over Irene, South Africa, from 1990 to 1994 and 1998 to 2002. *Journal of Geophysical Research D: Atmospheres*, *109*(20).

Demerjian, K. L. (2004). A review of national monitoring networks in North America. *Atmospheric Environment*, *34*. 1861-1884.

Duncan, B. N., Strahan, S. E., & Yoshida, Y. (2007). Model study of the cross-tropopause transport of biomass burning pollution. *Atmospheric Chemistry and Physics Discussions*, *7*(1), 2197–2248. http://doi.org/10.5194/acpd-7-2197-2007.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J. F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., & Kloster, S. (2010). Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4). *Geoscientific Model Development*, *3*(1), 43-67.

Fast, J. D., Gustafson, W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., Grell, G. A., & Peckham, S. E. (2006). Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. *Journal of Geophysical Research: Atmospheres*, *111*(D21).

Finlayson-Pitts, B. J. & Pitts, J. N., *Chemistry of the Upper and Lower Atmosphere*, Academic Press: San Diego, 2000. Print.

Fishman, J., Bowman, K. W., Burrows, J. P., Richter, A., Chance, K. V., Edwards, D. P., Martin, R. V., Morris, G. A., Pierce, R. B., Ziemke, J. R., Schaack, T. K., & Thompson, A. M. (2008). Remote sensing of tropospheric pollution from space. *Bulletin of the American Meteorological Society*, *89*(6), 805–821. http://doi.org/10.1175/2008BAMS2526.1

Fishman, J., Iraci, L. T., Al-Saadi, J., Chance, K., Chavez, F., Chin, M., Coble, P., Davis, C., DiGiacomo, P. M., Edwards, D., Eldering, A., Goes, J., Herman, J., Hu, C., Jacob, D. J., Jordan, C., Kawa, S. R., Key, R., Liu, X., Lohrenz, S., Mannino, A., Natraj, V., Neil, D., Neu, J., Newchurch, M., Pickering, K., Salisbury, J., Sosik, H., Subramaniam, A., Tzortziou, M., Wang, J., & Wang, M. (2012). The United States' Next Generation of Atmospheric Composition and Coastal Ecosystem Measurements: NASA's Geostationary Coastal and Air Pollution Events (GEO-CAPE) Mission. *Bulletin of the American Meteorological Society*, *93*(10), 1547–1566. http://doi.org/10.1175/BAMS-D-11-00201.1

Flynn, C. M., Pickering, K. E., Crawford, J. H., Lamsal, L., Krotkov, N., Herman, J., Weinheimer, A., Chen, G., Liu, X., Szykman, J., Tsay, S., Loughner, C., Hains, J., Lee, P., Dickerson, R. R., Stehr, J. W., & Brent, L. (2014). Relationship between column-density and surface mixing ratio: Statistical analysis of O3 and NO2 data from the July 2011 Maryland DISCOVER-AQ mission. *Atmospheric Environment*, *92*, 429–441. http://doi.org/10.1016/j.atmosenv.2014.04.041

Gégo, E., Porter, P., Gilliland, A., Rao, S., 2007. Observation-based assessment of the impact of nitrogen oxides emissions reductions on ozone air quality over the eastern United States. *Journal of Applied Meteorology and Climatology* 46, 994e1008.

Goldberg, D. L., Vinciguerra, T. P., Anderson, D. C., Hembeck, L., Canty, T. P., Ehrman, S. H., Martins, D. K., et al. (2016). CAMx ozone source attribution in the eastern United States using guidance from observations during DISCOVER-AQ Maryland. *Geophysical Research Letters*, *43*(5), 2249-2258.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, *39*(37), 6957-6975.

Grosjean, D., & Harrison, J. (1985), Response of chemiluminescence NO_X analyzers and ultraviolet ozone analyzers to organic air pollutants, *Environ. Sci. Technol.*, 19, 862–865.

Guenther, A. B., X. Jiang, C. L. Heald, T. Sakulyanontvittaya, T. Duhl, L. K. Emmons, and X. Wang (2012), The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5(6), 1471–1492, doi:10.5194/gmd-5-1471-2012.

