Utilizing Operational and Improved Remote Sensing Measurements to Assess Air Quality Monitoring Model Forecasts

by

CHUEN-MEEI GAN

A dissertation submitted to the Graduate Faculty in Electrical Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2011

© 2011

CHUEN-MEEI GAN

All Rights Reserved

This manuscript has been read and accepted for the
Graduate Faculty in Engineering in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Dr. Barry Gross

Date

Chair of Examining Committee

Dean Mumtaz Kassir

Date

Executive Officer

Dr. Fred Moshary

Dr. YongHua Wu

Dr. Mark Arend

Dr. Elie Bou-Zeid

Dr. Jorge Gonzelez
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Utilizing Operational and Improved Remote Sensing Measurements to Assess Air Quality Monitoring Model Forecasts

by

CHUEN-MEEI GAN

Advisers: Professor Barry Gross (EE) and Professor Fred Moshary (EE)

Air quality model forecasts from Weather Research and Forecast (WRF) and Community Multiscale Air Quality (CMAQ) are often used to support air quality applications such as regulatory issues and scientific inquiries on atmospheric science processes. In urban environments, these models become more complex due to the inherent complexity of the land surface coupling and the enhanced pollutants emissions. This makes it very difficult to diagnose the model, if the surface parameter forecasts such as PM2.5 (particulate matter with aerodynamic diameter less than 2.5 μm) are not accurate. For this reason, getting accurate boundary layer dynamic forecasts is as essential as quantifying realistic pollutants emissions. In this thesis, we explore the usefulness of vertical sounding measurements on assessing meteorological and air quality forecast models. In particular, we focus on assessing the WRF model (12km x 12km) coupled with the CMAQ model for the urban New York City (NYC) area using multiple vertical profiling and column integrated remote sensing measurements. This assessment is helpful in probing the root causes for WRF-CMAQ overestimates of surface PM2.5 occurring both predawn and post-sunset in the NYC area during the summer. In particular, we find that the significant underestimates in the WRF PBL height forecast is a key factor in explaining this anomaly. On the other hand, the model predictions of the PBL height during daytime when convective heating dominates were found to be highly correlated to lidar derived PBL height with minimal bias. Additional topics covered in this thesis include mathematical method using direct Mie scattering approach to convert aerosol microphysical properties from CMAQ into optical parameters making direct comparisons with lidar and multispectral radiometers feasible. Finally, we explore some tentative ideas on combining visible (VIS) and mid-infrared (MIR) sensors to better separate aerosols into fine and coarse modes.

Acknowledgement

It has been a long journey to achieve my Ph.D. degree. On this journey, I had encountered numerous difficulties but my family and mentors, Dr. Barry Gross and Dr. Fred Moshary, have given their full supports to me. I really appreciate their support and guidance which gave me the courage to complete my degree. There are two special persons that I would like to thank, Mrs. Regenia Cafini and Dr. Koh Yit Yan, because they encouraged me to pursue my Ph.D. I also want to thank Dr. YongHua Wu for his guidance and assistance in my research work.

Most importantly, I would like to thank my husband, William Jerez, for his patience and support throughout my life. He shows me that everything is possible; I will be able to accomplish anything. Furthermore, he has taken good care of me when I was sick or feeling depressed. I'm so grateful to have him by my side every day.

I would also like to express my thanks to the NSF MIRTHE and ISET programs for helping support my work and CUNY for providing me with fellowship support as well as allowing me to teach Electrical Engineering courses. This has further developed my self-confidence and appreciates teaching as well as learning.

Last but not least, I want to express my heartfelt appreciation to everyone who lent a hand to me in my PhD journey. Without them, I would not be in this position

Content

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 7

CHAPTER 8

List of Tables

List of Figures

INTRODUCTION

Air quality models such as the Weather Research and Forecast (WRF) model coupled with the Community Multiscale Air Quality (CMAQ) model are commonly used to provide near term (24 hours) forecasts that are used to support air quality applications ranging from regulatory health issues to purely scientific inquires on atmospheric science and boundary layer processes. Air quality parameters of interest that are driven by the meteorology and the transport processes (e.g. physical and chemical) include tropospheric ozone, acid deposition, particulate matter etc. While a number of large scale studies have been made to assess performance of these models in regard to surface parameters such as temperature, water vapor, winds, ozone, PM2.5 etc (Simpson et al., 2006; Hogrefe et al., 2007; Zhang et al., 2009; Doraiswamy et al., 2010), such studies do not provide insight into the anomalies that may appear in the model results. In particular, there is always the underlying question whether the forecast errors are due more to the errors in emission inventories, or in the transport processes. Furthermore, there has been comparatively little work focused analysis on assessing models in highly dense urban environments, but such studies are extremely valuable since application of these models in urban environments will depend strongly on the inherent complexity of the land surface coupling and more complex primary emission sources pushing the models to their limit.

In focusing on particulate matter, the model forecast of surface PM2.5 is not a simple matter. For example, it was shown (Dorismarty et al., 2010) that the urban PM2.5 forecasts from WRF-CMAQ particularly in summer; strong diurnal errors occurred in the PM2.5 forecasts even on a seasonally averaged scale. In particular, the model forecasts showed strong pollution spikes both during predawn and post-sunset periods, and the study was not able to unambiguously identify the source of this anomaly. Moreover, it is obvious to see that getting the meteorology and the boundary layer dynamic forecasts correct (e.g. Mixing Layer Height) is bound to be as strong factor as quantifying the pollutants emissions. Clearly, progress on this question would be made if the dynamics of the boundary layer can be studied directly using remote sensing measurement techniques. In this way, it should be possible to assess the different factors and provide guidance to further model development.

To address the above specific problem which was the underlying motivation of our activities; in this thesis, we focus on developing and improving remote sensing tools in order to directly address WRF-CMAQ performance difficulties encountered in the NYC area. However, note that this study was limited by the operational performance outputs of the New York State Department of Environmental Conservation (NYSDEC). Their operational forecast products at present even over the NYC area are at 12 km grid resolution. Of course, the NYSDEC as well as CCNY are experimenting with higher resolution "urbanized" products and these were tested against our measurements on small datasets. Our preliminary matchups with these models almost always were less accurate than the lower resolution product. This seems to be a symptom of the instability in the urban WRF model producing spatial vortices etc.

To this end, the extensive simulations of the WRF-CMAQ performed within the urban NYC area are assessed with the aid of vertical profiling and column integrated remote sensing measurements which allow us to critically assess the performance of the model and to interpret the root causes of possible anomalies. In making these assessments, we focus on several topics listed below.

1) Assessing planetary boundary layer (PBL) dynamics of WRF model under highly convective conditions.

To begin, we find that when turbulent mixing processes are dominant, the WRF derived PBL height exhibits a good linear correlation with lidar derived PBL height. In making these comparisons, we found that an excellent remote sensing approach was to estimate the PBL height from near-infrared (NIR) lidar measurements using a Wavelet Covariance Transform (WCT) approach, which is further modified to better isolate the convective layer from the residual and plume. In selecting the modified WCT approach, we compare the results against other PBL height retrieval methods (e.g. gradient and maximum variance algorithms) to illustrate the improvements. Furthermore, the WRF-Lidar PBL height comparisons are made using a number of different PBL parameterization schemes including Asymmetric Convective Model – version 2 (ACM2), Modified Blackadar (BLK) and Mellor-Yamada-Janjic (MYJ) schemes. Our finding shows that the correlations for all these runs were high.

2) Assessing CMAQ PM2.5 forecasts.

We apply continuous 24 hours / 7 days ceilometer measurements to assess WRF-CMAQ model forecasts of surface PM2.5 (particulate matter has diameter < 2.5 μm). We observe strong overestimations in the surface PM2.5 mass (μg/m^3) particularly in the summer during predawn and post-sunset periods, which seems to be more strongly connected to the underestimations of PBL heights and less to the enhanced pollutants emissions. This interpretation is consistent with the PM2.5 observations of TEOM (Tapered Element Oscillating MicroBalance). These PM2.5 measurements are better correlated to path integrated CMAQ PM2.5 mass during these periods. This is critical in urban model setting due to the obvious enhancements of primary emission sources and the density of individuals that can be adversely affected by high levels of PM2.5.

3) Using passive remote sensors to assess CMAQ aerosol optical properties.

While active measurements can provide estimates of intensive parameters, such measurements are very rare in the community. However, it is still possible to explore vertical performance of WRF-CMAQ forecasts using more common aerosol optical depth (AOD) column measurements. In addition, using multispectral information may provide better insight into aerosol speciation and production inconsistencies within the model. To that end, we develop a preliminary model that converts CMAQ particulate predictions into multispectral optical extinction (and backscatter) measurements base on Mie scattering theorem, which may be compared to passive radiometric measurements. This model is formulated based on the detailed microphysical size distribution available within the CMAQ aerosol outputs as well as the reflective index data for different aerosol classes obtained within the Optical Parameters of Aerosols and Clouds (OPAC) package. Given this microphysical formulation, the optical scattering (extinction) coefficients can be calculated at any wavelength and vertically integrated to obtain the relevant column integrated AOD. In this procedure, the effect of humidification on the aerosol is also included since for certain aerosol classes, high relative humidity (RH) will modify both the size and reflective index of these particles (e.g. hygroscopic processes). In particular, we make preliminary tests on column integrated optical properties (e.g. AOD and Angstrom coefficients) of the models against the AERONET radiometer measurements for multiple days and compare against the semi-empirical method of Malm approach. Our model results seem to have better correlation with the AERONET AOD measurements than the Malm model.

4) Application of mid-infrared (MIR) lidar system in aerosol study.

Finally, we explore possible uses of combining MIR measurements with existing near-infrared (NIR), visible (VIS) and ultra-violet (UV) measurements to better separate fine and coarse mode of aerosols. In particular, based on realistic system parameters, by combining a VIS-MIR system with high power MIR systems, aerosol in the troposphere can be better separated into fine and coarse modes since a wider range of wavelength spectra is applied.

The structure of the thesis is as follows: The fundamentals of aerosol formation and their physical properties together with the basics of PBL meteorology are introduced in Chapter 1. In addition, a review of current air quality monitoring systems and EPA standards are presented. In Chapter 2, a detailed description of the relevant remote sensing instruments is presented and their working principles discussed. In particular, since the CCNY lidar system is one of the most important instruments for observing aerosol and PBL dynamics, the details of lidar data processing including calibration methods are presented. In Chapter 3, a modified NIR lidar based PBL retrieval algorithm (WCT), which is designed to better separate convective from residual and plume layers, is described. Moreover, its improved performance compared to other lidar based PBL retrieval algorithm is presented. In Chapter 4, the introduction of the air quality models (WRF and CMAQ) are given together with the details of the parameterizations of both models that we will test is described. In Chapter 5, the model forecasts for the PBL height and the vertical structure of the PM2.5 are compared against a variety of remote sensing measurements, to better understand the capabilities and the limitations of the WRF-CMAQ model. In Chapter 6, the details in converting the microphysical parameterizations of CMAQ into optical scattering properties are given and the preliminary comparisons with AERONET data and lidar measurements are made. In Chapter 7, we explore the possible use of MIR sensors to improve aerosol speciation into fine and coarse modes. In particular, we study the feasibility of combining path integrated aerosol extinction from VIS and MIR to improve this separation. Finally, our summary and conclusions are given in Chapter 8.

CHAPTER 1

Meteorology Background and Air Quality Regulation

First, we go through the relevant properties of aerosol such as species, modes, and effects (direct and indirect) on climate, atmospheric environment, ecosystem and human health. In the following section, the structure and dynamic of PBL is explained in detail, including their characteristic for each component. In the same time, we will remark a few concerns of PBL features in this study. Lastly, current air quality monitoring system and standards are outlined.

1.1 Aerosol

Aerosols are defined as solid or liquid particles suspended in the air and are composed of a number of different natural or man-made substances. They are formed through the dispersal of material at the Earth's surface or the reaction of gases in the atmosphere. The three main types of aerosol are volcanic, desert dust, and anthropogenic (man-made) with their sizes ranging from roughly 1 nm to 100 µm, which are generally divided among three main aerosol size modes (Aitken, fine and coarse). A detail of the formation of these modes is given later. The lifetime of an aerosol particle in the atmosphere can be extend up to weeks, or in some extreme cases, it can be over a year or more (e.g. volcanic ash in the stratosphere). Normally, the smallest particles lie on the boundary between molecular or atomic cluster, and its lifetime is limited by break-up or coagulation with other particles. For the big particles, its lifetime depends on the deposition rate. Aerosols are of particular interest in climate studies since they form the largest uncertainties within existing climate models (URL [1 & 2]). In fact, aerosol uncertainties are much larger than uncertainties that are mainly prescribed to Green House Gases (see Fig. 1.1).

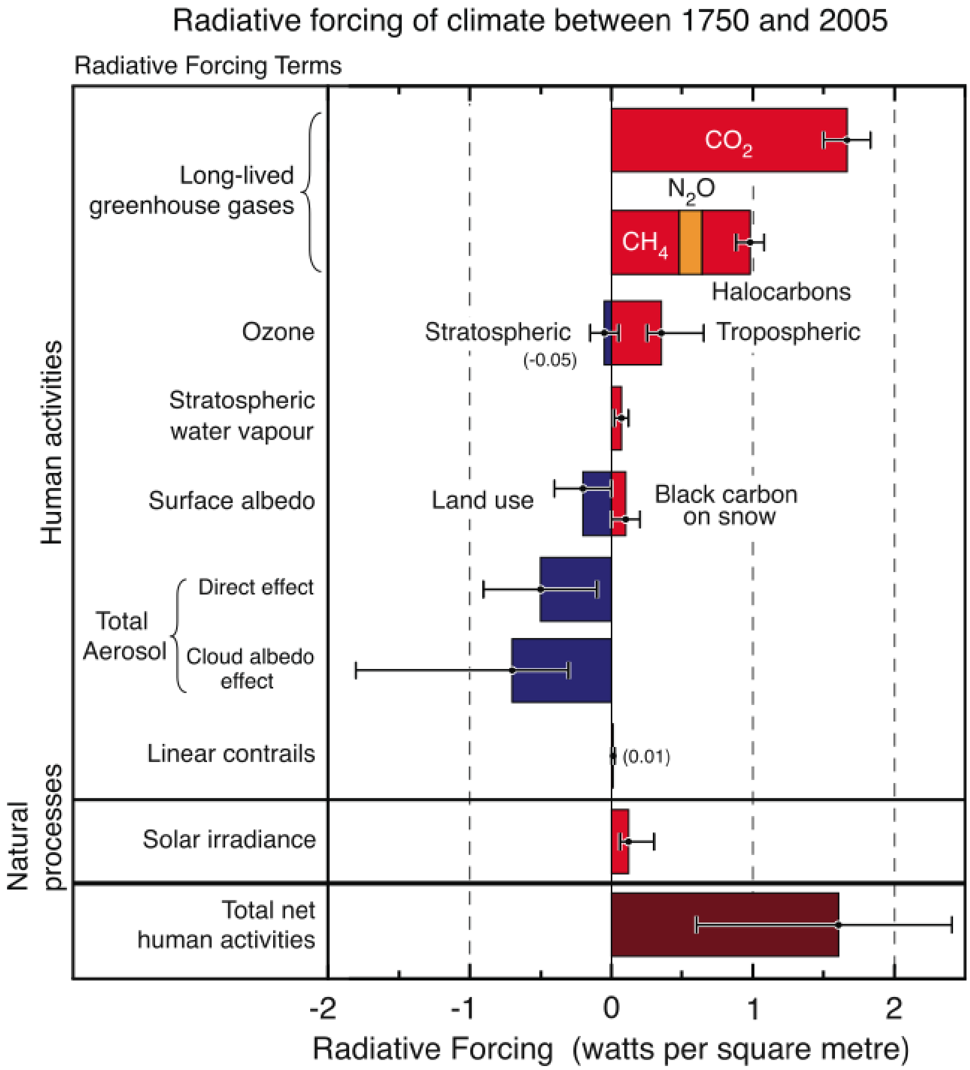


Figure 1.1: Summary of the principal components of the radiative forcing of climate change. All these radiative forcing result from one or more factors that affect climate and are associated with human activities or natural processes. The value represents the forcings in 2005 relative to the start of the industrial era (about 1750). Human activities cause significant changes in long-lived gases, ozone, water vapor, surface albedo, aerosols and contrails. The only increase in natural forcing of any significance between 1750 and 2005 occurred in solar irradiance. Positive forcings lead to warming of climate and negative forcings lead to a cooling. The thin black line attached to each colored bar represents the range of uncertainty for the respective value (adapted from Figure 2.20 of IPCC 2007 report).

Aerosols have a direct effect on planetary energy balance. Unfortunately quantifying this effect is very complicated since aerosols can either scatter sunlight back into space (resulting in a net cooling mechanism) or absorb sunlight with different efficiency (warming the atmosphere). In fact, separating non-absorbing aerosols from absorbing aerosols using satellites for global measurements is difficult because passive techniques based on solar reflected sunlight are ill suited to separate absorption from scattering. In addition to the direct effects, aerosols can strongly affect the climate indirectly by interacting and changing the properties of clouds. For example, aerosols are necessary for cloud formation, acting as seeds to start the formation of cloud droplets. Increases in aerosol concentrations cause water in a cloud to spread over many more particles, each of which is correspondingly smaller. Distributing the cloud droplets to smaller size distributions tends to make the clouds more optically scattering (e.g. reflective), thereby leading to an overall cooling mechanism. However, the indirect effects are even more subtle. Smaller particles fall more slowly in the atmosphere since they are lighter and tend to stay in the atmosphere longer and thereby decrease the amount of overall rainfall. Therefore, the changing of aerosols in the atmosphere can alter both the frequency of cloud occurrence as well as the optical reflection properties and associated rainfall amounts.

Furthermore, aerosols pose a number of unique health risks among humans, animals and plants. For example, deposition of aerosols can redistribute the chemicals which can lead to altering of ecological systems. For example, aerosol types such as sulfur dioxide and nitrogen oxides can form cloud droplets which cause acid rain forcing significant regulatory efforts (e.g. control the emission of sulfur dioxide from industries) to reduce acid rain.

In addition to possible aerosol toxicity, tiny aerosols can also affect human health directly. Fine particles can follow the inhaled air far enough down the respiratory system and deposit deep in the lung which can be harmful for human health (e.g. asthma, bronchitis, acute and chronic respiratory symptoms such as shortness of breath and painful breathing, and premature deaths). On the other hand, coarse particles do not follow the air flow so easily, but are deposited early in the respiratory system and then removed from the body with mucus. Because of the particular health risks associated with finer particles, the EPA has defined a number of aerosol classes which are required to be monitored continuously. For example, PM2.5 is defined as the particulate matter with aerodynamic diameter less than 2.5 µm and PM10 is for particles with diameter less than 10 µm. They are measured in term of particulate mass per volume (unit of µg/m3). Fig. 1.2 shows how the particles are realistically distributed between different size modes. These include the nucleation or Aitken mode (diameter ≤ 0.1 µm) and the accumulation mode ($0.1 \leq$ diameter ≤ 2.5 µm). The nucleation mode particles are formed from homogenous nucleation and gas-to-particle conversion and are very short lived. This is due to the fact that they are very mobile because of their ultrafine size, which results in rapid coagulation. This coagulation mechanism leads to the formation of bigger particles which eventually end up in the accumulation mode. Particles within the accumulation mode are less variable since the only effective deposition processes for the accumulation mode are rainout and washout which acts very slowly. Pollen, dust and sea salt are considered as coarse mode aerosols and normally come from natural sources. (Seinfeld & Pandis, 1988; Olofson, F., 2008)

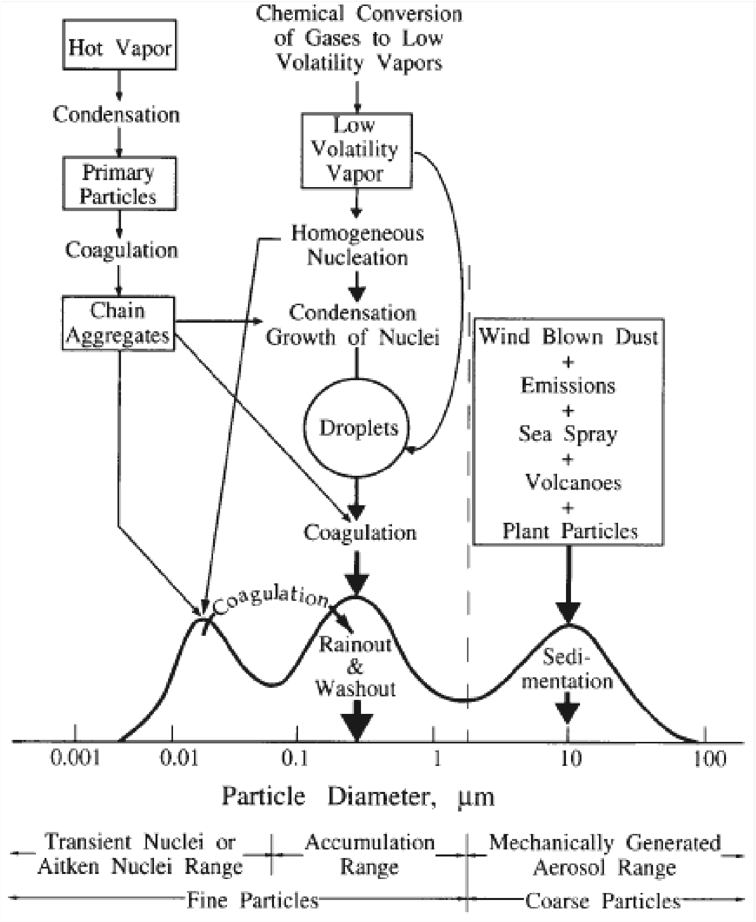


Figure 1.2: Overview of aerosol size classification with main sources and sinks (Atmospheric Physics and Chemistry, Seinfeld & Pandis, 1988).

1.2. Planetary Boundary Layer (PBL)

In order to better understand the transport processes of pollutants from primary emission sources (e.g. factories, vehicles, etc.); it is critical to know the structure and the dynamics of the PBL (also known as mixing layer) which is the lowest 100 to 3000 meters of troposphere. The remainder of the air in the upper level of troposphere is called the free atmosphere. Unlike the free atmosphere which is not coupled to surface processes, the PBL is defined as the part of troposphere which is directly influenced by the land / ocean surface, and responds to surface forcings with a timescale of about an hour or less (Stull, 1988). These forcings include frictional drag, evaporation and transpiration, heat transfer, pollutant emission and terrain induced flow modification. The depth of the PBL is complex, and varies in time and space quickly, ranging from hundreds of meters to a few kilometers depending critically on both the diurnal cycle as well as the land surface characteristics. The three major components of PBL (see Fig. 1.3) are the mixed layer, the residual layer and the stable boundary layer. Characteristics of each layer will be further discussed in the following sub-section (Stull, 1988; URL [3 & 4]).

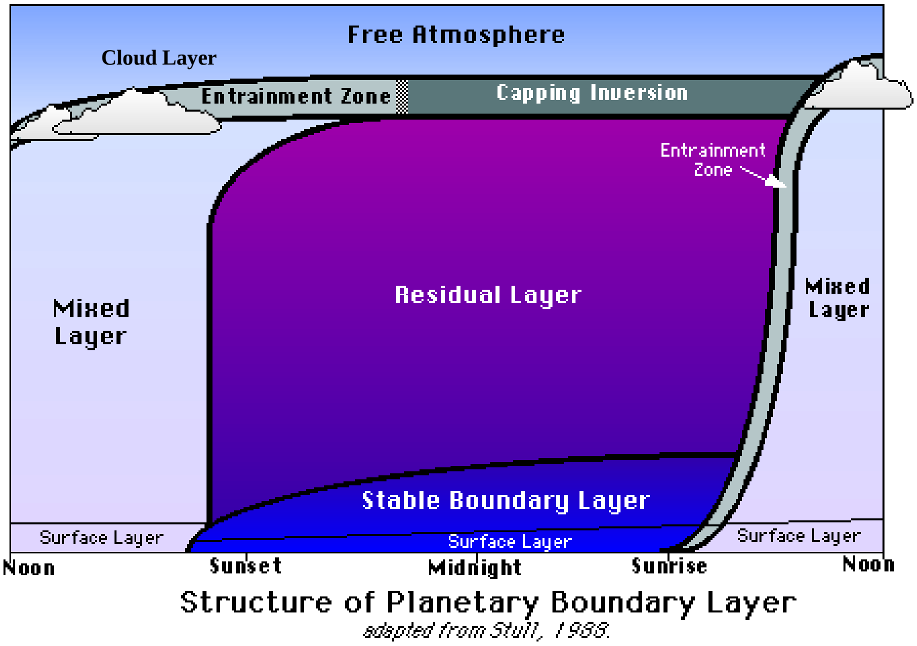


Figure 1.3: Structure of Planetary Boundary Layer (Stull, 1988).

1.2.1 Mixed Layer (ML)

The strongest coupling of the ground to the atmosphere is the diurnal heating / cooling of the ground surface which drives heat, moisture and energy into the atmosphere. It is important to note that the coupling is not caused by the direct forcing of solar radiation on the boundary layer which is a weak process, but by the warming of ground surface where absorptivities on the order of 90% occur which then couples advectively with the atmosphere providing a buoyant conduit in which energy, momentum and moisture (as well as pollutants) may be transported into the boundary layer.

In general, the process is qualitatively simple. Starting about a half hour after sunrise, a turbulent ML begins to grow in height as buoyant thermal forces create a temperature gradient which pushes surface air upward. This convective is characterized by intense turbulent mixing where warm thermal flows continuously rise from the surface. This mechanism in general continues until it reaches a maximum height during mid-late afternoon. At the same time, there is a simultaneous mixing process from the free atmosphere where the cooler (dryer) air tries to mix with the warmer-moister air from below. This process is called as entrainment and results in unstable turbulent processes which tend to mix heat, moisture and momentum uniformly over the vertical extent of the ML. In addition, during this entrainment process, enhanced moisture can mix with cold air forming cloud capping layers and initiate cloud formation. Furthermore, moisture in the air has the ability to retain large amounts of heat energy, which can then be transferred throughout the layer via turbulent mixing. Therefore, air with high humidity will tend to have a thinner ML than air with little moisture present, due to the fact that it takes longer time to heat the air as the moisture absorbs more energy. Thus, air with high moisture content also prevents heat energy from escaping during the night, the height of the ML will be greater that of air with low humidity. Moisture tends to decrease variability in air temperature.

The turbulence in the ML is convectively driven, but a nearly well mixed layer can form in regions of strong winds. Convective sources include heat transfer from a warm ground surface, and radiative cooling from the top of the cloud layer. The first situation creates thermals of warm air rising from the ground, while the second creates thermals of cool air sinking from cloud top. Both processes can occur simultaneously, particularly when a cool stratocumulus topped ML is being advected over warmer ground. Even when convection is the dominant mechanism, there is usually wind shear across the top of the ML that contributes to the turbulence generation.

Virtual potential temperature profiles are nearly adiabatic in the middle portion of the ML, while in the surface layer one often finds a superadiabatic layer adjacent to the ground. A stable layer at the top of the ML acts as a lid to the rising thermals, thus restraining the domain of turbulence. This stable layer is called the entrainment zone because cooler air mixes down from the upper layers often bringing moisture and particulates into the ML. This capping stable layer is strong enough to be classified as a temperature inversion (e.g. the absolute temperature increases with height). In fact, it is frequently called an inversion layer regardless of the magnitude of the stability.

Another feature of convective ML is the air flow. Wind speeds are subgeostrophic throughout the ML, with wind directions crossing the isobars at a small angle towards low pressure. The middle portion of the ML frequently has nearly constant wind speed and direction. However, wind speeds decrease towards zero near the ground, resulting in a wind speed profile that is nearly logarithmic with height in the surface layer. Wind directions cross the isobars at increasingly large angles as the ground is approached, with 45 degree angles not uncommon near the surface. In addition, the wind is strongly affected by surface roughness effects which result in a dragging force on the near surface winds as well as modifying the transport of water vapor and energy. However, the different mechanisms are affected differently with momentum being the most inhibited by large surface roughness parameters in relation to the others.

Moisture is another trace of the ML and strongly affects its dynamics. Since the evaporation of soil and plants occurs from below and the entrainment of drier air is from above, the water vapor mixing ratios tend to decrease with height, even within the center portion of the

ML. The decreasing of moisture across the top of the ML is very pronounced, and is often used together with potential temperature profiles to identify the ML top from radiosonde soundings.

The main concern for this study is the distribution of pollutants within the ML. Pollutant concentrations can build up in the ML while free atmosphere concentrations remain relatively low because most pollutant sources are near the ground surface. Pollutants are transported by eddies such as thermals; so the inability of thermals to penetrate very far into the stable layer means the stable layer acts as a lid to the pollutants, which limit the lifting of pollutants. Trapping of pollutants below such as "inversion layer" is common in high-pressure regions, and sometimes leads to pollution alerts in large communities. This sharp contrast in pollutant levels is critical in our use of lidar method to monitor the PBL dynamics.

1.2.2 Residual Layer (RL)

About a half hour before sunset, the thermals cease to form (in the absence of cold air advection), allowing turbulence to decay in the formerly well mixed layer. In the absence of the turbulence, some particulates will remain exist in the atmosphere since their deposition rates are slow and the resulting particulate layer in the atmosphere is called the residual layer. For example, in the absence of advection, passive tracers dispersed into the daytime mixed layer will remain aloft in the RL during the night. The RL is neutrally stratified, resulting in turbulence that is nearly of equal intensity in all directions. Therefore, smoke plumes emitted into the RL tend to disperse at equal rates in the vertical and lateral directions, creating a cone-shaped plume.

Nonpassive pollutants may react with other constituents during the night to create compounds that were not originally emitted from the ground. Sometimes gaseous chemicals may react to form aerosols or particulates which can precipitate out. The RL often exists in the mornings for a short time of period before being entrained into the new ML. During this period, the solar radiation may trigger photochemical reactions among the constituents in the RL. Moisture often behaves as a passive tracer. More moisture may be evaporated into the ML and will be retained in the RL each day. During succeeding days, the re-entrainment of the moist air into the ML might allow cloud formation to occur where it otherwise might not.

Virtual potential temperature usually decreases slowly during the night because of radiation dissipation. This cooling rate is on the order of 1°C/day. Since the cooling rate is more-or-less uniform throughout the height of the RL, allowing the RL virtual potential temperature profile to remain nearly adiabatic. The RL does not have direct contact with the ground, so when the top of the next day ML reaches the base of the RL, the ML growth becomes very fast. During the night, the nocturnal stable layer gradually increases in thickness by modifying the bottom of the RL. The remainder of the RL is not affected by turbulent transport of surface-related properties and hence does not really fall within the definition of a boundary layer.

In describing the RL, it is clear that it may lead to confusion with the ML when particulates (aerosol) are used for PBL height retrieval. For this reason, it is important to devise a method which can isolate the convective mixing layer from the residual layer which is not modeled with the WRF-CMAQ environment. Such an approach is explained in Chapter 3.

1.2.3 Stable Boundary Layer (SBL)

When the surface is cooler than the air, the boundary layer becomes stably stratified which always occur at night over land. This boundary layer is called as stable boundary layer, where it is known as a nocturnal boundary layer (NBL). It can also form by advection of warmer air over a cooler surface.

As the night progresses, the bottom portion of the RL is transformed by its contact with the ground into a SBL. This is characterized by statically stable air with weaker, sporadic turbulence. Even though the wind at ground level frequently becomes lighter or calm at night, the winds aloft may accelerate to supergeostrophic speeds in a phenomenon that is called the low-level jet or nocturnal jet. The statically stable air tends to suppress turbulence, while the developing nocturnal jet enhances wind shears that tend to generate turbulence. Thus, turbulence sometimes occurs in relatively short bursts that can cause mixing throughout the SBL. During the nonturbulent periods, the flow becomes essentially decoupled from the surface. As opposed to the daytime ML which has a clearly defined top, the SBL has a poorly-defined top that smoothly blends into the RL above. The top of the ML is defined as the base of the stable layer while the SBL top is defined as the top of the stable layer or the height where turbulence intensity is a small fiction of its surface value.

1.2.4 Seasonal Dependencies

In general, at night and in the cool season the PBL tends to be lower while during the day and in the warm season it tends to be higher. This is because of the wind speed and thickness of the air is a function of temperature. Since strong wind speeds allow for more convective mixing, this convective mixing will cause the PBL to grow. At night, the PBL contracts due to a reduction of rising thermals from the surface. Cold air is denser than warm air so the PBL will tend to be shallower in the cool season. Besides, the daytime is longer in the summer than in the winter which will cause surface heating occur longer. Indirectly, the temperature in the warm season is higher resulting higher PBL. Such behavior is amply demonstrated with our multiyear lidar measurements that demonstrate deep PBL during summer and shallow PBL during winter months.

1.2.5 Urban Modification to the Conventional Boundary Layer Dynamics

The above description needs refinement when urban areas are considered. In particular, the PBL in the urban area is less likely to collapse quickly during the night, even the solar heating is absent due to the large heat capacity of buildings and streets, and the heat released from vehicles and space heating. Early evening shallow SBL in the neighboring rural countryside are not observed in the city. Later at night, when the rural stable layer is lower than the height of urban buildings, a shallow layer of air within the city can remain well mixed, but it is capped by a stable layer. This phenomenon is called as urban heat island. (URL [5 & 6]) Such persistence in the PBL in urban conditions is indirectly observed in our lidar measurements (especially in the late afternoon) and is not well captured by the WRF model. However, due to the difficulties in running lidar at night, these datasets are incomplete (insufficient of long period of night measurements) so strong conclusions are hard to make.

In addition, the urban surface is rough and as such acts to inhibit transport processes such as horizontal momentum, moisture and kinetic energy. This roughness scale is an extremely important parameter in many PBL parameterization models and strongly modifies the latent heat and moisture fluxes during convective conditions (Flagg & Taylor, 2008). Roughness capture can often modify the water uptake making the air masses dryer and therefore allowing a greater thermal expansion of the PBL height. The increased latent heat together with this roughness mechanism can explain the large PBL height observed in the NYC area often approaching 3km.

1.2.6 Measurement Signatures of PBL Height

Since the PBL height during convective conditions is well mixed though turbulent mechanisms, the PBL height can often be determined using thermodynamic sounding. The top of the PBL is often marked with a temperature inversion, a change in air mass, a hydro lapse, and change in wind speed and / or a change in wind direction. Inversions traps air within the PBL and do not allow convection to occur into the middle and upper atmosphere. Inversions above the PBL are referred to as CAPs. The PBL is most definable in situations where differential advection is occurring or when a shallow front is at the surface. At the top of a front, there is an abrupt change in air mass. In some cases the transition between the PBL and free atmosphere is not well defined. However, the general height of the PBL can be determined by looking for subtle changes in dewpoint and wind speed / direction (e.g. HYSPLIT achieve meteorology sounding and microwave radiometer measurements). Two study cases are presented at Section 5.2. During the day, the PBL often mixes out to the dry adiabatic lapse rate, especially on clear days. At nights with clear skies the opposite occurs. The surface radiative cools, creating a large temperature inversion throughout the entire PBL

In additional, meteorologically, PBL height also can be found by using bulk Richardson number which is an approximation to the gradient Richardson number formed by approximating local gradient by finite difference across layers. The bulk Richardson number, R_B, defined as below is dimensionless number in meteorology relating vertical stability and vertical shear (generally, stability divided by shear).

$$R_B = \frac{\frac{g}{T_v}\Delta\overline{\theta_v}\Delta z}{\Delta\bar{U}^2+\Delta\bar{V}^2} \tag{1.1}$$

where g is gravitational acceleration, T_v is absolute virtual temperature, $\Delta\overline{\theta_v}$ is the virtual potential temperature difference across a layer of thickness Δz, and $\Delta\bar{U}$and $\Delta\bar{V}$ are the changes in horizontal wind components across that same layer. It represents the ratio of thermally produced turbulence and turbulence generated by vertical shear. In the limit of layer thickness becoming small, the bulk Richardson number approaches the gradient Richardson number, for which a critical Richardson number is roughly $Ri_c = 0.25$. Gradient Richardson numbers less than this critical value are dynamically unstable and likely to become or remain turbulent. Unfortunately, a critical value is not well defined for the bulk Richardson number, leading to uncertainty in turbulence likelihood for values near the critical value. This technique is implemented by the Blackadar scheme in PBL simulation. Also, ACM2 scheme use this technique together with eddy diffusion component for PBL simulation (Nielsen-Gammon, 2010; Pleim, 2007a & 2007b).

Meanwhile, the TKE method is implemented in the MYJ scheme to probe the turbulent terms directly. The prognostic TKE differential equation at each layer can be solving iteratively:-

$$\frac{d\left(\frac{q^2}{2}\right)}{dt} - \frac{\partial}{\partial z}\left(\ell q S_q \frac{\partial\left(\frac{q^2}{2}\right)}{\partial z}\right) = P_s + P_b - \varepsilon \tag{1.2}$$

$$P_S = K_M\left(\frac{\partial U}{\partial z}\frac{\partial U}{\partial z} + \frac{\partial V}{\partial z}\frac{\partial V}{\partial z}\right) \tag{1.3}$$

$$P_b = -\beta g K_H\left(\frac{\partial \theta_v}{\partial z}\right) \tag{1.4}$$

$$\varepsilon = \frac{q^3}{B_1 \ell} \tag{1.5}$$

where $\frac{q^2}{2}$ is the TKE, K_M (K_H) is the vertical turbulent exchange coefficient for momentum (heat), $S_q=0.20$, $\beta=1/273$, B_1 is an experimental constant and all other symbols have common meanings. The shear production (P_s), buoyant production (P_b) and dissipation (ε) and a small vertical diffusion term comprise the components of the prognostic TKE. A threshold of TKE = 0.2 m^2 s^{-2} determines the top of the PBL (Flagg & Taylor, 2008).

Finally, we would again emphasize that in this study, we are particularly interested in cases where strong convective turbulence occurs without a lot of interference by clouds and very strong wind. This allows us to reduce the complexity in the model interpretation as well target more severe heated and stagnant conditions not conducive to good air quality. These conditions occur mainly in the summer and further justify our focus on this period.

1.3 Air Quality Monitoring and Standards

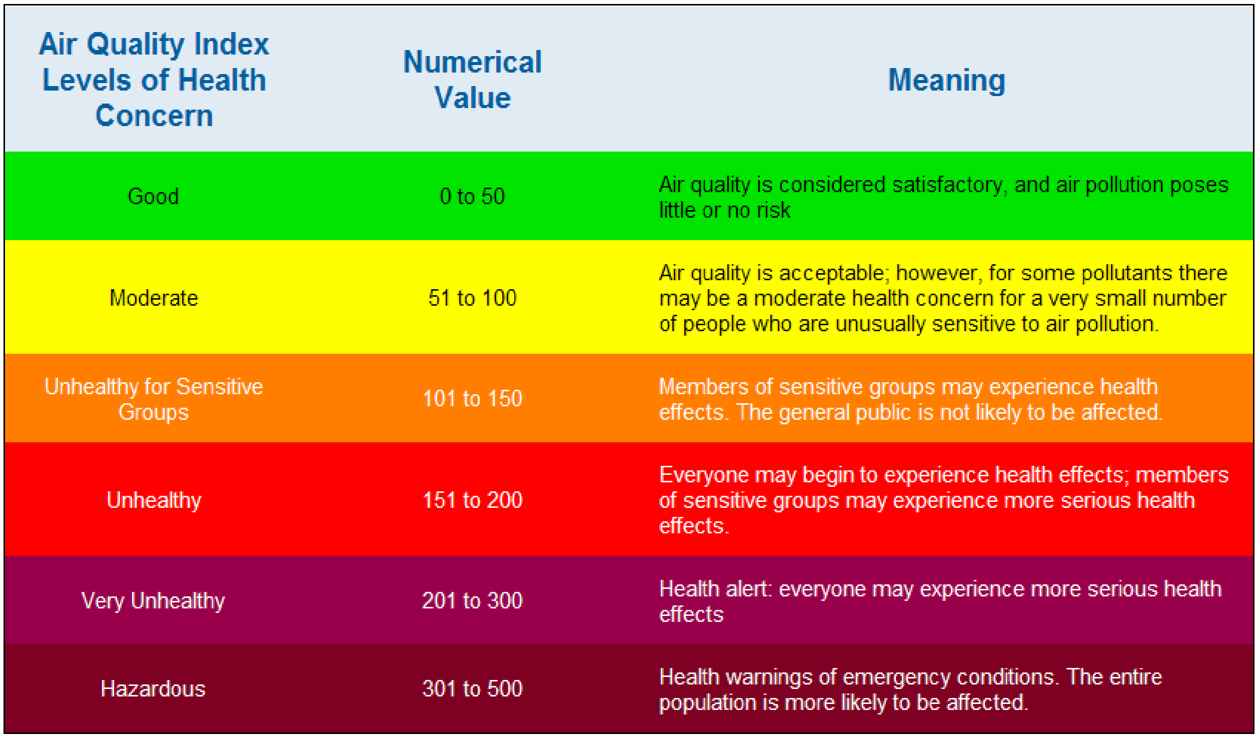
Based on the health concerns described in Section 1.1, under the authority of the Clean Air Act (CAA), the U.S. Environmental Protection Agency (EPA) has established National Ambient Air Quality Standards (NAAQS) to protect human health and the environment from high levels of criteria pollutants, such as ozone and PM. This NAAQS is often represented by a number of Air Quality Index (AQI) classes with different color so that public can understand the designation most easily (see Fig. 1.4).

The AQI was initially defined as piecewise linear function of the pollutant concentration and therefore has discontinuous jumps between different health risk classes. We can convert the pollutant concentration to AQI by this equation:

$$I = \frac{I_{high} - I_{low}}{C_{high} - C_{low}}(C - C_{low}) + I_{low} \tag{1.6}$$

where I is the air quality index, C is the pollutant concentration, C_{low} is the concentration breakpoint that is $\leq C$, C_{high} is the concentration breakpoint that is $\geq$ C, I_{low} is the index breakpoint corresponding to C_{low} and I_{high} is the index breakpoint corresponding to C_{high}.

Different pollutants have different breakpoints and Table 1.1 is the EPA's list of breakpoints for PM2.5.



Air Quality Index Levels of Health Concern	Numerical Value	Meaning
Good	0 to 50	Air quality is considered satisfactory, and air pollution poses little or no risk
Moderate	51 to 100	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
Unhealthy	151 to 200	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy	201 to 300	Health alert: everyone may experience more serious health effects
Hazardous	301 to 500	Health warnings of emergency conditions. The entire population is more likely to be affected.

Figure 1.4: EPA has assigned a specific color to each AQI category to make it easier for people to understand quickly whether air pollution is reaching unhealthy levels in their communities.

Table 1.1 List of breakpoints for PM2.5

C_{low}	C_{high}	I_{low}	I_{high}	***Category***
0	15.4	0	50	Good
15.5	40.4	51	100	Moderate
40.5	65.4	101	150	Unhealthy for Sensitive Groups
65.5	150.4	151	200	Unhealthy
150.5	250.4	201	300	Very Unhealthy
250.5	350.4	301	400	Hazardous
350.5	500.4	401	500	Extremely Hazardous

In addition, EPA, NOAA, NPS and local agencies have established an online network called as AIRNow, which is providing real-time national air quality information daily for the public over 300 cities across the U.S. This air quality information can be used to generate spatial maps and forecasts which are collected using either federal reference or equivalent monitoring techniques or techniques approved by the state or local agencies. The data are displayed on the website (http://airnow.gov) after the end of each hour so it can be “real-time”. Therefore the data as such are not fully verified and validated through the quality assurance procedures monitoring organizations, which are always officially submitted and certified on the EPA NAAQS, even though some preliminary data quality assessments are performed. Thus, these data are not used to formulate or support regulation, guidance or any other Agency decisions or positions as it is for the purpose of reporting the AQI. The AQI represents how clean or unhealthy of the air and what associated health effects might result (see Fig.1.4).

CHAPTER 2

Remote Sensing Instrumentations

This chapter will introduce remote sensing instruments (lidar, ceilometer, sunphotometer and microwave radiometer) which are used in this research. Their working principles are explained and specifications are listed. Since lidar is the key instrument in this research, its calibration methods and data processing are discussed.

2.1 Lidar

The principle of the lidar technique for finding the quantity and the location of target of interest, is to illuminate the atmosphere with pulsed electromagnetic radiation (e.g. light) and then collect the part of the emitted light which is scatter back to a receiver (e.g. at 180^o) in strong analogy with radar echo detection (Kovalev & Eichinger, 2004). However, due to the nature of the very small size of target (e.g. sub micron), radio wave (or microwaves) are not suitable and short wavelength sources are essential. For this reason, a pulse laser is used at specific wavelengths (e.g. Nd:YAG laser at 1064 nm). The choice of this wavelength (as opposed to UV or VIS) is that the molecular scattering is greatly reduced, making the aerosol signature much clearer. The instantaneous magnitude of the return signal provides information on the magnitude of the target (e.g. backscatter properties) while the height of the target is determined by the measured time delay of the pulse echo.