Hains, J. C., Taubman, B. F., Thompson, A. M., Stehr, J. W., Marufu, L. T., Doddridge, B. G., & Dickerson, R. R. (2008). Origins of chemical pollution derived from Mid-Atlantic aircraft profiles using a clustering technique. *Atmospheric Environment*, *42*(8), 1727–1741. http://doi.org/10.1016/j.atmosenv.2007.11.052

Hair, J. W., Hostetler, C. A., Cook, A. L., Harper, D. B., Ferrare, R. A., Mack, T. L., Welch, W., Izquierdo, L. R., & Hovis, F. E. (2008). Airborne high spectral resolution lidar for profiling aerosol optical properties. *Applied Optics*, *47*(36), 6734–6752. http://doi.org/10.1364/ao.47.006734

He, H., Loughner, C. P., Stehr, J. W., Arkinson, H. L., Brent, L. C., Follette-Cook, M. B., Tzortziou, M. A., Pickering, K. E., Thompson, A. M., Martins, D. K., Diskin, G. S., Anderson, B. E., Crawford, J. H., Weinheimer, A. J., Lee, P., Hains, J. C., & Dickerson, R. R. (2013). An elevated reservoir of air pollutants over the Mid-Atlantic States during the 2011 DISCOVER-AQ campaign: Airborne measurements and numerical simulations. *Atmospheric Environment*, *85*, 18-30.

Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M., & Abuhassan, N. (2009). NO2 column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation. *Journal of Geophysical Research: Atmospheres*, *114*(13), 1–20. http://doi.org/10.1029Z2009JD011848

Holton, J. R. (2004), *An Introduction to Dynamic Meteorology*, Academic Press, 9780123540157.

Holtslag, A. A. M., & Boville, B. A. (1993). Local versus nonlocal boundary-layer diffusion in a global climate model. *Journal of Climate*. http://doi.org/10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Monthly Weather Review*, *134*(9), 2318–2341. http://doi.org/10.1175/MWR3199.1

Hu, X. M., Nielsen-Gammon, J. W., & Zhang, F. (2010). Evaluation of three planetary boundary layer schemes in the WRF model. *Journal of Applied Meteorology and Climatology*, *49*(9), 1831–1844. http://doi.org/10.1175/2010JAMC2432.1

Jacob, Daniel, *Atmospheric Chemistry*, Princeton University Press, Princeton, 1999, Print.

Janjić, Z. I. (1990). The Step-Mountain Coordinate: Physical Package. *Monthly Weather Review*, *118*, 1429–1443.

Janjić, Z. I. (1994). The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. *Monthly Weather Review*. http://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2

Knepp, T., Pippin, M., Crawford, J., Chen, G., Szykman, J., Long, R., Cowen, L., Cede, A., Abuhassan, N., Herman, J., Delgado, R., Compton, J., Berkoff, T., Fishman, J., Martins, D., Stauffer, R., Thompson, A. M., Weinheimer, A., Knapp, D., Montzka, D., Lenschow, D., & Neil, D. (2013). Estimating surface NO2 and SO2 mixing ratios from fast-response total column observations and potential application to geostationary missions. *Journal of Atmospheric Chemistry*, (2), 1–26. http://doi.org/10.1007/s10874-013-9257-6

Kota, S. H., G. Schade, M. Estes, D. Boyer, and Q. Ying (2015), Evaluation of MEGAN

predicted biogenic isoprene emissions at urban locations in Southeast Texas, *Atmos. Environ.,* 110,54–64.

Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W. H., Pickering, K. E., Bucsela, E. J., Gleason, J. F., Martin, R. V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Szykman, J. J., & Knepp, T. N. (2014). Evaluation of OMI operational standard NO2 column retrievals using in situ and surface-based NO2 observations. *Atmospheric Chemistry and Physics Discussions*, *14*(10), 14519–14573. http://doi.org/10.5194/acpd-14-14519-2014

Lamsal, L. N., Martin, R. V., Padmanabhan, A., van Donkelaar, A., Zhang, Q., Sioris, C. E., Chance, K., Kurosu, T. P., & Newchurch, M. J. (2011). Application of satellite observations for timely updates to global anthropogenic NOx emission inventories. *Geophysical Research Letters*, *38*(5), 1–5. http://doi.org/10.1029/2010GL046476.

Lamsal, L. N., Martin, R. V., van Donkelaar, A., Steinbacher, M., Celarier, E. A., Bucsela, E., Dunlea, E. J., & Pinto, J. P. (2008), Ground-level nitrogen dioxide concentrations inferred from satellite-borne Ozone Monitoring Instrument, *J. Geophys. Res., 113*, D16308, doi: 10.1029/2007JD009235.