Unlike radar where the intervening atmosphere between the source and the target is negligible, while processing the raw lidar data, we must include the fact that the atmosphere attenuates the signal before and after hitting the target. To account for this, the instantaneous return signal strength is given by the LIDAR equation:

$$P(R,\lambda) = K.G(R).\beta(R).T(R) \tag{2.1}$$

The received power $P(R,\lambda)$ is a function of range (R) and wavelength (λ), which is depend on 4 terms that are listed and are explained in the following paragraphs (see Fig. 2.1).

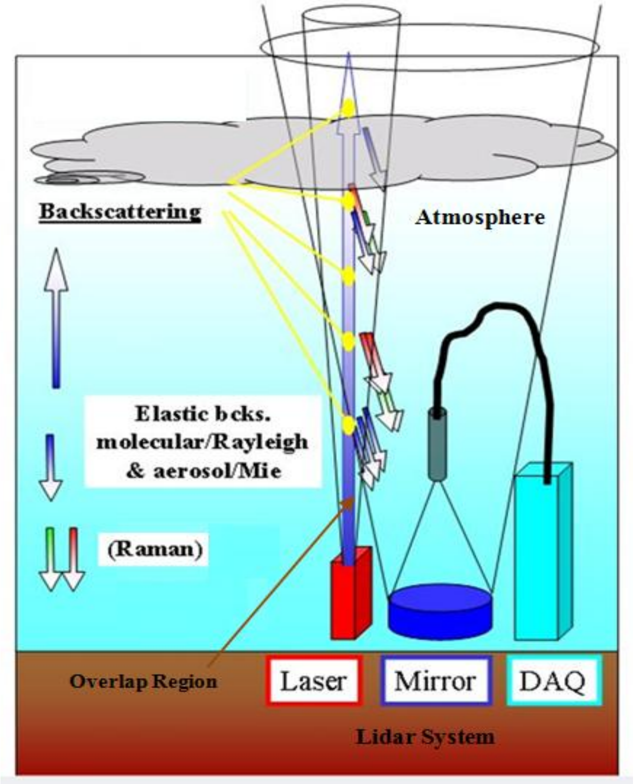


Figure 2.1: Lidar Principles.

***K* - System constant**

$$K = P_0 \frac{c\tau}{2} A\eta \tag{2.2}$$

Given that P_0 is the average power of a single laser pulse and τ the temporal pulse length, the pulse energy is derived by $E_0 = P_0\tau$. The effective pulse length is defined as $\Delta R = (c\tau)/2$ where c

is the speed of light. A is the area of the primary reception optics and η the overall system efficiency (optical transmission from emitter to receiver and detection efficiency). It is an experimentally controlled parameter (system specifications).

G(R) - Range-dependent measurement geometry

$$G(R) = \frac{O(R)}{R^2} \tag{2.3}$$

The geometry factor contains the overlap function of the laser beam with the receiver field of view, called $O(R)$, and an R^{-2} dependence. The latter is obtained from the fact, that the area of reception is part of a sphere's surface with radius R, centered in the scattering volume. It is also an experimentally controlled parameter (optical setup).

β(R) - backscatter coefficient

The backscatter coefficient is an atmospheric parameter which determines the strength of the lidar signal. It describes how much light is scattered in the backward direction (e.g. 180°). Given that N_j is the concentration of scattering particles of kind j and $d\sigma_{j,sca}(\pi,\lambda)/d\Omega$ is the differential backscatter cross section of the particles, the backscatter coefficient results in the sum of all kinds of scatterers:

$$\beta(R,\lambda) = \sum_j N_j(R) \frac{d\sigma_{j,sca}}{d\Omega}(\pi,\lambda) \tag{2.4}$$

In the atmosphere, the laser light is scattered by air molecules and particles. The backscatter coefficient $\beta(R,\lambda)$ can therefore be written as a sum:

$$\beta(R,\lambda) = \beta_{mol}(R,\lambda) + \beta_{aer}(R,\lambda) \tag{2.5}$$

Molecular scattering (index "mol") is primarily dependent on air density and decreases with increasing height. The influence by particles (Index "aer") is highly variable in space and time.

T(R) - Transmission term

That part of light which gets lost on the way from the lidar to the scattering volume is expressed by the transmission term:

$$\boldsymbol{T(R,\lambda) = e^{\left(-2\int_0^R \alpha(r,\lambda)dr\right)}} \tag{2.6}$$

This term is derived from the specific form of the Lambert-Beer-Bouguer law for lidar. The integral describes the path from the lidar to the location of backscattering at *R*. The factor 2 represents the forward and backward travel of the light. The extinction coefficient *α(R,λ)* is computed similarly to the backscatter coefficient as a product of concentration and extinction cross section $\sigma_{j,\,ext}$ for every type of scatterer:

$$\alpha(R,\lambda) = \sum_j N_j(R)\sigma_{j,ext}(\lambda) \tag{2.7}$$

The extinction results from absorption (index "abs") and scattering (Index "sca") of light by molecules and particles:

$$\alpha(R,\lambda) = \alpha_{mol,sca}(R,\lambda) + \alpha_{mol,abs}(R,\lambda) + \alpha_{aer,sca}(R,\lambda) + \alpha_{aer,abs}(R,\lambda)$$

As already mentioned, *β* and *α* depend on the wavelength of the laser light. This relation is determined by the size, refraction index, and shape of the scattering particles.

In summary, the lidar equation from Eq. 2.1 can be re-written as:

$$P(R,\lambda) = \left(P_0 \frac{c\tau}{2} A\eta\right)\left(\frac{O(R)}{R^2}\right)\left(\beta(R,\lambda)\right)\left(e^{\left(-2\int_0^R \alpha(r,\lambda)dr\right)}\right) \tag{2.8}$$

This study is using a ground-based vertically pointing multi-wavelength elastic-Raman scattering lidar that has been operating for aerosol, cloud and water vapor monitoring at CCNY (N40.8^{o}, W74.0^{o}), as part of NOAA-CREST lidar network. The 30-Hz pulsed Nd:YAG laser (Spectra-physics Quanta-Ray 320) system emits at three wavelengths (1064-532-355-nm) with a laser power drift less than 5% for routine daytime measurements (5~8 hours long per day). The laser beam divergence is less than 0.5 mrad (full angle at FWHM points). The complete lidar system descriptions are listed in Table 2.1.

Table 2.1 CCNY Lidar System Specifications

Laser	Q-switched Spectral-Physics Quanta Ray Pro230 Nd:YAG with variable output power up to 475 mJ at 532 nm, 950 mJ at 1064 nm and 300 mJ at 355 nm. Repetition rate of 30 Hz with 8-12 ns pulse duration at 1064 nm, <0.5 mrad divergence.
Telescope	Diameter 20 inch Newtonian Reflector, F3.5
Detectors	APD (silicon enhanced avalanche photodiode) for the 1064 nm (infrared) channel. PMT (Hammamatsu photomultiplier tubes) for the 532 nm (visible), 407 nm (Raman), 386 nm (Raman) and 355 nm (UV) channels.
Digitizing System	Lidar Transient Recorder TR 40-160 (LICEL) with 12-bit, 40 MHz A/D converter for signals between 10 MHz and 200 MHz, 64-level fast discriminator for signals in the high frequency domain above 200 MHz.
Range Resolution	500 m to 15 km
Data Acquisition	Acquisition system is configured by using the TR 40-160 Transient Recorder modules for all channels in a five-channel rack comprising power supplies and interface ports to a PC computer equipped with National Instruments digital I/O card, DIO-32F.

In addition to the better aerosol contrast against a molecular background, the use of the 1064 nm wavelength has other advantages. For instants, the solar signal is weaker at 1064 nm than it is at 532 nm. Also, there are more initial photons present, allowing an increase in signal to noise ratio (SNR), which is primary limited by the amount of photons present and thus making far range detection possible. Unfortunately, at this wavelength, the system is not eye-safe, meaning that its deployment is severely. Since this system is located in NYC area where two major airports are nearby and a lot of helicopters (e.g. police or tourist) are flying at low altitude, human observer is employed during operating period to cut off laser transmission if aircraft activity is spotted inside the scan area. Besides, a radar system which is used to automatically detect targets is employed in the same time, but such system is not foolproof. Thus, both monitoring methods are needed to run legally in the NYC area with FAA approval.

The basic setup of the lidar system in CCNY remote sensing laboratory is shown in Fig. 2.2. Unfortunately, with coaxial transmitter-receiver geometry, the complete geometric overlap begins at approximately 0.5 km altitude making it ill-suited for near surface observations. Initial lidar return profiles are recorded at 1 minute time steps with a range gate resolution of 3.75 meters but are suitably averaged to reduce noise since most model estimates have significantly less spatial and temporal resolution requirements. Furthermore, the high power of the 1064 nm channel makes it easy to see high altitude features (including clouds), with aerosol vertical profiles up to 12 km. This Nd:YAG laser system can transmit the fundamental beam at 1064 nm but suitable non-linear crystals allow for the generation of second (532 nm) and third harmonic (355 nm) beams. The ability to simultaneously observe in the NIR, VIS and UV helps to better identify the class of the aerosols. For example, if the optical backscatter properties decrease dramatically with increased wavelength, the particles are likely to be fine mode dominated. If the decrease is less, a large coarse mode component is present. This information is particularly striking for aloft plumes and is a robust method to separate aloft smoke plumes from dust plumes. However, the speciation potential of the lidar data is not critical to the material of this thesis.

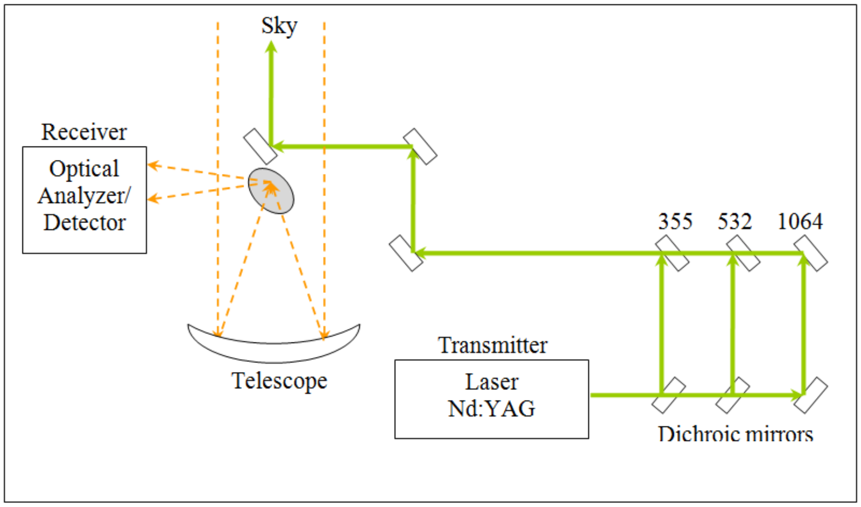


Figure 2.2: A schematic diagram of the lidar in CCNY.

At the receiver end, a telescope collects the photons backscatter from the atmosphere and sends it to the optical analyzer / detector. To reduce the amount of sky light that passes through to the detector thereby generating noise, a small pinhole on the receiver is used. Besides the three fundamental wavelengths, (1064 nm, 532 nm, 355 nm), two additional inelastic (Raman scattering) scattered photons are scattered back to the telescope. These Raman channels include the 407 nm channel which results in inelastic conversion of water vapor molecules, and the 387 nm channel due to inelastic Raman scattering off of nitrogen. The Raman processes are valuable since the ratio of these Raman signals determines the ratio between water vapor and nitrogen. Since nitrogen is well mixed, this is equivalent to retrieving the water vapor mixing ratio. All 5 channels can be selected from the collected light by suitably using interference filters and dichroic mirrors. The selected wavelengths are directed onto different detectors where the received optical signal is converted into electrical signals as shown in Fig. 2.3.

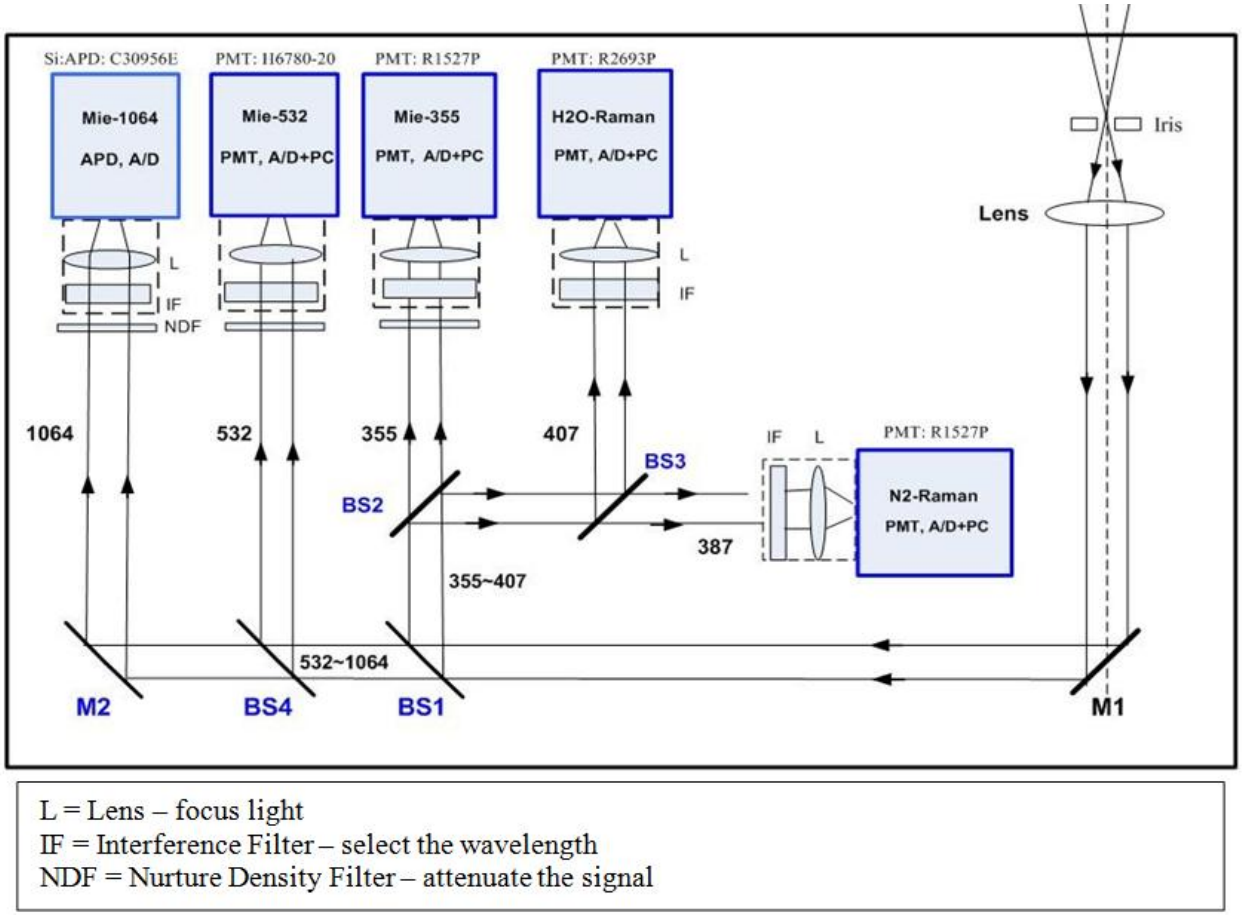


Figure 2.3: A schematic diagram of the optical receiver.

As mentioned earlier, we use the 1064 nm channel measurement to derive the PBL height since this channel is less sensitive to molecular backscatter and attenuation, which means the lidar signal is mostly contributed by aerosols. As we know, the PBL typically contains greater aerosol concentration than the overlaying troposphere and hence has larger backscatter. This is because the aerosols are trapped in the PBL by a potential temperature inversion. Therefore, the backscatter signal strength is dramatically reduced when it leave the PBL into the free troposphere. This sharp change of signal is used to estimate the PBL height. These changes in backscatter signal caused by the strongly varying aerosol concentration provide a useful tool for remotely observing the 2-dimensional (2D) structure of the PBL.

2.1.1 Calibration

Lidar calibration is important for the elastic-scattering lidar system to provide absolute aerosol backscatter profile which is important if we intend to compare them against model outputs (see Chapter 6). For the 355 nm and 532 nm channels, calibration is performed by matching the lidar return signals to a sufficiently high range atmospheric layer which is dominated by molecular (Rayleigh) scattering. However, this method of calibration is ill–suited for 1064 nm channel.

According to Wu et al. (2010), the CCNY elastic lidar (1064 nm wavelength) can be calibrated using different cloud targets because clouds often occur in the lidar observation and their optical properties (e.g. extinction-to-backscatter ratio based on Mie scattering calculations over a diverse range of droplet size distributions) are reasonably stable. The single-scattering elastic lidar equation with both molecular and aerosol components is given as

$$\begin{aligned} P(R) &= C \times \beta(R) \times T^2(R) / R^2 \\ &= C \times [\beta_m(R) + \beta_p(R)] \times \exp[-2\int_{R_o}^{R} (\alpha_m(R') + \alpha_p(R')) dR'] / R^2 \end{aligned} \tag{2.9}$$

where $P(R)$ is the lidar return at altitude R, R_0 is the lidar near end reference altitude; C is the lidar system constant (the quantity to be determined) which is a complex and uncertain function of the receiver and transmitter characteristics; β and α are the atmospheric backscatter and extinction coefficients, where the subscripts 'm' and 'p' denote molecule and particle (aerosol and/or cloud), respectively; and $T^2(R)$ is the total two-way transmittance. In the analysis, molecular backscatter and extinction coefficients can be calculated with the temperature and pressure profiles and therefore, only the aerosol profiles are considered unknown. Since the aerosol attenuation at 1064 nm is always small, (with AOD < 0.05) one can approximately calculate the aerosol backscatter coefficient under the clean sky by neglecting the attenuation term if one knows the lidar system constant. Consequently, the required assumption of lidar ratio and far-end boundary value of aerosol used in the conventional inversions is not needed, but must be replaced by another form of constraint which is defined through the calibration methods to be briefly discussed in the following.

2.1.1.1 Calibration using low-level water clouds

This calculation use the fact that water cloud droplets are usually spherical shape with known refractive indices, and the particle size distribution is taken to be a gamma distribution, which has been shown to provide a realistic representation of the actual droplet size distributions in water clouds (Miles et al., 2000; Pinnick et al., 1983). Assuming that the clouds are indeed water phase clouds with “mature” droplet structure, we can formally integrate the lidar signal in the clouds as below:

$$\frac{1}{C}\int_{R_b}^{R_t}\left[P(R)R^2\right]dz = T^2(R_0,R_b)\int_{R_b}^{R_t}\left[\beta_c(R)\exp(-2\tau_c(R)\eta(R)\right]dR = T^2(R_0,R_b)\left[1-\exp(-2\tau_c\eta(R))\right]/[2S_c\eta(R)] \qquad (2.10)$$

where R_b and R_t are cloud base and top, respectively; $T^2(R_0, R_b)$ is the two-way transmittance of aerosol and η is the cloud multiple scattering factor ($\eta=1$ when no multiple scattering or only single scattering is present) and can be assumed as a constant for optically thick low-level water cloud (Platt, 1981; Spinhirne et al., 1989; Kunkel & Weinman, 1976). It follows that the lidar system constant C can be expressed as:

$$C = (2S_c\eta)\times\int_{R_b}^{R_t}\left[P(R)R^2\right]dR/[1-\exp(-2\tau_c\eta)]/T_p{}^2(R_0,R_b) \qquad (2.11)$$

Based on the above formula, if an estimate of the S ratio, the multiple-scattering parameter (η) and the aerosol transmittance below the cloud can be estimated, and then the system constant can be determined. Fortunately, for low altitude water clouds, the S ratio for mature droplets is very stable with a value approximately 20. This can be seen in Fig. 2.4 where the S ratio for two different cloud droplet gamma distribution variances are plotted as a function of the variable effective radius.

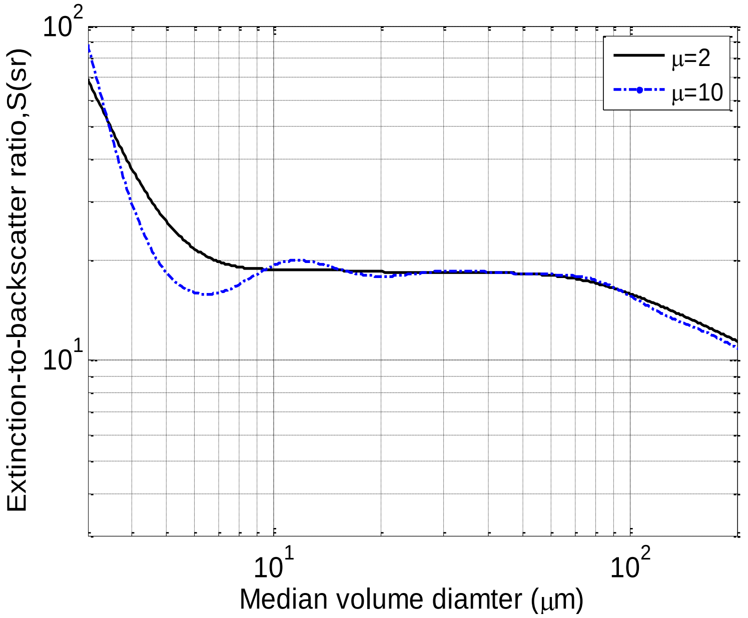


Figure 2.4: Water cloud lidar ratio at 1064 nm as a function of median volume diameter for gamma droplet size distribution with two different values of μ.

For optically thick clouds *($\tau_c \geq 3$)*, the denominator $[1-\exp(-2\tau_c\eta)]$ can be safely ignored. In addition, the atmospheric transmittance below the cloud at 1064 nm is quite small, but these errors can be further reduced by estimating the AOD at 1064 nm under the low cloud using Raman-channel derived aerosol extinction profile (Wu et al., 2009) at 355 nm and a sunphotometer measured Angstrom exponent (Holben et al., 1998) to extrapolate back to 1064 nm. Another source of possible error is the multiple scattering influences. To estimate this effect, for each individual case, Eloranta's multiple scattering model (1998) can be used to calculate the factor η with the lidar derived cloud extinction profile and MODIS derived effective radius for the cloud droplets (Hogan, 2006; Wandinger, 1998; King et al., 1997).

However, there are a few limitations of this method. For instance, this method cannot be used in precipitating clouds since their droplet diameters lay outside the range of low lidar ratio variability (O’Connor et al., 2004). Similarly, this approach is not appropriate for long wavelength CO_2 laser lidars due to changes in the droplet size to lidar ratio (O’Connor et al., 2004; Pinnick et al., 1983; Dubinsky et al., 1985). According to Eq. 2.11, the uncertainties of calibration constant are mainly due to uncertainties in the multiple-scattering factor, the lidar ratio and the aerosol transmittance under the cloud. Since these factors are not correlated, total uncertainty for calibration constant is estimated as:

$$\frac{\Delta C}{C} = \sqrt{(\frac{\Delta S_c}{S_c})^2 + (\frac{\Delta \eta}{\eta})^2 + (\frac{\Delta T_p{}^2}{T_p{}^2})^2} \qquad (2.12)$$

2.1.1.2 Calibration using high cirrus clouds

On the other hand, water clouds are not always present especially in winter, so it is very important to develop another calibration method to provide consistency checks. The main idea of this method is to use cirrus clouds at high altitude as our calibration target. If we already calibrate the 532 nm channel so we have an absolute measure of the backscatter, we can assume the backscatter of the cloud is white, so the backscatter at 1064 nm is equal to the backscatter at 532 nm. With this value, we can then get the 1064 nm system calibration constant.

As summarized above, the first step is to obtain a calibration coefficient for the 532 nm channel based on the normal calibration against a molecular target together with the total column sunphotometer measured AOD (Welton et al., 2002; Voss et al., 2001; Currie et al., 2004). Here, the lidar system constant is re-written from Eq. 2.9 as:

$$C = \frac{P(R)R^2}{S(R)\beta_m(R)T_m(R)^2 T_a(R)^2} \qquad (2.13)$$

where, $S(R)$ is the aerosol scattering ratio defined as $S(R)= [\beta_a(R)+\beta_m(R)]/\beta_m(R)$. If we assume $S(R_{ref})=1.01$ at 532 nm in the upper troposphere (10~13 km) and estimate aerosol transmittance $T^2_a(R_{ref})=exp(-2*\tau)$ with the sunphotometer measured optical depth τ, then the lidar constant C_{532} can be estimated from Eq. 2.13. The AOD at the lidar wavelength is calculated according to the Angstrom power law from the sunphotometer multiple-wavelength AOD. Error for the 532nm lidar calibration constant is found to be approximately ±5% since C_{532} is not very sensitive to the assumed value of the scattering ratio $S(R_{ref})$.

The next step is to estimate lidar constant at 1064 nm by assuming the wavelength independence of cirrus cloud backscatter and extinction. The ratio of lidar signals at 532 nm and 1064 nm in the cirrus cloud under suitable approximations can be written as:

$$C_{1064} \approx \frac{P(1064, R_c)}{P(532, R_c)} \times C_{532} / K_m(R_0, R_c) / K_a(R_0, R_b) \tag{2.14}$$

where R_c is the cirrus cloud altitude; K_m and K_a are the two-wavelength transmittance ratios for molecules and aerosols under the cloud, respectively and K_c is cloud transmittance ratio.

Here, the molecular transmittance ratio K_m can be precisely calculated from the radiosonde measured temperature and pressure profiles while the ratio K_a of aerosol transmittances is estimated from the sunphotometer measured AOD in an adjacent clean-sky patch. In this case, the multiple-scattering influence can be safely neglected because the ratio of cirrus transmittances at 532-1064 nm is used and the transmittances are independence of wavelengths. To ensure the clouds are sufficiently scattering, we only use strongly scattering returns with the attenuated scattering ratios S'_{532} greater than 10, which determines cloud segments with sufficient signal intensity and minimizes the influence of non-cloud return (Yang & Liou, 1996; Tao et al., 2008). The main errors in the cirrus cloud method are from the uncorrelated uncertainties of C_{532}, Ka, and the individual lidar cirrus returns so the total uncertainty can be written as:

$$\frac{\Delta C_{1064}}{C_{1064}} = \sqrt{(\frac{\Delta C_{532}}{C_{532}})^2 + (\frac{\Delta K_a}{K_a})^2 + (\frac{\Delta P(1064)}{P(1064)})^2 + (\frac{\Delta P(532)}{P(532)})^2} \tag{2.15}$$

To compare both methods, we need cases where both water clouds and high altitude clouds are both present. One example is given in Fig. 2.5 (a-d). The calibration constants obtained from both methods are close to each other (≈1.4x10^4) in this case (see Fig 2.5 (b and d)). To ensure the precision of lidar constant at 532 nm, the AOD results derived from lidar and sunphotometer for the clear sky are compared in Fig 2.5 (c) and it shows a good agreement.

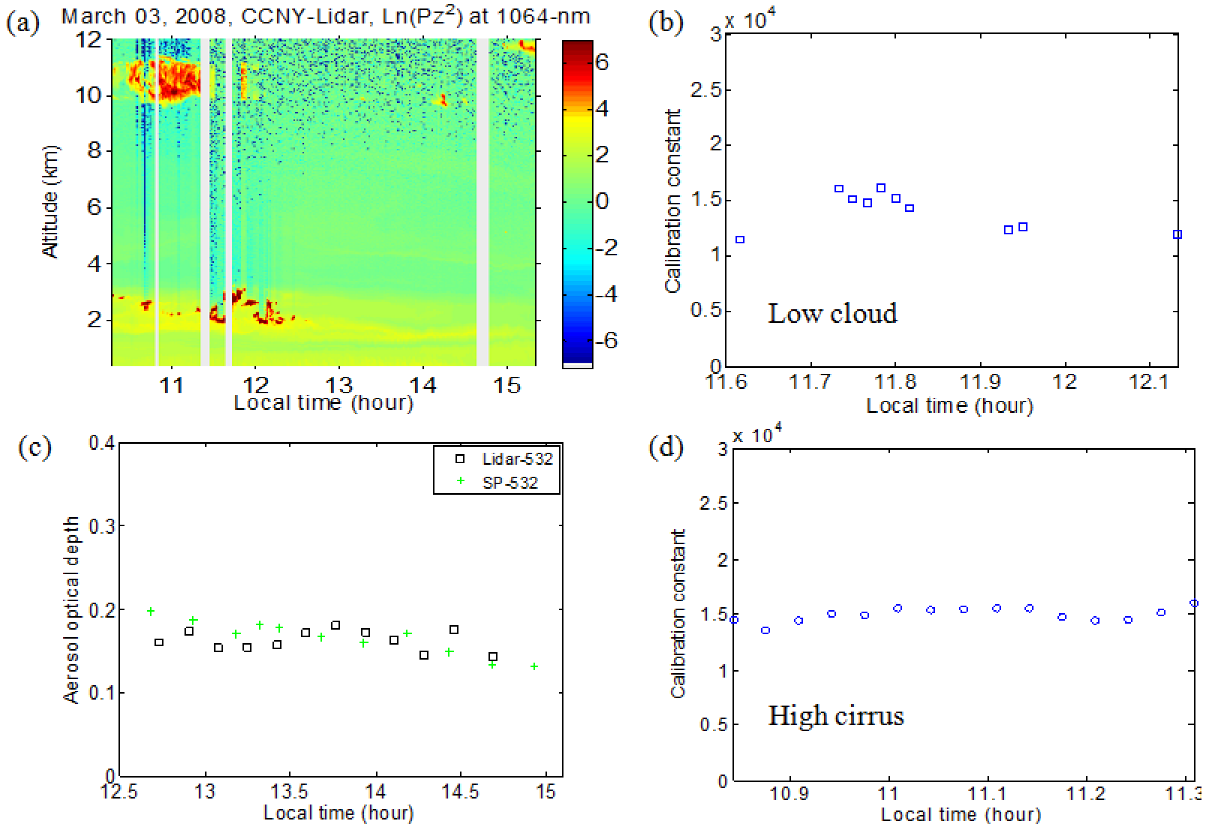


Figure 2.5: (a) Time-height cross section of the range-corrected lidar returns, (b) Calibration constant from the low-level water cloud, (c) Comparison of lidar- and sunphotometer (SP)-AOD, and (d) Calibration constant from the high-cirrus cloud method on March 3, 2008.

Second example is shown in Fig. 2.6 (a-c), the calibration results (≈1x10^4) from both methods are in good agreements again. The ratio of the 1064 nm backscatter when compared to the 532 nm value is very reasonable. In particular, we find that between the two methods, calibrations < 10% for the 1064 nm channel can be made as long as the necessary ancillary data from sunphotometer are available.

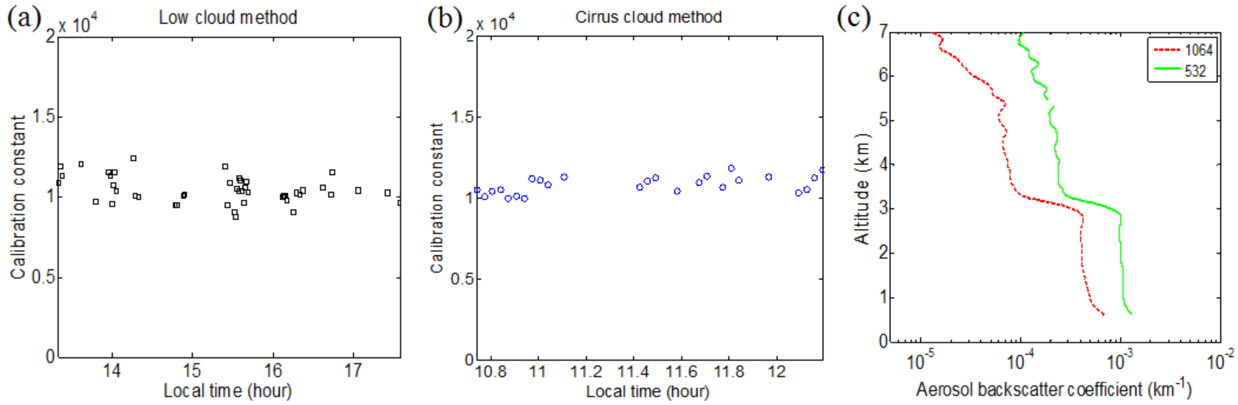


Figure 2.6: Lidar calibration constants with (a) the low-altitude water clouds, (b) the high cirrus clouds, and (c) aerosol backscatter coefficients at 1064-532-nm on June 16, 2006.

2.2 Ceilometer

As we mentioned earlier, the lidar is not suitable for low altitude measurements or 24 hours / 7 days operation. To address this, we use a high repetition rate low power lidar commonly called a ceilometer (model CL31) which is manufactured by Vaisala Inc. Its single lens optics that uses the inner part of the lens for transmitting and its outer part for receiving light provides sufficient overlap of the transmitter and the receiver FOV over the whole measuring range (see Fig. 2.7). This results in far superior near-range performance compared to two lens systems and allows reliable detection of very low aerosol structures below 200 meters, that not seen by other instruments. Besides, this instrument is good for mobile operation because it is small and light weight. The main performance characteristic of ceilometer is listed in Table 2.2.

The ceilometer is a device that is originally designed to determine the height of cloud base. Due to the higher backscatter signal of cloud, the laser pulse power which is needed is dramatically less and therefore it can be made eyesafe. This instrument is located at the same location as CCNY lidar system. Furthermore, unlike the lidar transmitter which has a particularly unsafe visible channel at 532 nm, the ceilometer utilizes a single NIR (Near Infra-Red) channel at 910nm. While designed for cloud detection, it is reasonable to believe that ceilometers can also be used to measure the aerosol concentration within the atmosphere (Eresmaa et al., 2006; Haji et al., 2009; Münkel & Roininen, 2008, 2010). Signal return profiles are recorded at 15 seconds averaged with a range resolution of 10 meters. However, it is also clear that the low signal pulse power will result in higher noise and significant sliding temporal and spatial averaging would need to be applied to bring the noise contamination down to acceptable levels (Münkel & Roininen, 2010). Still, noise limitation of this instrument makes it impractical to observe PBL height that is above 1.5 km, where strong sky noise is dominated at higher altitude.

Figure 2.7: Vaisala ceilometer model CL31

Table 2.2 Main performance characteristics of Vaisala ceilometer CL31

Minimum range resolution	5 m
Typical range resolution for boundary layer scans	10 m
Minimum report interval	2 s
Typical report interval for boundary layer scans	16 s
Measuring range for cloud base detection	0 … 7500 m
Backscatter profile range	0 … 7500 m
Range for boundary layer fine structure profiling	0 … 3000 m
Eye-safety class	1 M
Total height	1190 mm
Total weight	31 kg

2.3 Sunphotometer

In order to get better calibrations as well as test the CMAQ output data directly, column integrated multi-wavelength optical depth measurements are very useful. A CIMEL Electronique 318A spectral radiometer manufactured in Paris, France which is a solar-powered weather hardy robotically pointed sun and sky spectral radiometer is used in this study. This instrument is located at CCNY (same location as lidar and ceilometer), as part of the AERONET (**AE**rosol **RO**botic **NET**work). Details of this instrument could be found in reference (Holben et al., 1998).

Two basic measurements are made by the radiometer. This includes direct sun which result in multispectral column integrated optical depths (1640 nm, 1020 nm, 532 nm, 500 nm, 440 nm, 380 nm and 340 nm). These data can be obtained from AERONET website (http://aeronet.gsfc.nasa.gov/) and an example product is shown in Fig. 2.8. In general, the system provides nominal cloud clearing but does not provide quality assurance (level 1.5). Note that the 1640 nm channel is not standard product but was acquired by us to improve the separation aerosols into fine and coarse modes. More details of this application where MIR is also considered are presented in Chapter 7. Besides the direct sun measurements, full scanning sky measurements are performed at 440 nm, 670 nm, 870 nm and 1020 nm. These measurements provide multi-angle diffuse transmission measurements which can be used to extract the particle size distribution (bimodal) and the complex refractive index. However, these full sky measurements are much harder because any cloud passing will destroy the measurements. The instrument deployment is on the roof of CCNY Engineer School building (same location as ceilometer) and is shown in Fig 2.9.

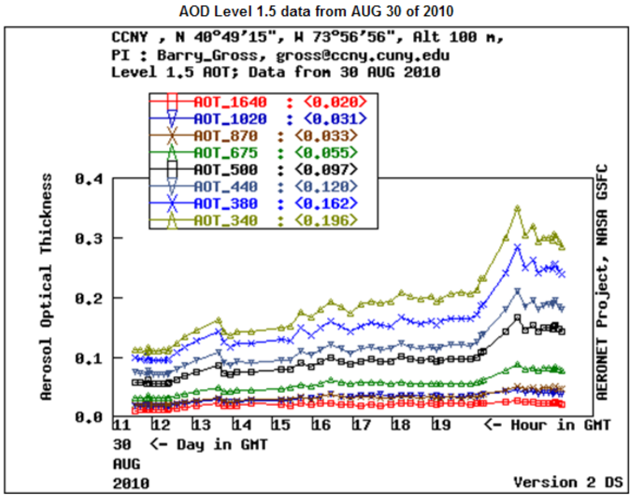


Figure 2.8: AOD Level 2 data (1040 nm, 1020 nm, 870 nm, 675 nm, 500 nm, 440 nm, 380 nm and 340 nm) from CCNY sunphotometer measurements.

Figure 2.9: Sunphotometer (CIMEL Electronique 318A spectral radiometer) located at CCNY

2.4 Microwave Radiometer (MWR)

In trying to assess WRF meteorological fields such as temperature and water vapor profiles, we have made use of a microwave radiometer (model MP-3000A) which is manufactured by Radiometrics (see Fig. 2.10). The MWR is a passive sensing device which looks at the sky brightness with 50 microwave channels from 22 to 182.3 GHz. The specification of this instrument is given in Table 2.3. The multispectral observations in the microwave region are capable of producing high resolution temperature (Kelvin), relative humidity (%) and water vapor profiles (g/m^3) and low resolution liquid profiles (g/m^3). They are measured every 2 minutes with a nonlinear spatial grid (0-0.5 km with a 0.05 km interval; 0.6-2 km with a 0.1 km interval; 2.25-10 km with a 0.25 km interval). When doing the WRF-MWR comparisons, the data is interpolated for 0-4 km at 100 meter resolution and is averaged hourly.

Figure 2.10: MP-3000A Hyper-Spectral Temperature, Humidity and Liquid Water Profiler.

Table 2.3 MP-series (Advanced) Specifications

Calibrated Brightness Temperature Accuracy[1]	0.2 + 0.002*|TkBB-Tsky|[2]
Long Term Stability	< 1.0 K/year typical
Resolution (depends on integration time)[3]	0.1 to 1 K
Brightness Temperature Range[4]	0-400 K
Antenna System Optical Resolution and Side Lobes	
• 22-30 GHz	4.9 – 6.3° -24 dB
• 51-59 GHz	2.4 – 2.5° -27 dB
Integration Time (user selectable in 10 msec increments)	0.01 to 2.5 seconds
Frequency Agile Tuning Range	
• Water Vapor Bands	22-30 GHz 179-183.3 GHz
• Oxygen Band	51-59 GHz
• Minimum Frequency step size	2.0 MHz @ 30 GHz 4.0 MHz @ 60 GHz 12.0 MHz @ 180 GHz
Standard Factory-Calibrated Channels	
• 22-30 GHz Band	21 Channels
• 51-59 GHz Band	14 channels
• 170-183.3 GHz Band	15 channels
Pre-detection channel bandwidth (effective double-sided)	
• 22-30 and 51-59 GHz Bands	300 MHz
• 170-183.3 GHz Band	1000 MHz
Surface Sensor Accuracy	
• Temperature (-50° to +50° C)	0.5° C @ 25° C
• Relative Humidity (0-100%)	2 %
• Barometric Pressure (800 to 1060 mb)	0.3 mb[5]

1 Specified accuracy for instrument calibrated with an external target with no error.

2 Absolute accuracy is best for sky brightness temperatures close to ambient, such as for the highest V and channels, and degrades as the absolute difference between the black body reference and sky temperatures increases.

3 Typical resolution for 250 msec integration time is 0.25 K for MP-3000A, 2500A, 1500A, and 1 K for MP-183A.

4 Wider ranges are available. 0-400K is optimum for meteorological applications.

5 Optional extended range barometer: 600 to 1060 mb, +/- 2 mb.

CHAPTER 3

Data Analysis and Methodology

In this chapter, lidar based PBL height retrieval methods (maximum gradient, maximum variance and WCT) are presented. Based on our comparisons, WCT method, one of the best candidates, is modified so that it can better determine PBL height especially in complex atmospheres (e.g. multiple layers occurs in the same time).

3.1 Methodologies for Planetary Boundary Layer (PBL) Height Determination

Several PBL height detection techniques using air-borne or in situ lidar system have been suggested in the past years such as curve fitting, threshold, maximum gradient, maximum variance and wavelet methods (Emeis et al., 2008). For instance, the curve fitting method tries to fit the backscatter profile to an assumed shape (Steyn et al., 1999). However the main problem of this method is to find a functional form in which the lowest altitudes will have a high backscatter with a sharp transition to lower levels of backscattering in the layers above. The functional form must be robust enough to accommodate the many variations in shape that may be found. Eresmaa et al. (2006) modified this method by utilizing the whole backscatter profile rather than just the portion surrounding the top of the PBL. In the threshold method, the PBL height is determined as the highest data point where the backscatter intensity was some fraction higher than the average backscatter value in the free atmosphere above (Melfi et al., 1985; Boers et al., 1987). However, due to the natural variability of the atmosphere, defining an appropriate threshold value is difficult. An inappropriate value of the threshold will bias the result. Moreover, it is difficult to identify the PBL and the residual layer if they have similar backscatter value.

However, this method is useful in screening for cloud because the cloud backscatter is much higher and consistent.

For our applications where we need to separate between convective and residual layers, and be fairly robust, the above methods are not serious contenders. In the next section, we focus on a modified wavelet covariance transform (WCT) method we develop and compare its performance against more standard methods such as maximum gradient and maximum variance methods. In particular, we find that the improved WCT technique is capable of determining the PBL height under a wide variety of conditions with minimal adjustment to suitable algorithm parameters. Moreover, we find our approach is consistent with those obtained from radiosonde measurement (Culf, 1989; Parlange & Brutsaert, 1990; Hennemuth & Lammert, 2006).

3.1.1 Maximum Variance Method

The top of the PBL is called as entrainment zone because the entrainment of clear air masses from the free troposphere into the PBL occurs here. The entrainment process is variable in time and leads locally to considerable fluctuations in the aerosol concentration. Therefore the maximum in the vertical profile of the variance of the backscatter intensity can be an indicator for an entrainment layer on top of the PBL (Hooper & Eloranta, 1986; Piironen & Eloranta, 1995; Lammert & Bosenberg, 2006; Martucci et al, 2007). This method is called as variance centriod method by Menut et al. (1999). The detail of this method is described as follow.

The variance of the backscatter return signal is defined as:

$$Var(R)_i = \frac{1}{N}\sum_{k=i+1-N/2}^{k=i+N/2}\left[B_k(R) - \overline{B(R)_i}\right]^2 \tag{3.1}$$

where N is the number of profiles and the length of k index is equal to N. For example, if we have a period of 60 min, the external i index goes from 5 to 55 and N=10. The internal k index of summation represents the individual measurements $k = i\text{-}4$, ….. , i+5 in the N interval. The mean value of $\overline{B(R)}_i$ is obtained by averaging the single B_k profiles over N. The variance indicates how much the backscatter signal at a fixed R varies in respect to $\overline{B(R)}_i$ during the N interval. Eq. (3.1) applies to all i profiles; passing from profile i to profile i+1, the N interval shift by 1 minute and makes reference to $\overline{B(R)}_{i+1}$. During daytime convective condition, the entrainment process on top of the mixing layer cause a variability in the backscatter signal, so the Var_i maxima are expected to match the PBL top. The final *Var* profile is obtained as the arithmetic average of the single Var_i profiles.

3.1.2 Maximum Gradient Method

This method calculated the gradient of the signal with height and used the maximum gradient as an indicator of the PBL top. In our convention where height (R) in Eq. 3.2 increase from the surface, the PBL top is defined at which the first derivative has the largest negative value (Kaimal et al., 1982; Hayden et al, 1997; Flamant et al., 1997).

$$BL_{GM} = \min\left(\frac{\partial B(R)}{\partial R}\right) \tag{3.2}$$

where $B(R)$ is the backscatter intensity and R is the height.