Lamsal, L. N., Martin, R. V., van Donkelaar, A., Celarier, E. A., Bucsela, E. J., Boersma, K. F., Dirksen, R., Luo, C., Wang, Y. (2010), Indirect validation of tropospheric nitrogen dioxide retrieved from the OMI satellite instrument: Insight into the seasonal variation of nitrogen oxides at northern midlatitudes, *J. Geophys. Res., 115*, D05302, doi:10.1029/2009JD013351.

Lamsal, L. N., Janz, S., Krotkov, N. A., Pickering, K. E., Spurr, R., Kowalewski, M., Loughner, C., Crawford, J., Swartz, W., and Herman, J. (2016), High-resolution NO_2 observations from the Airborne Compact Atmospheric Mapper: Retrieval and Validation, *J. Geophys. Res.*, submitted.

Lee, C. J., Brook, J. R., Evans, G. J., Martin, R. V., & Mihele, C. (2011). Novel application of satellite and in-situ measurements to map surface-level NO2 in the Great Lakes region. *Atmospheric Chemistry and Physics*, *11*(22), 11761–11775. http://doi.org/10.5194/acp-11-11761-2011

Lin, J. T., & McElroy, M. B. (2010). Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: Implications to satellite remote sensing. *Atmospheric Environment*, *44*(14), 1726–1739. http://doi.org/10.1016/j.atmosenv.2010.02.009

Lin, J. T., Youn, D., Liang, X. Z., & Wuebbles, D. J. (2008). Global model simulation of summertime U.S. ozone diurnal cycle and its sensitivity to PBL mixing, spatial resolution, and emissions. *Atmospheric Environment*, *42*(36), 8470–8483. http://doi.org/10.1016/j.atmosenv.2008.08.012

Liu, X., Bhartia, P. K., Chance, K., Spurr, R. J. D., & Kurosu, T. P. (2010). Ozone profile retrievals from the Ozone Monitoring Instrument. *Atmos. Chem. Phys. Atmospheric Chemistry and Physics*, *10*, 2521-2537. Retrieved from www.atmos-chem-phys.net/10/2521/2010/

Liu, C., Liu, X., Kowalewski, M. G., Janz, S. J., González Abad, G., Pickering, K. E., Chance, K., & Lamsal, L. N. (2015). Characterization and verification of ACAM slit functions for trace-gas retrievals during the 2011 DISCOVER-AQ flight campaign. *Atmospheric Measurement Techniques*, *8*(2), 751-759.

Loughner, C. P., Tzortziou, M., Follette-Cook, M., Pickering, K. E., Goldberg, D., Satam, C., Weinheimer, A., Crawford, J. H., Knapp, D. J., Montzka, D. D., Diskin, G. S., & Dickerson, R. R. (2014). Impact of bay-breeze circulations on surface air quality and boundary layer export. *Journal of Applied Meteorology and Climatology*, *53*(7), 1697–1713. http://doi.org/10.1175/JAMC-D-13-0323.1

Mao, Q., Gautney, L. L., Cook, T. M., Jacobs, M. E., Smith, S. N., & Kelsoe, J. J. (2006). Numerical experiments on MM5-CMAQ sensitivity to various PBL schemes. *Atmospheric Environment*, *40*(17), 3092–3110. http://doi.org/10.1016/j.atmosenv.2005.12.055

Martin, R. V. (2008). Satellite remote sensing of surface air quality. *Atmospheric Environment*, *42*(34), 7823–7843. http://doi.org/10.1016/j.atmosenv.2008.07.018

Martins, D. K., Stauffer, R. M., Thompson, A. M., Halliday, H. S., Kollonige, D., Joseph, E., & Weinheimer, A. J. (2015). Ozone correlations between mid-tropospheric partial columns and the near-surface at two mid-atlantic sites during the DISCOVER-AQ campaign in July 2011. *Journal of Atmospheric Chemistry*, *72*(3-4), 373-391.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, *102*(D14), 16663-16682.

Natraj, V., Liu, X., Kulawik, S.S., Chance, K., Chatfield, R., Edwards, D.P., Eldering, A., Francis, G., Kurosu, T.P., Pickering, K., Spurr, R., Worden, H., 2011. Multispectral sensitivity studies for the retrieval of tropospheric and lowermost tropospheric ozone from simulated clear sky GEO-CAPE measurements. Atmospheric Environment, 7151-7165. http://dx.doi.org/10.1016/j.atmosenv.2011.09.014.