One may also look for the minimum of the second-order derivative of the backscatter intensity (also known as inflection point method). This method usually gives slightly lower values for PBL top than the PBL top value determined by Eq. (3.2). In general, maximum derivative or inflection point methods have the advantage of being independent of any arbitrary threshold values and show good accuracy when turbulent fluctuations are present. However, running derivatives are difficult to calculate in noisy dataset, it may result in false maxima. Therefore, spatial and / or temporal averaging is required to increase the SNR. This averaging may dramatically reduce the resolution of the measurements and may also bias the result. Moreover, aerosol layers above or below the PBL often have strong backscatter which may have more significant change than those of the PBL. The change in backscatter with height is larger at the edges of these layers which will cause false PBL top identification.

3.1.3 Wavelet Covariance Transform (WCT) Method

From extended studies of the performance of the available lidar methods applied to a large set of lidar observations, we found the WCT method to be a robust technique for semi-automated PBL top detection (Davis et al., 2000; Cohn & Angevine 2000; Brooks, 2003; Baar et al., 2008). Based on the similarity between the sharp, coherent changes in backscatter of each profile and a step function, also known as Haar function, the location of the PBL top can be found by using the convolution of both of them. The Haar function is defined as:

$$h\left(\frac{R-b}{a}\right) = \begin{matrix} 1: & b-\frac{a}{2} \le R \le b \\ -1: & b \le R \le b+\frac{a}{2} \\ 0: & elsewhere \end{matrix} \tag{3.3}$$

where *R* represents distance in the vertical in this application, and *a* and *b* describe the dilation and translation of the function, respectively. Fig. 3.1 gives its plot.

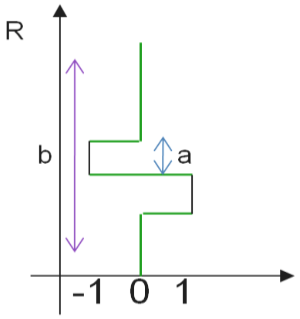


Figure 3.1: Haar function

The covariance transform defined by Gamage and Hagelberg (1993) is the convolution or localized transform, $W_f(a,b)$, of the Haar function with the lidar backscatter profile.

The wavelet covariance transform, $W_f(a,b)$, is defined as

$$W_f(a,b) = a^{-1}\int_{Rb}^{Rt} B(R)h\left(\frac{R-b}{a}\right)dR \tag{3.4}$$

where *Rb* and *Rt* are the bottom and top altitudes in the lidar backscatter profile, and *B(R)* is the lidar backscatter as a function of altitude, *R*. The locations of these maxima and minima, b_i^{max} and b_i^{min} , and the associated values of the covariance transform, $W_f(a_{max}, b_i^{max})$ and $W_f(a_{max}, b_i^{min})$, are the locations and relative strength of step like boundaries in the lidar backscatter profile, *B(R)*. The index *i* refers to the case of multiple local minima or maxima. For clear conditions, $W_f(a,b)$ takes on a clear local maximum at the PBL height because of the high backscatter values in the mixing layer and significantly lower backscatter values in the free atmosphere (Fig. 3.2).

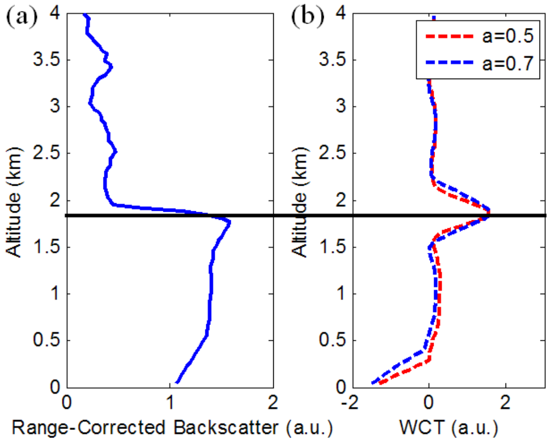
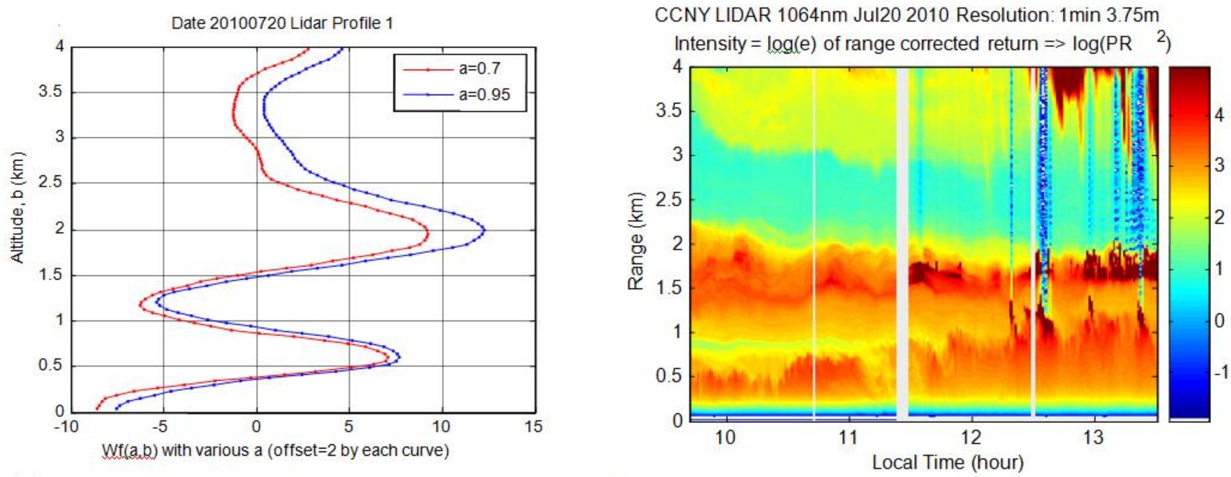


Figure 3.2: (a) Single lidar signal (log scale) profile, (b) Resulting WCT.

However, when strong residual layer or plume entrainment occurs, $W_f(a,b)$ will have multiple local minima and maxima as shown in Fig. 3.3 (a). These multiple maxima confuse the algorithm making it difficult to extract a unique PBL height. For example, Fig. 3.3 (b) illustrates four hours lidar measurements with an aloft plume occurring in the morning. Both layers have significant backscatter. In such cases, the actual PBL height may be biased and distorted so this ambiguity needs to be addressed.



(a)

(b)

Figure 3.3: (a) Single lidar WCT profile, (b) Lidar measurements on Jul 10, 2010.

3.1.4 Modifications applied to Wavelet Method

Our modified WCT is essentially a semi-automated method consists of several procedures to determine the PBL height: a) low-level cloud screening, b) choosing an optimal height-dependent dilation and translation parameter, and c) continuity.

As discussed before, the PBL is growing and decaying indirectly with the change of the solar radiation, which result complex structures of the PBL. To account for the different structures expected in the PBL height, we separate the lidar profiles into two classes. The “transition” class occurs between 8:00-12:00 EST and 18:00-22:00 EST when the PBL is growing / decaying and residual layers maybe present, while the “mature” PBL class extends from 12:00-18:00 EST.

In order to decrease the noise and to smooth the atmosphere transition, a sliding spatial and temporal average is applied to the lidar backscatter dataset. The averaging temporal interval Δt is 20 minutes with a sliding window of 10 minutes, while the averaging spatial interval Δz is 80 meters with a sliding window of 40 meters, is sufficient in comparison to temporal (1 hour) and vertical spatial (≈300 meters at range 1–2 km) resolutions of WRF dataset.

a) Cloud screening to the dataset before the WCT is applied:

Clouds are characterized by a steep increase of the range-corrected lidar signal at the cloud base, followed by a strong decrease of the signal with increasing cloud penetration depth. From these properties, the use of the WCT together with a minimum backscatter threshold is sufficient to easily identify cloud pixels in the lidar profile. For such marked cases, no PBL height is provided.

b) Choosing an optimal height-dependent dilation and translation parameter:

An optimal spatial resolution a is important for a successful retrieval of the PBL height. A very small value of a is sensitive to the noise, in other words, a may cause the $W_f(a,b)$ accidently defined the noise as the top of PBL from a noise dominates vertical profile. On the other hand, a large value of a may fail to resolve the PBL height when further aerosol layers are present in the lower free troposphere. In our analysis, we found that a reasonable smaller a (0.5 km) is needed during the transition regime because of the faster change of the PBL height (e.g. growing / decaying process). While a larger a (0.7 km) is optimum for the afternoon time dataset (see Table 3.1) where the PBL height change is slower. The larger value of a allows us to balance the effects of lower SNR due to the higher altitude where the signal strength is degraded.

In addition, to isolate the convective growth layer from the residual layer, we make use of general trends (e.g. climatology) in the PBL "searching range" to limit the PBL height retrieval. Here, "searching range" is represented by *b* in $W_f(a,b)$ function. Based on statistics, we found that the optimal PBL searching range was constrained below 1 km in the morning and night period while a range from 1 to 3.5 km was used for the afternoon period. The height restriction in morning is necessary since the residual layer is often very deep from the previous day, mainly in the summer.

Table 3.1 Numerical Definitions of Wavelet Parameters and Periods when used in summer.

Period / **Parameter**	**Morning (before 12:00)**	**Afternoon (12:00-18:00)**	**Night (after 18:00)**
a (width of function)	0.5 km	0.7 km	0.5 km
b (transition of function)	0.5-1 km	0.8-3.5 km	0.5-1 km

c) Continuity:

To insure the PBL continuity, a given interior point is compared to its adjacent points. If the discontinuity in the heights of these points is too large (e.g. ± 250 meters in this study with spatial resolution of 40 meters and temporal resolution of 10 minutes) on both sides, it is discarded. If only one of the external values is outside the range, a mean value of both points will be calculated. This scheme is iterated until no further changes occur. This reduces the number of false hits caused by noise or other layers (residual or plume) as well as internal turbulence.

3.2 Performance Comparisons and Validation

First, we want to assess our modified WCT performance with the radiosonde measurements. Fig. 3.4 (a-b) shows two days comparison of the PBL height determined by CCNY lidar (wavelet method), WRF model (simulation) and radiosonde (maximum gradient method). The PBL height determined by radiosonde is about ±250 meters of the PBL height determined by CCNY lidar for both days. The difference of the height might be caused by the different location since CCNY is in the uptown Manhattan while the radiosonde station is about 60 km away from CCNY. Another cause of this bias could be the coarse range resolution of radiosonde data. However, the result from the sounding shows that the lidar estimated PBL height is consistent. On the other hand, lidar derived PBL height is slightly suppressed from the WRF derived PBL height. More investigation of the lidar and model derived PBL measurements are presented in the following section. In general, they have good agreement with each other.

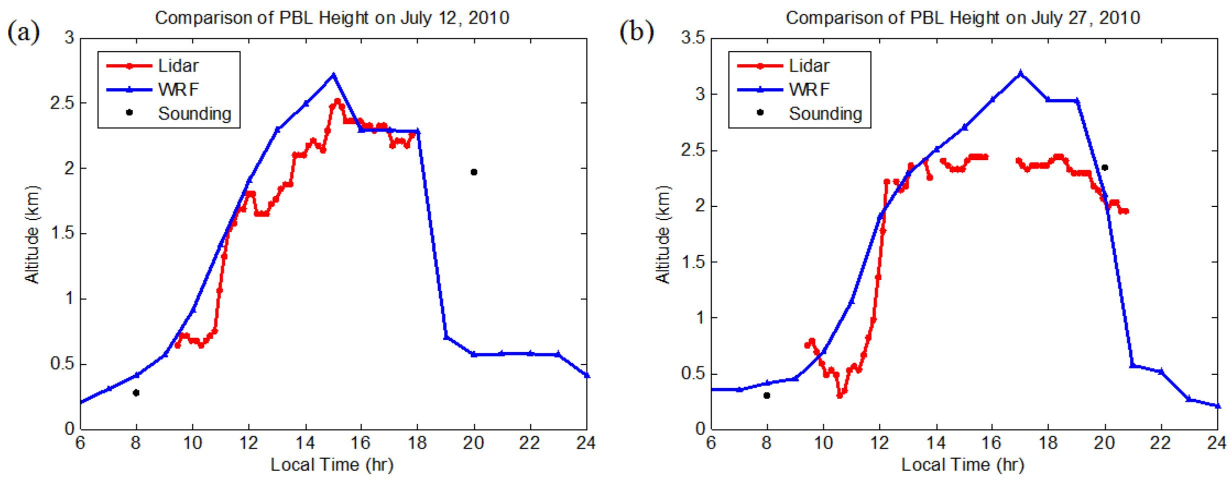


Figure 3.4: (a) July 12, 2010 (b) July 27, 2010. PBL height determined by Lidar (wavelet method), WRF model (simulation) and radiosonde (maximum gradient method).

To understand the superior performance of the modified WCT, we compare our modified WCT method with both maximum gradient and maximum variance methods for two different cases in Fig. 3.5 (a-b). Fig. 3.5 (a) shows a case where the aerosols trapped inside the mixing layer has much stronger backscatter than the aerosol backscatter from the residual layer.

Normally, most of the lidar derived PBL height methods work well when the layer is well mixed and residual aerosol layers are relatively weak. In case (a), the maximum gradient method works well since it is not followed by the weak residual layer. However, the strong inhomogeneous structure (e.g. uneven distribution of aerosols) within the boundary layer tends to result in variance fluctuations that can modify the maximum variance approach. Fig. 3.5 (b) illustrates second case where the aerosols backscatter magnitude is very similar in both the PBL and residual layer. While both standard methods (maximum gradient and maximum variance) are perturbed by the residual layer, our modified WCT is able to determine the PBL height during the initial growth.

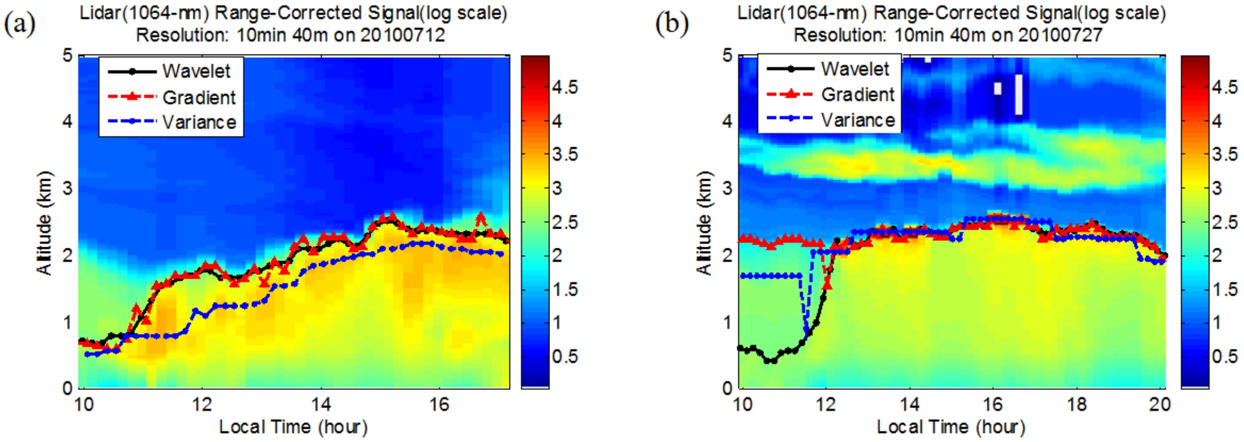


Figure 3.5: Lidar (Range-Corrected) measurements on (a) July 12, 2010 and (b) July 27, 2010.

We further illustrate the performance of the algorithm with particular emphasis on how we can separate the residual and convective layers using height restrictions on the morning mixing height in Fig. 3.6. The black trace represents the WCT estimated PBL height with the height restrictions applied while the blue trace is the WCT estimated PBL height without the height restriction applied. As shown in Fig. 3.6, without height restriction, the WCT picked the residual layer boundary as the PBL height instead.

In order to better quantify the improvement of modified WCT, we summarize the performance (e.g. correlation coefficient, mean and standard deviation) of original WCT; modified WCT and maximum variance methods with WRF model PBL data in Table 3.2. Obviously, the modified WCT method has the highest correlation coefficients (R≈0.85) among three different PBL schemes while maximum variance method has the lowest correlation coefficients (R≈0.4). The only difference between original WCT and modified WCT is the application of height restriction. Moreover, our analysis includes simple and complex PBL cases so this is one of the reasons, variance method has very low correlation coefficients.

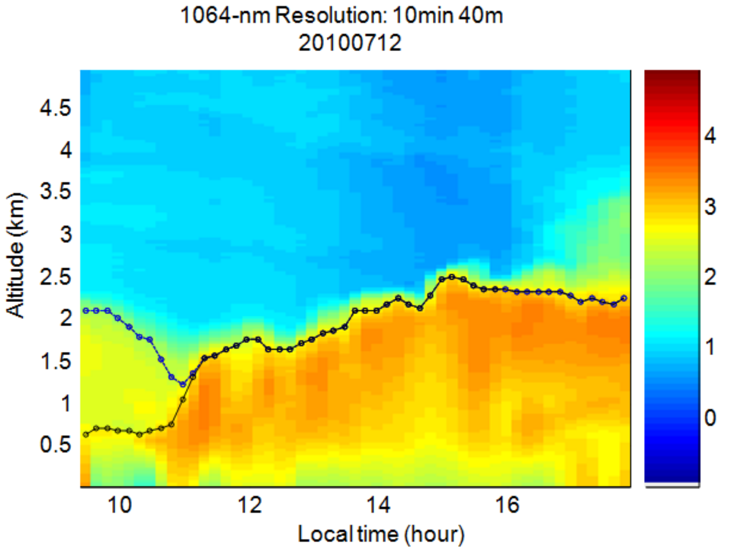


Figure 3.6: Both black dot and blue dot line are representative PBL height retrieval using WCT overlayed on lidar backscatter signal. Blue trace does not use any restrictions while black trace uses morning PBL height restrictions to separate residual and convective layers.

Table 3.2 Correlation Coefficients (R) and height differences between WRF model and all Lidar PBL retrieval methods (mean and standard deviation in km).

WRF schemes **Lidar PBL Methods**	**Blackadar (2007)**			**ACM2 (2007)**			**MYJ (2010)**		
	R	**Mean**	**Std**	**R**	**Mean**	**Std**	**R**	**Mean**	**Std**
Wavelet (with height limit)	0.87	0.19	0.30	0.84	0.34	0.28	0.84	0.23	0.40
Wavelet (without height limit)	0.72	0.10	0.46	0.71	0.25	0.40	0.60	0.06	0.61
Variance	0.32	0.12	0.71	0.39	0.30	0.63	0.48	0.34	0.62

Similarly, the WCT algorithm can be used for ceilometer to derive PBL height without changing ‘*b*’ range since its observation range is 300-1100 meters. We evaluate the performance of WCT in the ceilometer measurements by comparing with lidar WCT derived PBL height. Two days measurements shown in Fig. 3.7 (a-d) are used for this purpose. On Sep 6, we found that the PBL height in ceilometer observation was about 1 km with a deviation of ±100 meters during 11:00-16:00 EST while lidar also observed a similar PBL pattern in the same time (see Fig. 3.7 (a-b)). For the second case on Sep 27, a PBL height at 0.6 km with a deviation of ±100 meters during 10:00-14:00 EST was also found in both observations. Therefore, the ceilometer PBL height retrieval results agree well with lidar derived PBL height. However, note that the return backscatters of ceilometer need to be strong enough for accurate PBL observation which is not often the case. In general, the ceilometer is not suitable for detecting the PBL that is higher than 1.5 km and will not be used for this purpose.

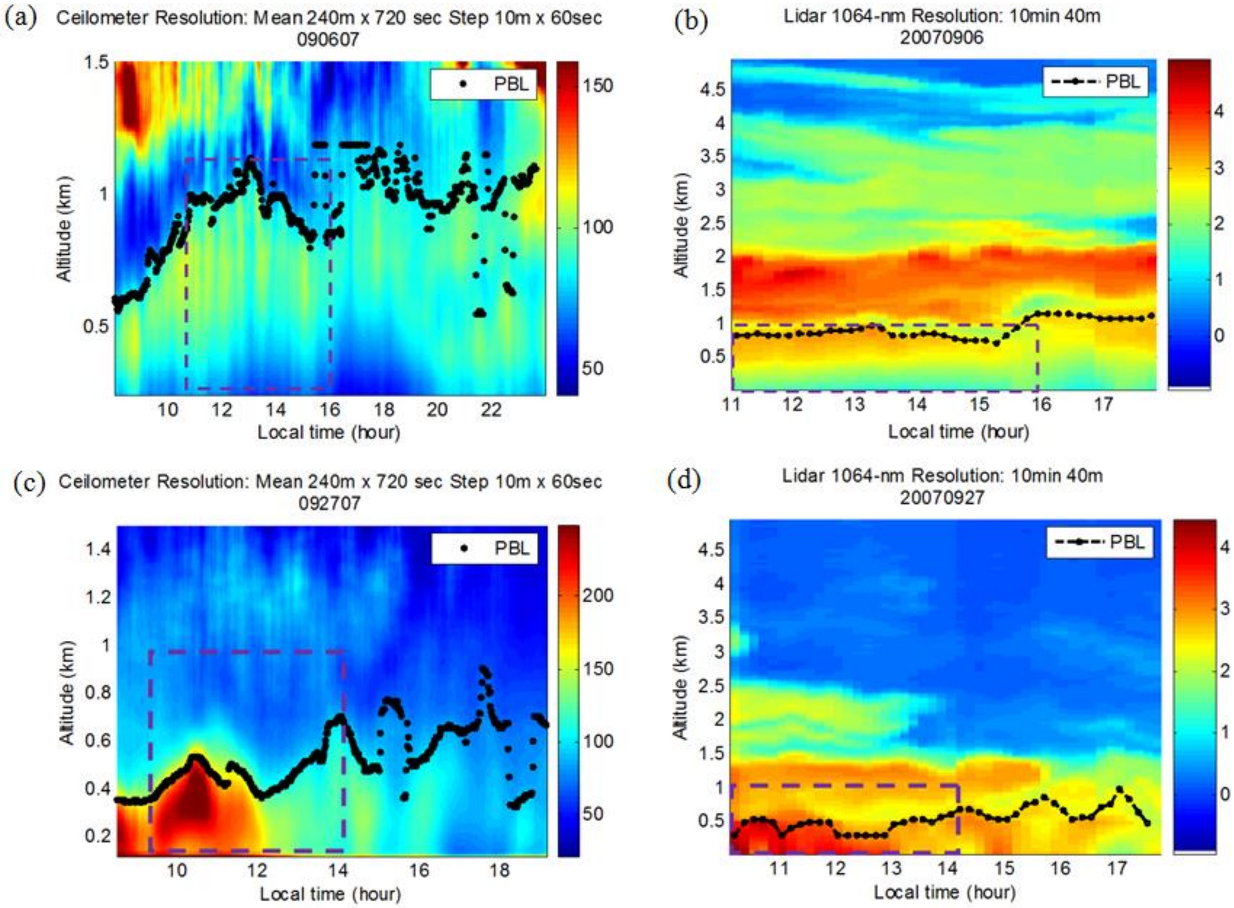


Figure 3.7: PBL height derived from WCT method on ceilometer observations (a) Sep 6, 2007, (c) Sep 27 ,2007 and lidar observations on (b) Sep 6, 2007, (d) Sep 27 ,2007.

CHAPTER 4

Air Quality Model

The fundamentals of the WRF and CMAQ models are introduced followed by the descriptions of the parameterizations and settings that are applied in this thesis.

4.1 Introduction of Air Quality Model

4.1.1 Weather Forecast and Research (WRF) Model

In general, the WRF model is a mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. It can be used for a broad spectrum of applications across different scales ranging from meters to thousands of kilometers. It features multiple dynamic cores, a 3-dimensional variational (3DVAR) data assimilation system and a software architecture allowing for computational parallelism and system extensibility. The effort to develop WRF has been a collaborative partnership, principally among the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (the National Centers for Environmental Prediction (NCEP) and the Forecast Systems Laboratory (FSL), the Air Force Weather Agency (AFWA), the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA).

The WRF model is able to conduct simulations reflecting either real data or idealized configurations, and is flexible and computationally efficient. In addition to the operational forecasts, WRF also is evolving integrating advances in physics, numerical approaches, and data assimilation techniques contributed by the research community. Fig 4.1 illustrates the flowchart for the WRF Modeling System which consists of these major programs:-

1. The WRF Preprocessing System (WPS)
    - Real-data interpolation for NWP runs.
    - New obsgrid program for adding more obs to analysis.
2. WRF Model (ARW and NMM dynamical cores)
    - Initialization programs for real and (for ARW) idealized data.
    - Numerical integration program.
3. Post-processing
    - Graphics and verification tools including MET.

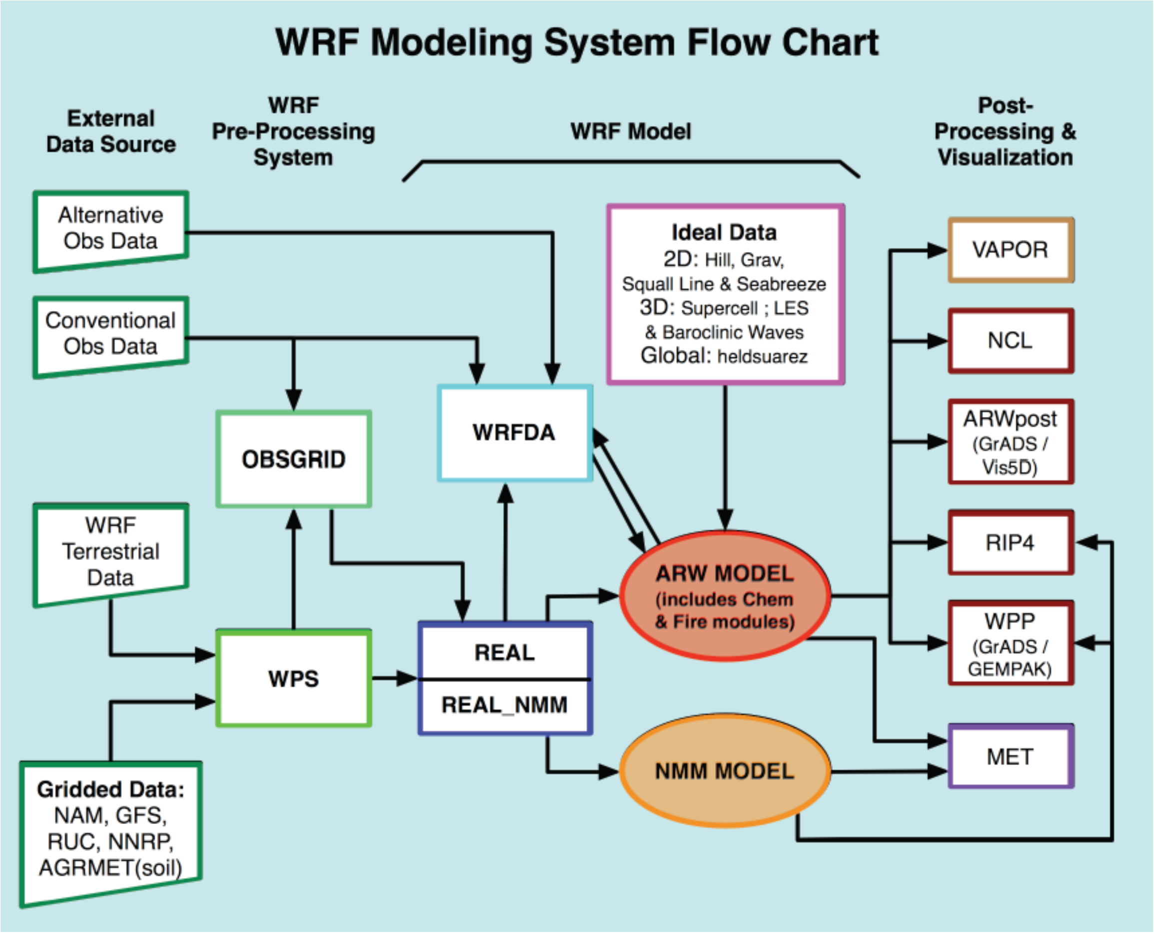


Figure 4.1: WRF Modeling System Flow Chart.

The main functions of WPS are to define the simulation domain area and produce terrain, land use, soil type etc on the simulation domain. It also de-grid GRIB files for meteorological data (u, v, T, q, surface pressure, soil data, snow data, sea-surface temperature etc.) and interpolate meteorological data to the WRF model grid in horizontal direction. The outputs from the WPS are injected to the WRF model. The WRF model runs the simulation with run-time selected namelist switches such as physics choices, timestep and length of simulation etc. In real data cases, it uses boundary conditions from REAL program. Lastly, the output from the WRF model can be interpolated and plotted for study in the post-processing. Detailed explanation of each program can be found in reference (WRF ARW Version 3 Modeling System User's Guide – Chapter 1).

The Nonhydrostatic Mesoscale Model (NMM) core of the WRF model is a fully compressible and non-hydrostatic mesoscale mode with a hydrostatic option (Janjic et al, 2001; Janjic 2003a, 2003b). The model is extremely versatile and supports a variety of capabilities which includes:-

a) Real-data simulations
b) Non-hydrostatic and hydrostatic (runtime option)
c) Full physics options
d) One-way and two-way nesting
e) Applications ranging from meters to thousands of kilometers

WRF offers multiple physics options that can be combined in many ways. Below is the list of all available physics options and the detail on each physics package options can be found in reference (Chapter 5 of NMM-WRF version 3 Modeling System User's Guide).

a) Microphysics

b) Longwave Radiation
c) Shortwave Radiation
d) Surface layer
e) Land Surface
f) Planetary Boundary layer
g) Cumulus Parameterization

However, it is not our intention to try to find the best model over the entire ensemble of possibilities but to add our expertise in vertical sensing to try to quantify the performance of WRF upper atmosphere forecasts with profiling sensors which are not available to the general scientific community especially in an urban environment. Details and rationale of the particular choices made are given in Section 4.2.

4.1.2 Community Multiscale Air Quality (CMAQ) Model

The EPA CMAQ modeling system is a third-generation air quality model and is designed for applications ranging from regulatory and policy analysis to understanding the complex interactions of atmospheric chemistry and physics. It is a 3D Eulerian (e.g. gridded) atmospheric chemistry and transport modeling system that simulate ozone, particulate matter (PM), toxic airborne pollutants, visibility and acidic and nutrient pollutants species through the troposphere. CMAQ model is designed as a “one-atmosphere” model so that it can address the complex couplings among several air quality issues simultaneously across spatial scales ranging from local to hemispheric.

The relationship and purpose of each of the CMAQ processors (and requisite interfaces) and their relation to the chemical transport modeling system are shown in Fig. 4.2. The arrows show the flow of data through the modeling system. Two additional functional features of the CMAQ system are included, one for process analysis, which is primarily for model diagnostic analyses and another one is an aggregation methodology for estimating longer term averaged fields. Descriptions of each of these processors can be found in reference (Ching & Byun), and the associated chapter numbers are also listed to note where details discussions can be found on the topics.

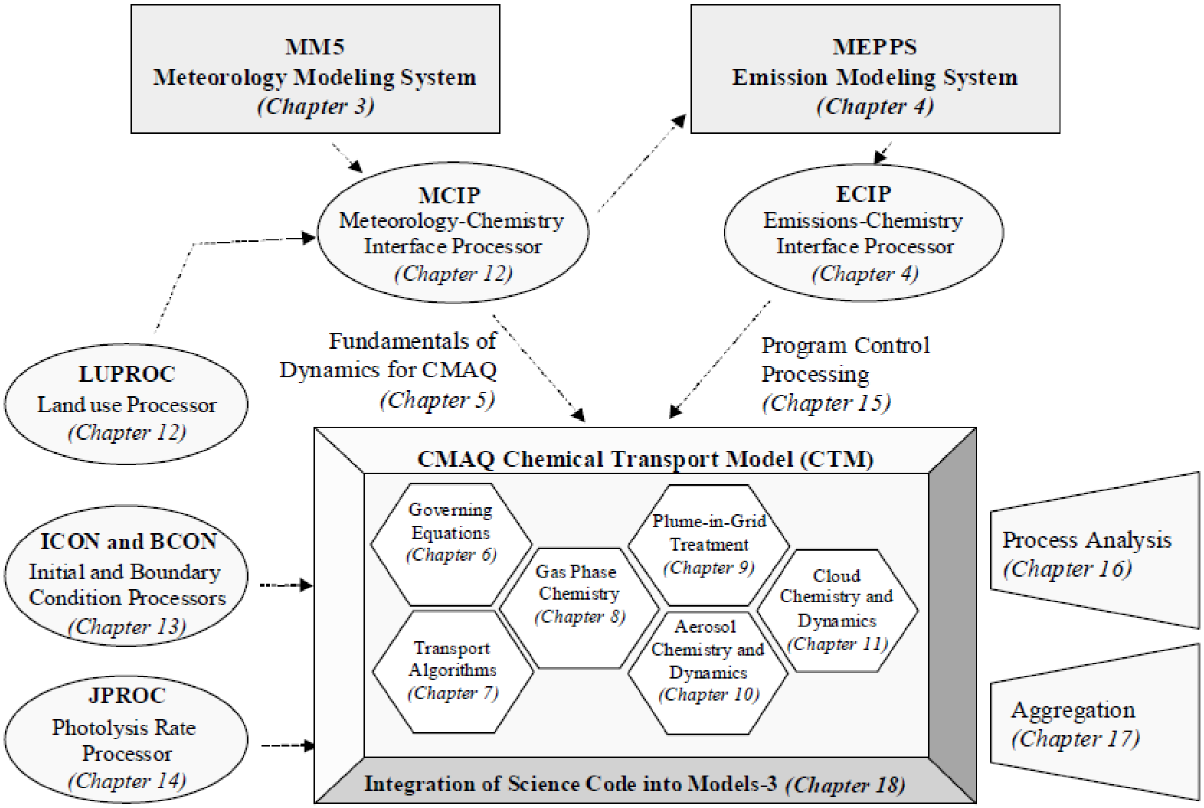


Figure 4.2: Emissions and Meteorological modeling systems and the CMAQ Chemical Transport Model and Interface Processor.

CMAQ uses coupled mathematical representations of actual chemical and physical processes to simulate air quality. It is based upon the underlying concept of preserving mass through a series of contiguous 3D grid cells covering a fixed model grid (e.g. x-y-z array that is fixed in space and covers a particular domain). Thus, CMAQ is also belongs to the Eulerian class of mathematical models. It calculates a mass balance within each grid cell by solving the transport across each cell boundary and chemical transformations within each cell during a given period. Two primary types of inputs: meteorological information and emission rates from sources of emissions that affect air quality are required in CMAQ. In this study we coupled the WRF-CMAQ model to run the simulations as WRF model is one of the meteorological models that are compatible with CMAQ (another meteorological model that is also compatible is Mesoscale Model, MM5). The meteorology inputs dictate the following CMAQ configuration parameters:

a) Horizontal grid coordinate system (e.g. latitude-longitude) and map projection (e.g. Lambert Conformal Conic)
b) Horizontal grid resolution (e.g. the size of the cells composing the grid)
c) Maximum spatial coverage (horizontal geographic extent, e.g. domain) of the grid
d) Maximum vertical grid extent (model top)
e) Temporal extent (the starting and ending dates and time and the meteorology update frequency)

The CMAQ model needs an emission model to estimate the magnitude, location and temporal variability of pollutions sources. In this study, SMOKE emission model was used for computing emissions inputs during 2007 simulations while PREMAQ emission model was used to input 2010 simulations. These emissions inputs must be on the same horizontal and vertical spatial scales and cover the same time period as those used in the air quality model simulation.

Additional details about CMAQ model can be found in reference (Operational Guidance for the CMAQ Modeling System Version 4.7.1.).

4.2 Model Descriptions

4.2.1 WRF PBL Parameterization

As we mentioned earlier, two sets of WRF-CMAQ simulation data for summer 2007 and 2010 were provided by New York State Department of Environmental Conservation (NYSDEC) in this study with the main difference of the datasets of most interest to us being the PBL parameterization scheme. Also, the 2007 simulations were run in the hindcast mode while the 20101 simulations were run in the forecast mode. The details of all physics options for both datasets are listed in Table 4.1 as well as a brief description of the PBL parameterization for each scheme is listed in Table 4.2.

Table 4.1 WRF-CMAQ Model Parametric Schemes

Physic Options	**2007 WRF Run1**	**2007 WRF Run2**	**2010 WRF Run**
PBL	Modified Blackadar	ACM2	MYJ 2.5
Surface Layer	Pleim-Xiu	Pleim-Xiu	NOAH Unified 5-layer Land-surface
Microphysics	WSM6	Morrison II	Ferrier Gridscale
Cumulus	Kain-Fritsch	Kain-Fritsch	Cumulus-Betts-Miller-Janjic
Shortwave Radiation	Dudhia	Dudhia	Lacis-Hansen
Longwave Radiation	RRTM	RRTM G	Fels-Schwartzkopf

Table 4.2 PBL description for each scheme.

PBL Scheme	**Description**
BLK	Symmetric nonlocal scheme. Determines the K-coefficient by the Richardson number and mixing length is first model layer thickness.
ACM2	Asymmetric nonlocal scheme. Non-local upward mixing and local downward mixing with an eddy diffusion algorithm.
MYJ	Eta operational scheme. One-dimensional prognostic turbulent kinetic energy scheme with local vertical mixing.

In performing our validation exercise, we focus primarily on the Modified Blackadar (BLK) and the Asymmetric Convective Model - version 2 (ACM2) Pleim-Xiu (PX) schemes in summer 2007 data. The choice of these schemes was motivated by extensive tests made by Ozone Transport Commission (OTC) Modeling Committee (Baker et al., 2009) in which BLK and ACM2 (PX) were better in general for 2 m surface temp and winds, compared to Yonsei University (YSU) and Mellor-Yamada-Janjic (MYJ) PBL schemes when run in retrospective mode with data nudging applied (Otte, 2008a, 2008b). Furthermore, within the WRF model, it is possible to better observe the differences in the PBL scheme itself since the modular structure of WRF allows different PBL parameterizations to use the same underlying surface layer schemes which in our case is the Pleim-Xiu Surface Layer Scheme (Pleim & Gilliam, 2009). Consequently, we focus only on the ACM2 and BLK schemes. While the retrospective nudging may improve the model error for surface temperature, humidity, and wind (although this is not an independent comparison), but it is not expected to significantly improve upper air parameters since the upper air observation network is very sparse. In addition, the OTC study did not focus their assessment on urban centers. During these retrospective analyses, when using the PX Land Surface Model, the pixel nearest the urban CCNY validation site is classified as 88% "Urban and Build-Up land" as defined by the USGS 24 land use category (Anderson et al., 1976).

In addition to the summer 2007 simulations which used retrospective analysis, summer 2010 forecast data were made using the MYJ PBL scheme which uses the 1.5-order (level 2.5) turbulence closure model of Mellor and Yamada (1982) to represent turbulence above the surface layer (Janjic 1990, 1994, 2001). This is a popular model for PBL representation and the details of the model parameters are also given in Table 4.1. Moreover, the WRF meteorology was used to drive the CMAQ model allowing us to explore how PBL height can be related to

CMAQ forecasts of surface PM2.5. However, note that the WRF model calculated the PBL height is based on the Bulk Richardson number in BLK and ACM2 schemes. Usually in the predawn periods, with the stratified boundary layer and low wind condition the Richardson number is always higher than critical level of 0.25. In this case, the WRF model would put a zero for mixing height, which causes a problem in running the CMAQ model. In processing the WRF data for CMAQ inputs, the zero mixing height will be replaced with the models layer 1 height (depending on the model setup, could range from 10 to 75 meters height). This is a significant issue that can explain some of the very low PBL heights obtained during the predawn periods. In addition, the summer 2010 cases have less bias since the first layer height was larger than the 2007 runs and the PBL height was calculated base on TKE method.

For operational mesoscale air quality applications, the NYSDEC WRF-CMAQ runs are resolved on a 12- by 12-km grid with 1 hour resolution. Although the PBL predictions from WRF are hourly; we are able to reduce the variability in lidar point measurements by similar one hour averaging of the lidar PBL retrievals. Due to natural wind flow, the one hour average is considered as a reasonably representative of the 12 km spatial footprint.

4.2.2 CMAQ PM2.5 Parameterization

Another model variable that we want to assess and to compare with TEOM surface observations is PM2.5. In these comparisons, PM2.5 mass in CMAQ is obtained by simply summing the mass of all particulate components of the Aitken mode (i-mode) and accumulation mode (j-mode) (EPA, 1999a, 1999b; Jiang et al., 2006). The simulation for summer 2007 used CMAQ version 4.6 with the “aero3” aerosol submodule, while the summer 2010 simulation used CMAQ version 4.7.1 with the “aero4” aerosol submodule. In the summer averages, only 13 days for summer

2007 and 14 days for summer 2010 were suitable after all filters were applied in the dataset. The variability of the PM2.5 mass from the TEOM data was available for three urban sites within the CMAQ model pixel (see Table 4.3). Furthermore, we filtered the TEOM data to only those cases in which the standard deviation of the three sites is less than 20 % of the mean.

Table 4.3: TEOM Stations Locations

Site Name	**Latitude**	**Longitude**
CCNY	40.81976^{o} N	-73.94825^{o} W
Manhattanville PO	40.81133^{o} N	$-73.9532\ ^{o}$ W
IS 52	40.81618^{o} N	-73.90200^{o} W

CHAPTER 5

WRF-CMAQ Model Assessments base on Instrument Observations

To begin, we first characterize the seasonal and diurnal trends of the PBL height observed by our intensive lidar (four years) measurements. After that, we focus on the PBL analysis in summer conditions which are the most critical to air quality concerns. First, the PBL height predictions from WRF for the different PBL parameterizations discussed in Chapter 4 are statistically assessed. In general, we find impressive performance matchups during highly convective conditions (e.g. during daytime). On the other hand, to address certain anomalies in PM2.5 predictions occurring during the summer before dawn and after sunset, we directly measure the diurnal structure of the particulates using 24 hours / 7 days ceilometer measurements and deduce that the pollution from CMAQ is generally distributed too close to the surface during these periods as a consequence of underestimations in the WRF PBL height. Lastly, we test the accuracy of WRF outputs such as temperature and water vapor mixing ratio with MWR measurements at low range.

5.1 Lidar Derived PBL Characteristic and Statistic

In this introductory section, we discuss general observation characteristics of the PBL which are obtained from CCNY lidar observations during entire year of 2007 (Fig. 5.1 (a-f) and Table 5.1) (Gan et al., 2010a). To begin, we first categorize the measurements with the PBL height. We use “Low” to signify that the PBL is lower than 1 km which always occurs during cold conditions, “Medium” signifies that the PBL is at 1-2 km, while “High” signifies that the PBL is larger than 2 km which only occurs during warm conditions. We further subdivide each group into strong and weak backscatter for better connecting observations to the aerosol concentration and relative humidity (RH) effects. Note also that strong residual layers occur often during the warm season, which mean the temperature does affect the boundary layer dynamic processes indirectly. In general, the RH in NYC is consistent over the whole year ~55%, while the maximum temperature was about 32^{o}C (June, July and August) and minimum temperature was about -14^{o}C (January and Feb). Refer to Fig. 5.2 for the average surface temperature and RH over 2007.

In order to study the inner structure of the PBL, we divide them into two groups, single layer and multiple layers. According to Table 5.1, single layer BL occur through the whole year, but multiple layers occur only during warm seasons. We believe that the earth radiation, convective and diurnal variations are the main causes of the multiple layers. Moreover, we need to filter those days which contain a lot of clouds on top of the PBL or slightly higher than the PBL, because those clouds create problem such as capping the PBL from growing and unable to define the height of the PBL (e.g. cloud top or cloud base).