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research Atmospheres*, *116*(12).

Nowlan, C. R., Liu, X., Leitch, J. W., Chance, K., González Abad, G., Liu, C., Zoogman, P., Cole, J., Delker, T., Good, W., Murcray, F., Ruppert, L., Soo, D., Follette-Cook, M. B., Janz, S. J., Kowalewski, M. G., Loughner, C. P., Pickering, K. E., Herman, J. R., Beaver, M. R., Long, R. W., Szykman, J. J., Judd, L. M., Kelley, P., Luke, W. T., Ren, X., & Al-Saadi, J. A. (2015). Nitrogen dioxide observations from the Geostationary Trace gas and Aerosol Sensor Optimization (GeoTASO) airborne instrument: retrieval algorithm and measurements during DISCOVER-AQ Texas 2013. *Atmospheric Measurement Techniques Discussions*, *8*(12), 13099-13155.

Ordóñez, C., Richter, A., Steinbacher, M., Zellweger, C., Nüß, H., Burrows, J. P., & Prévôt, A.

S. H. (2006). Comparison of 7 years of satellite-borne and ground-based tropospheric NO2 measurements around Milan, Italy. *Journal of Geophysical Research: Atmospheres*, *111*(5), 1–12. http://doi.org/10.1029/2005JD006305

Palmer, P. I., Jacob, D. J., Chance, K., Martin, R. V., Spurr, R. J. D., Kurosu, T. P., Bey, I., Yantosca, R., Fiore, A., & Li, Q. (2001). Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals form the Global Ozone Monitoring Experiment. *Journal of Geophysical Research*, *106*(D13), 14539–14550.

Pleim, J. (2006). A simple, efficient solution of flux-profile relationships in the atmospheric surface layer. *Journal of Applied Meteorology and Climatology*, (1), 341–347. http://doi.org/10.1175/JAM2339.1

Pleim, J. E. (2007a). A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *Journal of Applied Meteorology and Climatology*, *46*(9), 1383–1395. http://doi.org/10.1175/JAM2539.1

Pleim, J. E. (2007b). A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. *Journal of Applied Meteorology and Climatology*, *46*(9), 1396–1409. http://doi.org/10.1175/JAM2534.1

Pouliot, G., and T. E. Pierce (2009), Integration of the Model of Emissions of Gases and Aerosols from Nature (MEGAN) into the CMAQ modeling system. http://www3.epa.gov/ttnchie1/conference/ei18/session3/pouliot.pdf

Reed, A. J., Thompson, A. M., Kollonige, D. E., Martins, D. K., Tzortziou, M. A., Herman, J. R., Berkoff, T. A., et al. (2015). Effects of local meteorology and aerosols on ozone and nitrogen dioxide retrievals from OMI and pandora spectrometers in Maryland, USA during DISCOVER-AQ 2011. *Journal of Atmospheric Chemistry*, *72*(3-4), 455-482

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P. Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., & Woollen, J. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, *24*(14), 3624-3648.

Russell, A. R., Perring, A. E., Valin, L. C., Bucsela, E. J., Browne, E. C., Wooldridge, P. J., & Cohen, R. C. (2011). A high spatial resolution retrieval of NO2 column densities from OMI: Method and evaluation. *Atmospheric Chemistry and Physics*, *11*(16), 8543-8554.

Scarino, A. J., Obland, M. D., Fast, J. D., Burton, S. P., Ferrare, R. A., Hostetler, C. A., Berg, L. K., Lefer, B., Haman, C., Hair, J. W., Rogers, R. R., Butler, C., Cook, A. L., & Harper, D. B. (2014). Comparison of mixed layer heights from airborne high spectral resolution lidar, ground-based measurements, and the WRF-Chem model during CalNex and CARES. *Atmospheric Chemistry and Physics*, *14*(11), 5547-5560.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, X.Y. Huang, W. Wang, and

J. G. Powers, 2008: A description of the Advanced Research WRF version 3. *NCAR Tech. Note* NCAR/TN-475+STR, 113 pp.

Shin, H. H., & Hong, S. Y. (2011). Intercomparison of Planetary Boundary-Layer Parametrizations in the WRF Model for a Single Day from CASES-99. *Boundary-Layer Meteorology*, *139*(2), 261–281. http://doi.org/10.1007/s10546-010-9583-z

Stauffer, R. M., A. M. Thompson, & G. S. Young (2016), Tropospheric ozonesonde profiles at long-term U.S. monitoring sites: 1. A climatology based on self-organizing maps, *J. Geophys. Res. Atmos.*, 121, 1320–1339, doi:10.1002/2015JD023641.

Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., Prévôt, A.S.H., Hueglin, C. (2007), Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques, *J. Geophys. Res.-Atmos.*, D11307, doi:10.1029/2006JD007971.

Strode, S. A., Duncan, B. N., & Yegorova, E. A. (2015). Implications of model bias in carbon monoxide for methane lifetime. *Atmospheric Chemistry and Physics Discussions*, *15*, 20305–20348. http://doi.org/10.5194/acpd-15-20305-2015

Stull, R. B. (1988), *An Introduction to Boundary Layer Meteorology*, Springer, 9789027727688.

Sukoriansky, S., Galperin, B., & Perov, V. (2005). Application of a new spectral theory of stably stratified turbulence to the atmospheric boundary layer over sea ice. *Boundary-Layer Meteorology*, 1–30. http://doi.org/10.1007/s10546-004-6848-4

Tang, W., Cohan, D. S., Morris, G. a., Byun, D. W., & Luke, W. T. (2011). Influence of vertical mixing uncertainties on ozone simulation in CMAQ. *Atmospheric Environment*, *45*(17), 2898–2909. http://doi.org/10.1016/j.atmosenv.2011.01.057

Taubman, B. F., Hains, J. C., Thompson, A. M., Marufu, L. T., Doddridge, B. G., Stehr, J. W., Piety, C. A., & Dickerson, R. R. (2006). Aircraft vertical profiles of trace gas and aerosol pollution over the mid-Atlantic United States: Statistics and meteorological cluster analysis. *Journal of Geophysical Research: Atmospheres*, *111*(10), 1–14. http://doi.org/10.1029/2005JD006196

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., & van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys.*, 10, 11707-11735, doi:10.5194/acp-10-11707-2010, 2010.

Welton, E. J., Voss, K. J., Quinn, P. K., Flatau, P. J., Markowicz, K., Campbell, J. R., Spinhirne, J. D., Gordon, H. R., & Johnson, J. E. (2002). Measurements of aerosol vertical profiles and optical properties during INDOEX 1999 using micropulse lidars. *Journal of Geophysical Research: Atmospheres*, *107*(D19).

Wild, O., Zhu, X., & Prather, M. J. (2000). Fast-J: Accurate simulation of in-and below-cloud photolysis in tropospheric chemical models. *Journal of Atmospheric Chemistry*, *37*(3), 245-282.

Xiu, A., & Pleim, J. (2001). Development of a land surface model. Part I: Application in a mesoscale meteorological model. *Journal of Applied Meteorology*, 192–209. http://doi.org/10.1175/1520-0450(2001)040<0192:DOALSM>2.0.CO;2

Zaveri, R. A., & Peters, L. K. (1999). A new lumped structure photochemical mechanism for large-scale applications. *Journal of Geophysical Research: Atmospheres*, *104*(D23), 30387-30415.

Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., & Yao, Z. L. (2009). Asian emissions in 2006 for the NASA INTEX-B mission. *Atmospheric Chemistry and Physics*, *9*(14), 5131-5153.

Zhang, Y., Wang, Y., Chen, G., Smeltzer, C., Crawford, J., Olson, J., Szykman, J., et al. (2016). Large vertical gradient of reactive nitrogen oxides in the boundary layer: Modeling analysis of DISCOVER-AQ 2011 observations. *Journal of Geophysical Research: Atmospheres*, *121*(4), 1922-1934.

Zyrichidou, I., Koukouli, M. E., Balis, D. S., Katragkou, E., Melas, D., Poupkou, A., Kioutsioukis, I., van der A, R., Boersma, F. K., van Roozendael, M. & Richter, A. (2009). Satellite observations and model simulations of tropospheric NO2 columns over south-eastern Europe. *Atmospheric Chemistry and Physics*, *9*(2), 6119–6134. http://doi.org/10.5194/acp-9-6119-2009