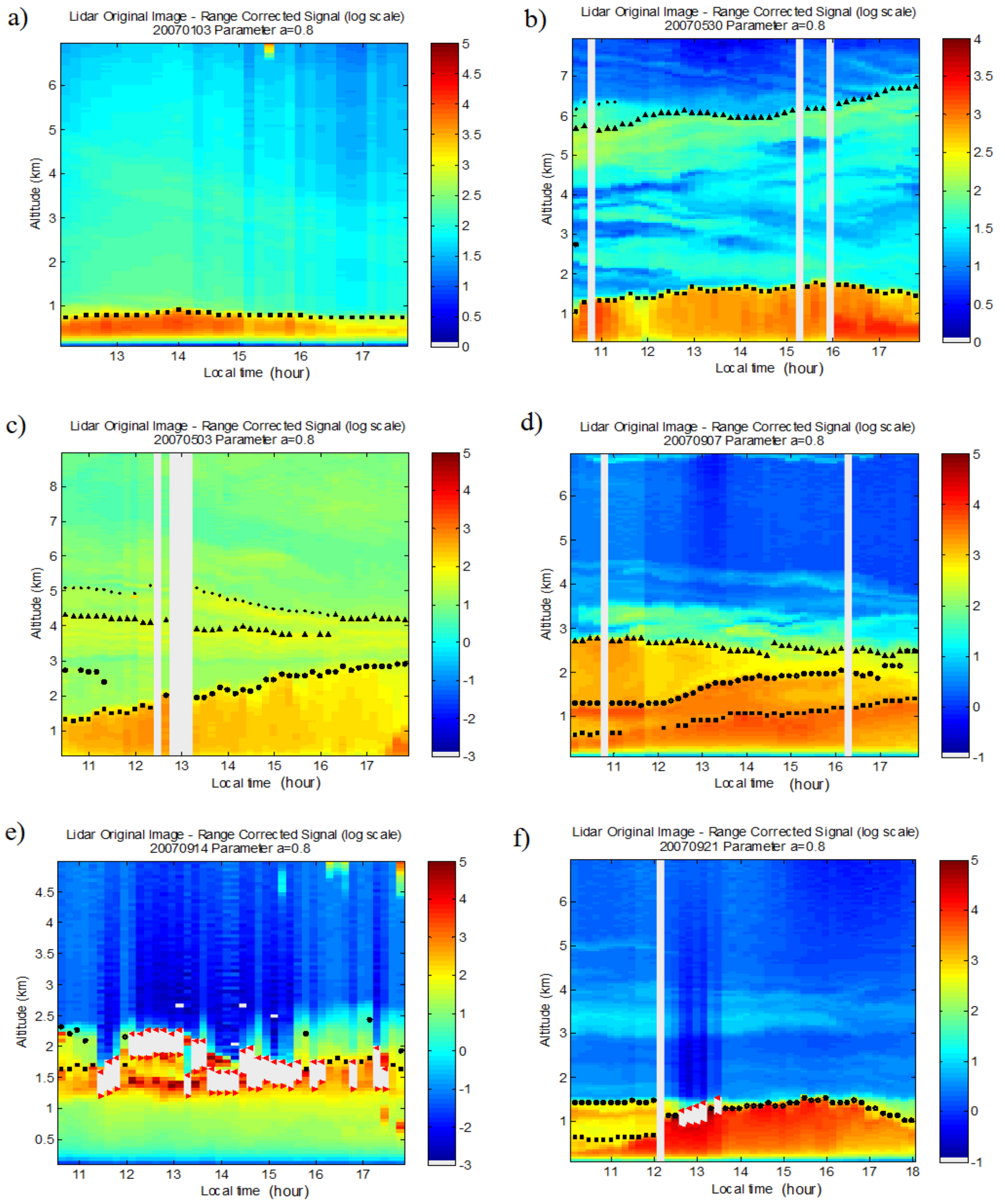


Figure 5.1: Lidar (1064nm) measurements that illustrate different kind of PBL characteristic:- a) Single layer and Low PBL, b) Median PBL with aloft plume occur at 5-7 km, c) High PBL with aloft plumes occur 3-5 km d) Multiple layers and Median PBL with residual layer, e) Median PBL with clouds (removed) attach on top, f) Median PBL with residual layer.

Table 5.1 Statistic Distribution and categories of PBL estimated by lidar over year 2007

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Total**
Observation Days	3	4	7	7	13	2	9	7	7	9	7	2	***77***
Low (<1km)	2	1	3	0	2	0	2	2	4	4	4	2	***26***
Median (1-2km)	1	3	4	6	7	1	3	4	3	5	2	0	***39***
High (>2km)	0	0	0	1	4	1	4	1	0	0	1	0	***12***
Single layer	3	4	6	5	10	1	4	7	4	7	7	2	***60***
Multiple layers	0	0	1	2	3	1	5	0	3	2	0	0	***17***
Cloud topped PBL	3	1	1	2	2	5	6	2	5	2	2	2	***33***
Plume	1	3	5	6	13	1	6	7	7	5	4	0	***58***
Residual	0	0	0	1	2	1	2	6	4	3	1	0	***20***

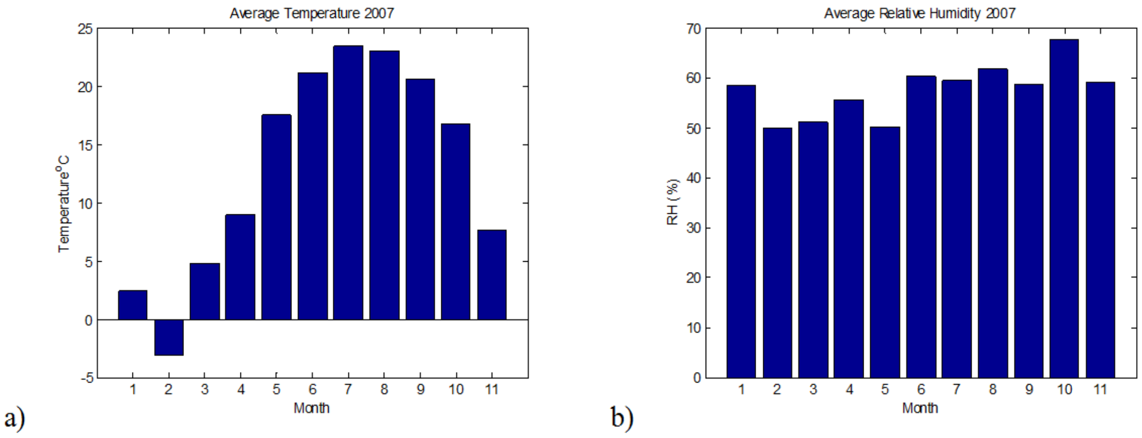


Figure 5.2: Average surface temperature and relative humidity were measured in CCNY weather station. Note that December measurements were missing because of the instrument broke down.

In order to have a general look of the PBL temporal dynamics in an urban area such as NYC, we overlay the PBL trend grouped by seasons using four years (2007-2010) CCNY lidar observations. Due to the weather and human factors, the lidar observations are normally limited from 10:00-18:00 EST. Therefore, the results as shown in Fig. 5.3 (a-b) only illustrate the daytime trend. Table 5.2 summarizes the number of observation days and monthly averaged of maximum PBL height that occur during daytime. The highest of the maximum PBL reached in the afternoon (≈2 km with a deviation of ±400 meters) occur in the summer as seen in Fig. 5.3 (c) while the lowest of the maximum PBL reached (≈1.2 km with a deviation of ±300 meters) occur in the winter. Other than seeing the general pattern of increasing of the maximum PBL height, we can conclude that the increasing of diurnal from morning through late afternoon is much more dramatic during summer and spring due to the surface heating process.

Table 5.2 Summary of observation days and monthly average of maximum PBL height (2007-2010)

Month	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Total Days	12	16	29	28	30	16	32	24	26	25	22	11
Max Height (km)	0.99	1.30	1.88	1.78	1.80	2.05	2.23	2.02	2.21	1.11	1.18	1.48

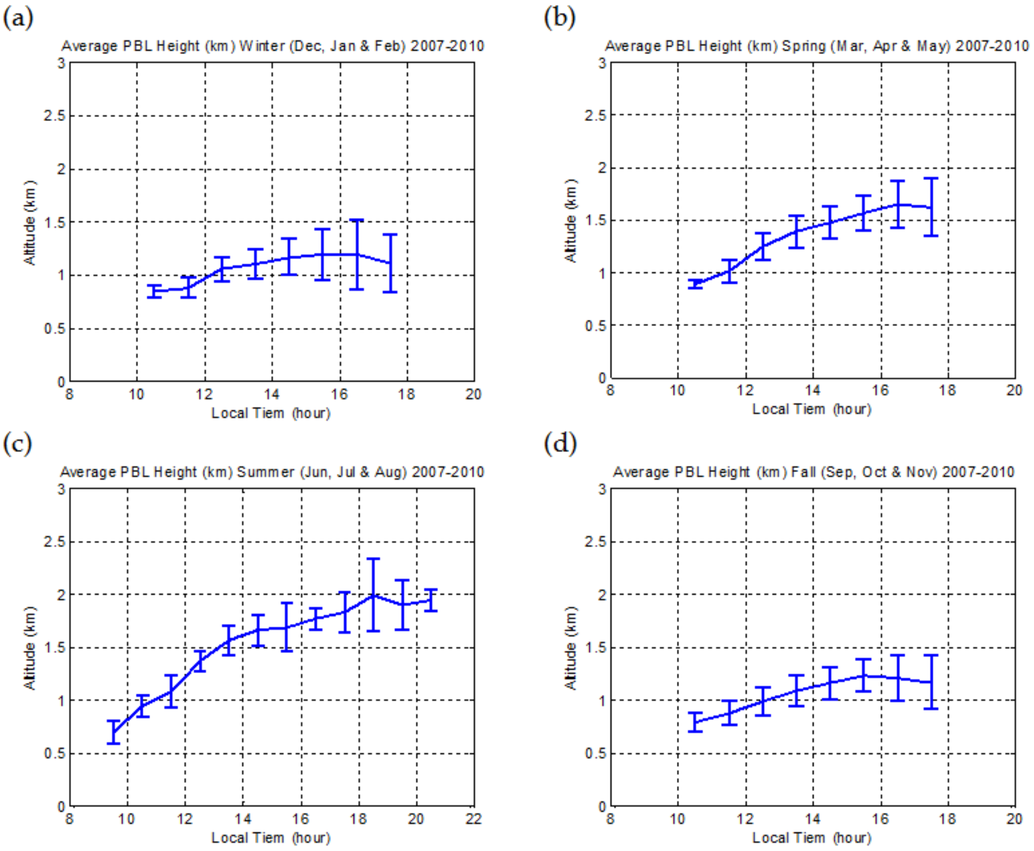


Figure 5.3: Monthly PBL trend average over 4 years (2007-2010).

5.2 WRF PBL Height Comparison Studies

In this section, we present our statistical comparisons between the lidar derived PBL height and the WRF model outputs. As a reminder, even though we run the lidar over the entire year, the focus of our comparisons for the thesis is the summer period. There are several motivations:-

1) The lidar measurement density is highest in the summer.
2) The focus on convective heating mechanisms makes the summer particularly important and leads to easier interpretation of the results.
3) The NYSDEC results for the WRF-CMAQ found the main anomalies of surface PM2.5 during summer.
4) The CMAQ data itself were only available to us over summer.

Therefore, intensive study for multiple summers (2007-2010) was the most efficient approach to illustrate vertical measurement potentials for model validation.

5.2.1 Data Filters to Improve Comparisons

To improve the data matchup quality, the following filters are applied into the lidar matchups.

1) The WCT method was applied to all non-cloud marked time slices independently and a continuity filter (described in Section 3.1.4) was used. Note that the lidar observation for PBL retrieval is limited to height at 5 km. If there is high cloud occurs above 5 km, the WCT is still applied to the measurement because the high cloud should not have significant effect on the PBL dynamic theoretically.

2) Only cloud free cases are considered. In fact, this is a limitation of the lidar measurements since operation under cloudy conditions can result in strong backscatter signals that would often saturate the detector. Since we are ultimately concerned with cases where pollution exceedence is of major concern and to avoid local coastal contamination issues, we remove from our comparisons all sea-breeze cases identified through wind data from Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) Model.

3) We restrict comparisons to cases where at least five lidar PBL measurements are made within the hour and the standard deviation is less than 20% of the mean in lidar observation. This limits comparisons to cases where temporal (e.g. spatial) homogeneity is reasonably met.

4) In addition, a further filter based on global performance was used which required that for any give day, at least 4 hours of data passed Quality Control.

5.2.2 Statistical Comparisons

The main results for the comparison of both the BLK and ACM2 predictions with lidar derived PBL height are given in Fig. 5.4 (a-d).

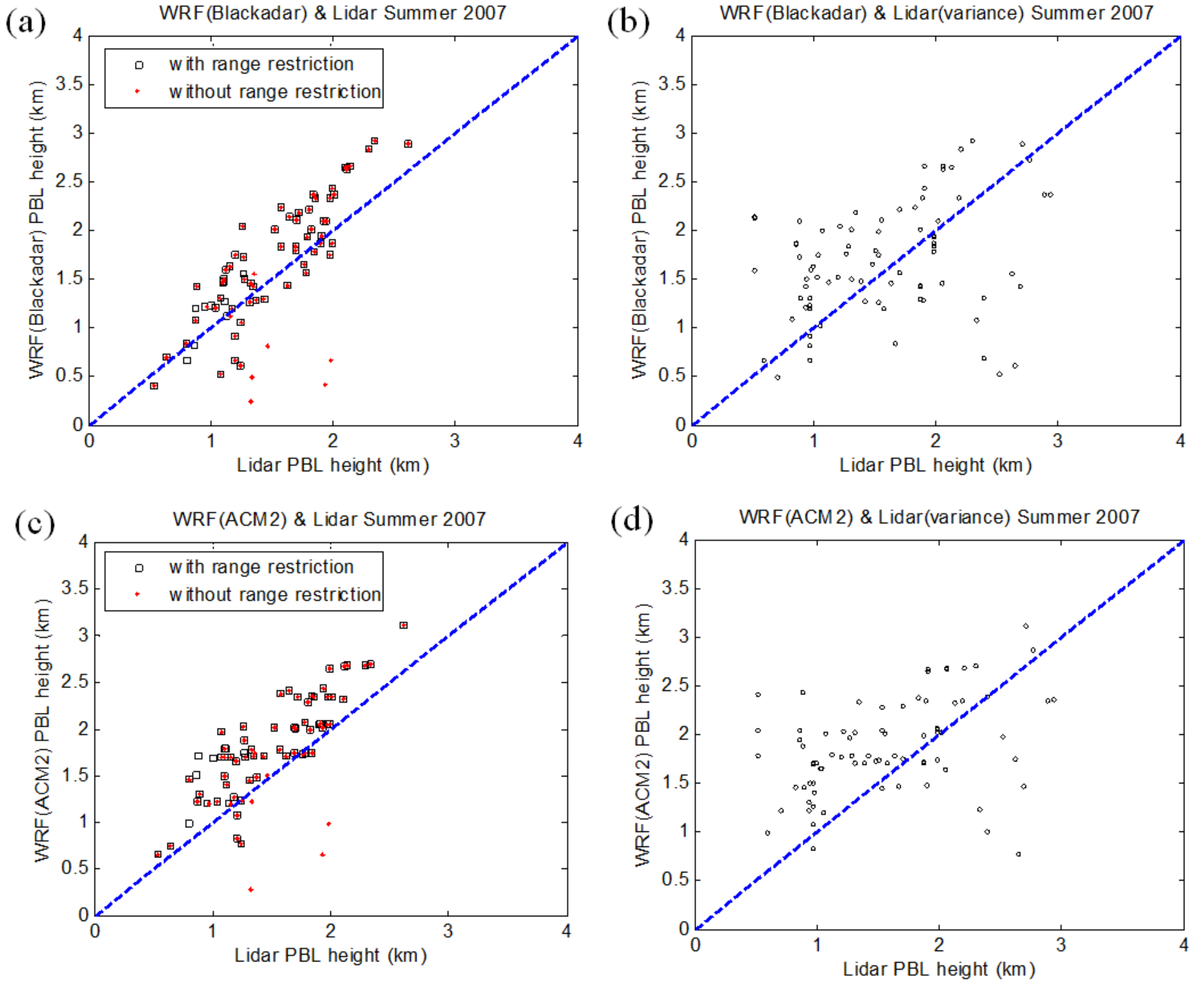


Figure 5.4: WRF versus Lidar PBL heights. (a) WRF-Blackadar scheme vs. lidar-WCT method, (b) WRF-Blackadar scheme vs. lidar-variance method, (c) WRF-ACM2 scheme vs. lidar-WCT method, (d) WRF-ACM2 scheme vs. lidar-variance method in summer, 2007.

Before making these comparisons, we test the lidar derived PBL height using multiple methods to further demonstrate the advantage in applying range restrictions to the WCT in comparison to the WCT without range restriction and the maximum variance method. In particular, as discussed in Chapter 3, the maximum variance method suffers both by being fooled by residual layers as well as sub layer structure within the PBL which results in the poor correlation seen in Fig. 5.4 (b) and (d). In addition, the improvement (e.g. WCT without height restriction is spreading over a larger area) of the WCT includes the height restriction is apparent in panels (a) and (c).

With the above considerations, we can now make our comparisons based primarily on the modified WCT approach in assessing model performance. In presenting the statistics, we make use of 3 parameters; the linear correlation coefficient (R), as well as the mean and standard deviation of the error instead of the slope and y-intercept of the regression line. The reason we need all of these statistics is that the linear correlation coefficient has meaning even in the presence of a constant bias. For example, if the output fits exactly to a linear model with arbitrary slope and y intercept, the correlation coefficient is still 1. Therefore, simply providing the correlation coefficient is not sufficient to completely describe the quality of the model. For this reason, we include the mean error and the error standard deviation defined on the data $\varepsilon(km) \equiv H_{Lidar} - H_{WRF}$. We believe this is a more direct measure than simply providing the slope and y intercept of the best fit regression line which in its own way provides some estimate of bias but can be misleading.

First, we note that the BLK has slightly better correlation (R=0.87) than the ACM2 scheme (R=0.84) (see Table 3.2 and Fig. 5.4). However, as seen in the histogram analysis of Fig. 5.5, the bias of the BLK scheme is smaller than that observed in the ACM2 scheme with a mean bias of m=190 meters for BLK and a mean bias of m=340 meters for the ACM2 scheme. On the other hand, the fluctuation of the deviation is quite high for the BLK scheme which is about σ_{BLK}=300 meters while the ACM2 schemes standard deviation is about σ_{ACM2}=280 meters. To place these biases and deviations in their proper context, keep in mind, while the lidar based methods matched against meteorologically based methods such as the parcel method, show no appreciable bias and average deviations on the order of ±200 meters are to be expected (Hennemuth & Lammert, 2006). Besides, because of the limitation of the nonlinear vertical grid resolution (spatial resolution ≈100 meters at range 0.5-1 km and ≈300 meters at range 1-2 km), WRF model use linear interpolation to locate the PBL height. Therefore, additional systematic bias between the WRF meteorological profiles and the lidar aerosol profile may be expected. Therefore, we may conclude that, in general, the observed biases are reasonable within the errors that might be expected. Still, the results illustrate that at least for ACM2, the overbias is more significant so we tentatively conclude the BLK scheme is considered performs best under the 2007 study period.

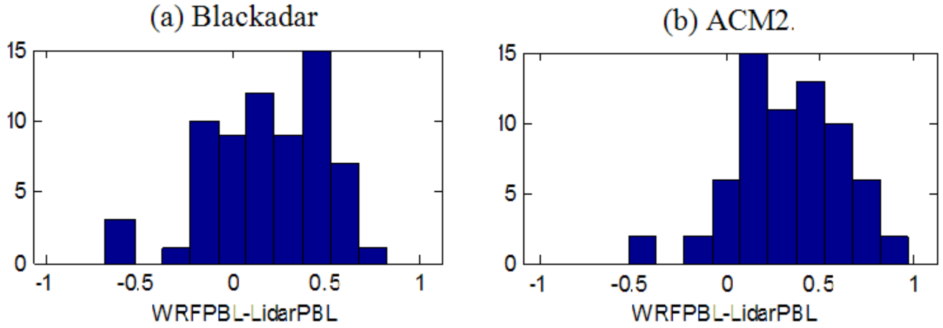


Figure 5.5: Histogram of PBL height difference WRF PBL-Lidar PBL (a) Blackadar (b) ACM2.

The results for the 2010 PBL height comparisons using the MYJ scheme in forecast mode are given in Fig. 5.6 (a). Even though the correlation is high (R~0.84), the fluctuations for the 2010 forecast data is much larger as seen in the histogram data as shown in Fig. 5.6 (b) with a standard deviation of σ_{MYJ}~400 meters in comparison to σ_{BLK}~300 meters and σ_{ACM2}~280 meters (see Table 3.2). This is consistent with the fact that forecast uncertainties are expected to significantly increase the fluctuations of all dynamical parameters.

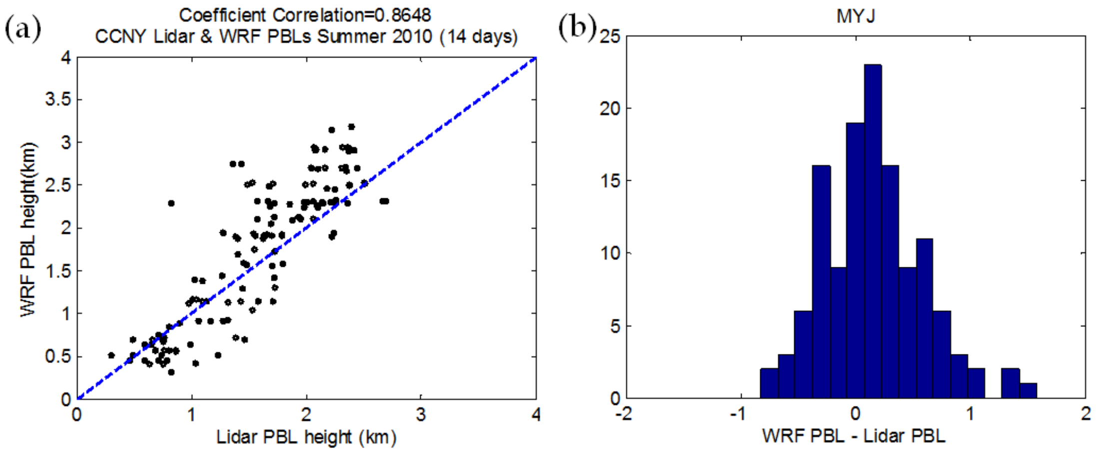


Figure 5.6: (a) Linear regression of WRF PBL height with lidar derived PBL height for selected 14 days on June, July and August, 2010 (b) Statistics of PBL height model versus measurement errors (WRF PBL – Lidar PBL in unit km) summer 2010.

5.2.3 Discussion of Model Biases

In an effort to try to better understand the systemic overbiases which seem to occur in comparing the lidar derived PBL against the WRF model outputs, we first attempt to explore possible diurnal patterns in the relative error $\varepsilon(km) \equiv H_{Lidar} - H_{WRF}$. Mean and standard deviation (STD) statistics of error between lidar and model observations as function of time of day are listed in Table 5.3. As shown in the Fig. 5.7 for both the BLK and ACM2 schemes observed in 2007, the overall bias was on the order of the uncertainty of the measurements although the ACM2 bias was significantly larger. However, we find no discernable trend in the bias statistics over the course of the day. Such an overall biases seems to indicate that perhaps a simple refining of the PBL parameterization model may be needed. On the other hand, we note that the bias observed in our 2010 MYJ comparison was relatively small in the morning and increased slightly in the afternoon with associate uncertainty. Such a trend may be in part due to the fact that the 2010 dataset was based on forecast runs and not the hindcast runs used in the 2007 data comparisons. However, care must be made in making any conclusions since the overall statistics was very sparse in each of these runs so only the most preliminary observations can be made.

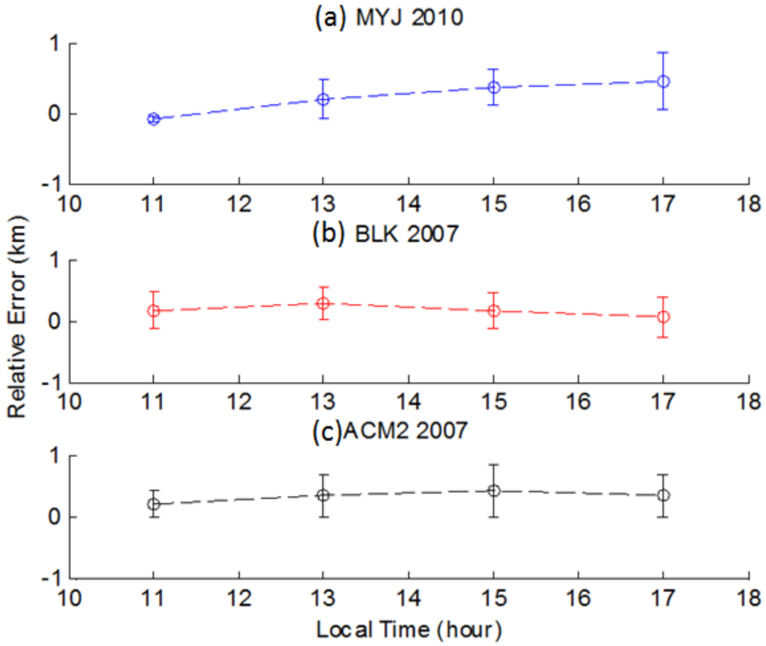


Figure 5.7: Statistical Bias trends for all cases considered (a) MYJ 2010 – forecast mode, (b) BLK 2007 – hindcast mode and (c) ACM2 – hindcast mode.

Table 5.3 Mean and standard deviation (STD) statistics of error between lidar and model observations as function of time of day.

PBL Scheme	**Time**	**10-12**	**12-14**	**14-16**	**16-18**
2010 MYJ	**Mean(x-y)**	-0.078	0.198	0.366	0.450
	STD(x-y)	0.40	0.282	0.256	0.405
2007 BLK	**Mean(x-y)**	0.163	0.283	0.158	0.051
	STD(x-y)	0.309	0.271	0.303	0.326
2007 ACM2	**Mean(x-y)**	0.208	0.348	0.422	0.352
	STD(x-y)	0.245	0.263	0.287	0.345

Since the diurnal study was inconclusive regarding the underlying cause of the bias, we also considered assessment based on meteorological conditions including surface and upper-air temperature, moisture (RH), and wind speed - wind direction. In particular, it would be reasonable to believe that the best model performance would be based on conditions where mechanical shear is limited and convective conditions dominate. This would be accompanied by low wind speed in general which does not undergo any changes due to frontal movement (e.g. a hot stagnant summer day). In Fig. 5.8, such a study is considered in an effort to separate a good model fit (Aug 30) against a strong model overbias (Aug 20) condition. The meteorological data are obtained from the HYSPLIT data in archive (e.g. analysis) mode (NAM 12 km, 3hourly) to eliminate forecast error issues that would accompany the use of WRF are shown in Fig. 5.9 (a-d) and Fig. 5.10 (a-d). The following observations can be made.

1. There seems to be little significant difference between the overall wind speeds and the direction during the PBL mixing is reasonably stable.
2. The surface and upper atmosphere temperatures seem to be similar as well.
3. On the other hand, the upper atmosphere water vapor (e.g. RH much higher) loading seems to be much larger for the strongly overbias case. This is not surprising since a strong vertical mixing of moisture is a symptom of a high model PBL height.

Therefore, it is reasonable that a strong factor in many of the overbias cases is an overestimate of turbulent water vapor flux made by the models. Of course, much more statistical assessment is needed which may be possible in summer 2011.

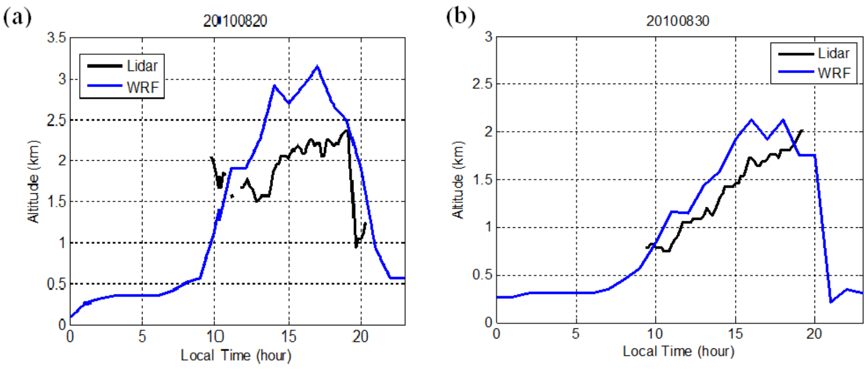


Figure 5.8: WRF versus Lidar PBL comparison on (a) Aug 20, 2010 for overbias case and (b) Aug 30, 2010 for good match case.

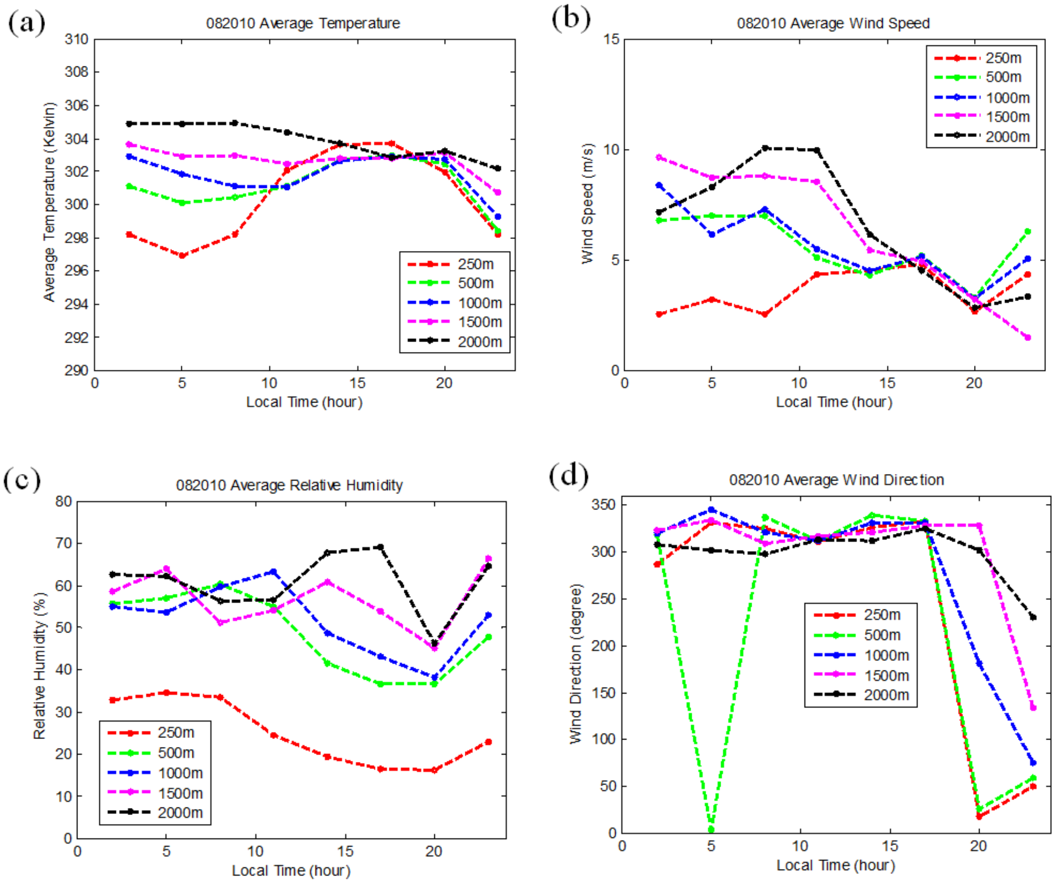


Figure 5.9: Meteorology and lidar derived boundary layer parameter diurnal cycle for low bias day – Aug 30 2010 (a) Temperature, (b) Horizontal Wind Speed, (c) Relative Humidity and (d) Horizontal Wind Direction*.

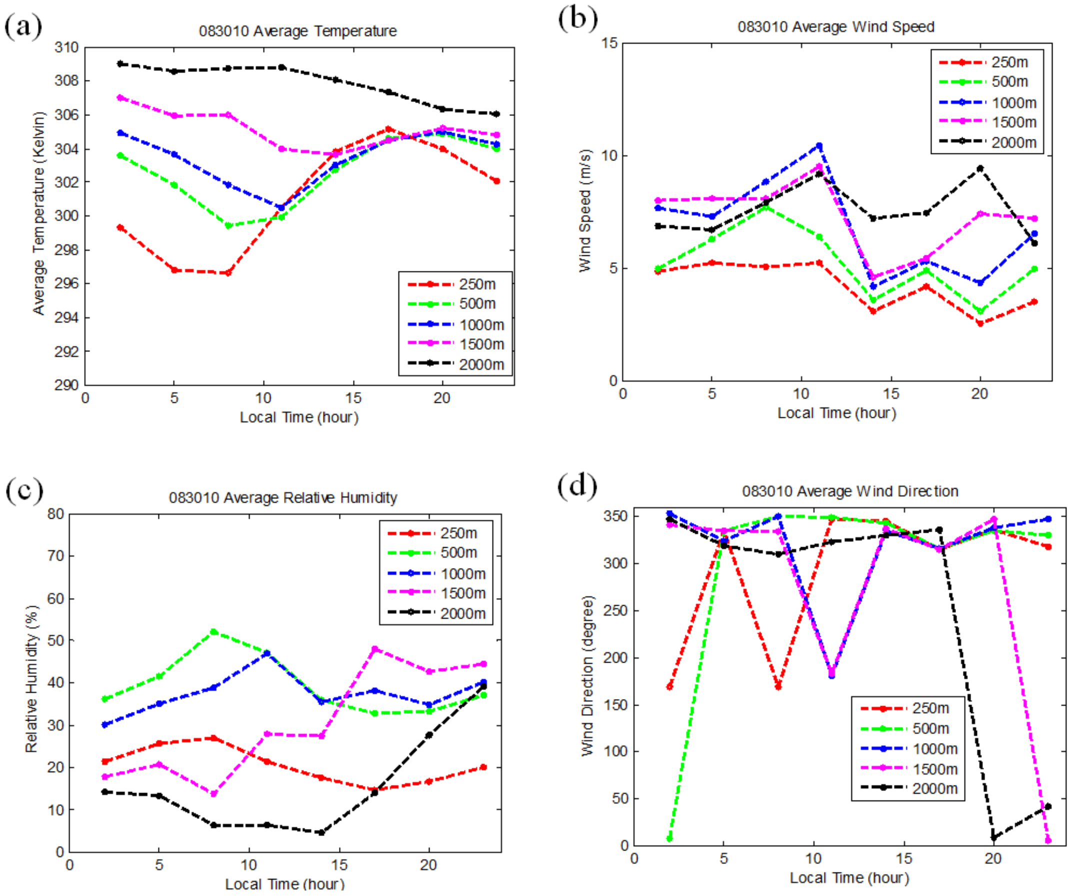


Figure 5.10: Meteorology and lidar derived boundary layer parameter diurnal cycle for low bias day – Aug 30 2010 (a) Temperature, (b) Horizontal Wind Speed, (c) Relative Humidity and (d) Horizontal Wind Direction*.

*wind direction is represented as a reference direction start from North (0^o) which is measured clockwise and the direction from which the wind is blowing.

5.3 CMAQ PM2.5

As mentioned before, PBL height has a strong affect on air quality forecasts and it is natural to assess the performance of WRF-CMAQ for surface PM2.5 simulations. In Doraiswamy et al. (2010) study, strong anomalies were observed between the CMAQ PM2.5 forecasts and the TEOM measurements from the AIRNow network. The most dramatic discrepancies were seen in the urban summer matchups where predawn overestimations from CMAQ were observed. These biases are presented in Fig. 5.11 (a-b) where the summer averaged PM2.5 values from the TEOM network were compared to CMAQ predictions for both summer 2007 and summer 2010. In restricting to the lidar cases (e.g. clear and sunny days), we assured that complications arising from precipitation and frontal activity were avoided.

While the overestimations appear most intense for summer 2007, significant overestimates are also seen in the summer 2010 comparisons. The direct source of this over bias was not clear but it was suggested by Doraiswamy et al. (2010) that errors in the PBL height could play a significant part of this anomaly. No proof of this was available so we attempt to use the vertical backscatter data to test the distribution of particulates inside the PBL.

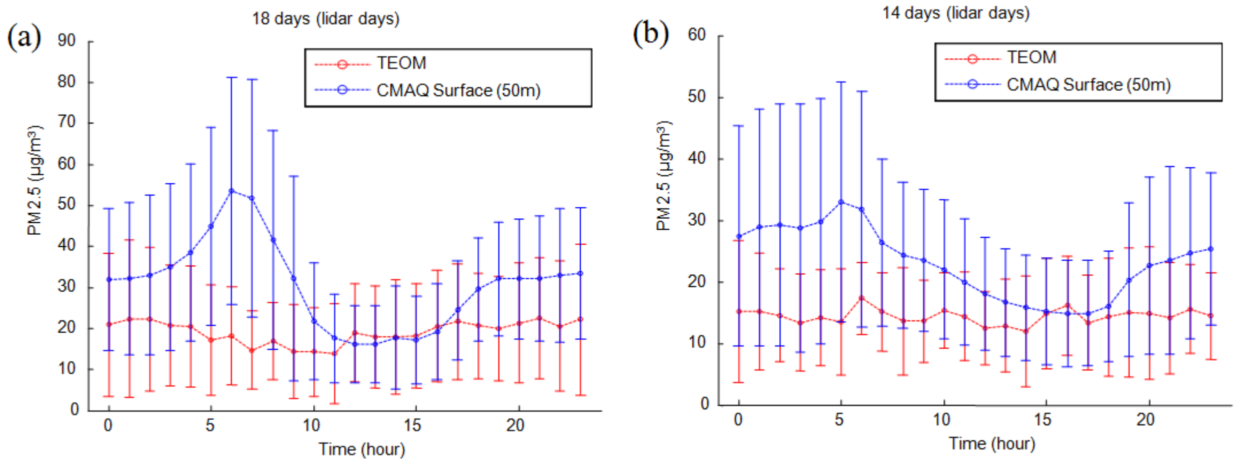


Figure 5.11: Selected days averaged PM2.5 mass concentration (μg/m^3) diurnal cycle comparison of CMAQ and TEOM during (a) Summer 2007, b) Summer 2010.

If the enhancement of the surface particulate matter (PM) measurements during predawn periods is mainly due to the imprecise modeling of the emissions transportation and sources, then the vertical profiles of the aerosol loading can be used to assess this directly. To do this, we make use of an eyesafe ceilometer (Vaisala CL31), which provides low altitude 24 hours / 7 days aerosol backscatter observations. Unfortunately, due to poor SNR, ranges for suitable PBL height retrieval are extremely limited (< 1.2 km). However, for our low altitude analysis, this limitation is not critical and the ceilometer is most useful in probing the diurnal structure of the near surface pollution patterns from both the TEOM and CMAQ PM2.5 results. As an instructive example, we plot in Fig. 5.12 the correlation coefficient for the TEOM PM2.5 - Ceilometer backscatter measurement (month average) for the summer months July and Aug 2007 as a function of height. In doing this, a 300 meters sliding averaging window was used to process the ceilometers data to reduce noise artifacts. Most importantly, we note that the near surface correlations (R≈0.85) persist to heights ≈500 meters followed by a strong reduction of correlation as expected for higher altitudes, was consistent with other studies (Münkel et al., 2007). This demonstrates strong correlation between PM2.5 and ceilometer backscatter. These high correlations near the surface justify the use of the ceilometers backscatter as a reasonable probe of PM2.5.

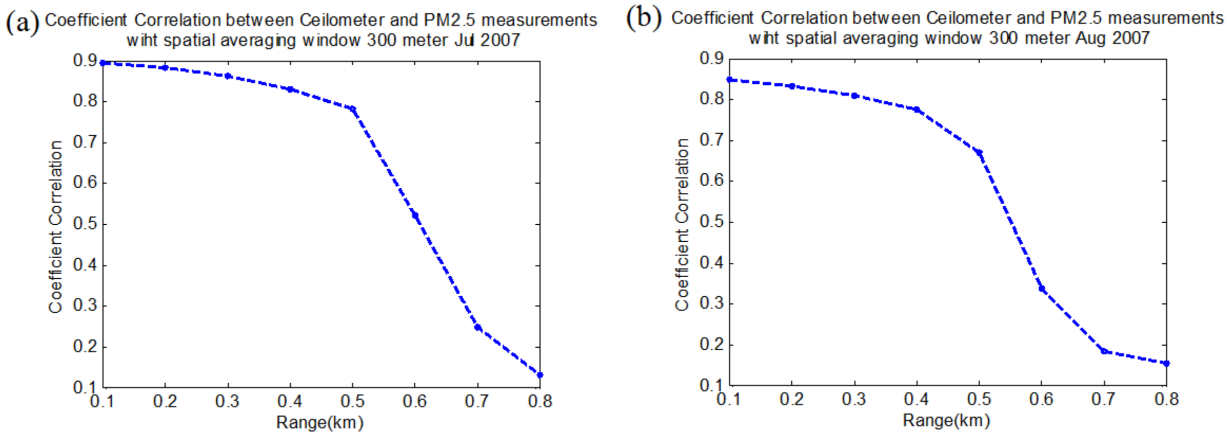


Figure 5.12: Coefficient correlation between ceilometer near-surface backscatter signal and PM2.5 measurements in the function of range for (a) July 2007 and (b) August 2007.

In Fig. 5.13 (a-b), the path averaged of CMAQ PM2.5 and the ceilometer backscatter over the diurnal cycle for different altitude ranges is displayed. In panel (a), the results illustrate how the CMAQ PM2.5 distribution is compressed near the surface during the morning spike anomaly. In fact, as the altitude increases, the PM2.5 diurnal distribution becomes peaked during mid-day and evidence of the spike behavior is removed. On the other hand, the ceilometers path averaged backscatter in panel (b) does not observe these pollutant spikes even for the lowest vertical bins. While a small increase is seen in the ceilometer backscatter relative to mid day, the contrast between morning and mid-day is smaller with a contrast coefficient $C(X) \equiv (X_{\max} - X_{\min})/(X_{\max} + X_{\min})$ of approximately 12% compared to ~ 50% for the CMAQ near surface measurements which is much more consistent with the TEOM measurements. On the other hand, it is interesting to see that as the height of the ceilometer path increases, we recover the general diurnal trend seen in the CMAQ data retrieval which is consistent with the hypothesis that the observed spikes are mainly due to vertical apportionment of the PM2.5. In other words, PM2.5 of the CMAQ model seems to be distributed too close to the surface.

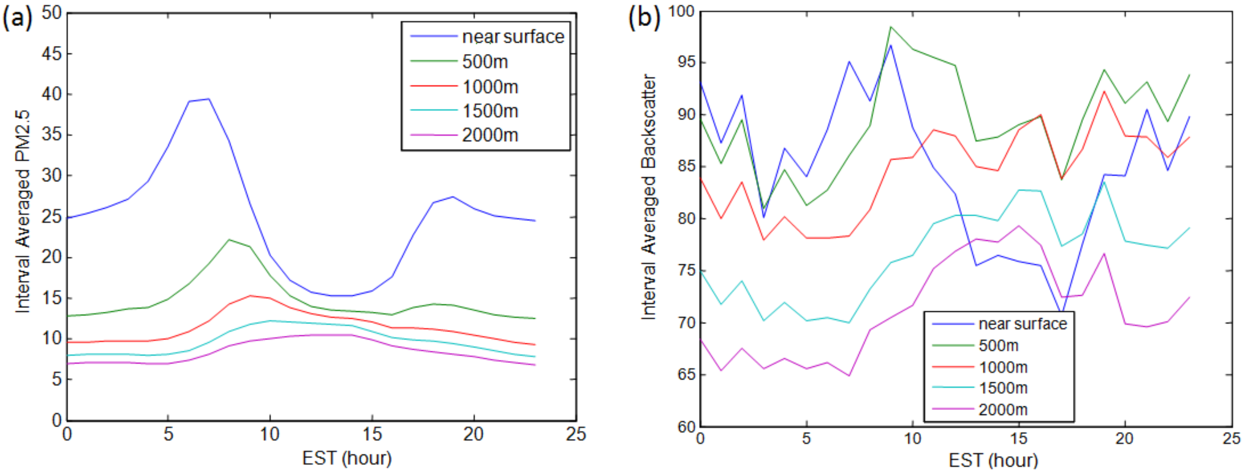


Figure 5.13: Cumulative path averaged parameters for different altitude ranges over the diurnal cycle (a) CMAQ PM2.5 (μg/m^3) and (b) Un-calibrated ceilometer backscatter (unitless).

Further evidence that total emissions is not the main issue is presented in Fig. 5.14 where we plot for the summer 2007 case, the linear correlation coefficient (R) binned by hour between the TEOM PM2.5 measurements and the path averaged PM2.5 mass from the CMAQ model at different vertical height levels. Most relevant is the enhancement of the correlation for the predawn PM emission case when the altitude is larger (black dash line for R=3000 meters). This illustrates again that at some extent, the primary emission is not the dominant issue. In practice, these emissions are being more evenly distributed in the PBL. This is proven by a larger surface correlation (red dash line for R=250 meters) in the afternoon as a result of a well mixed PBL.

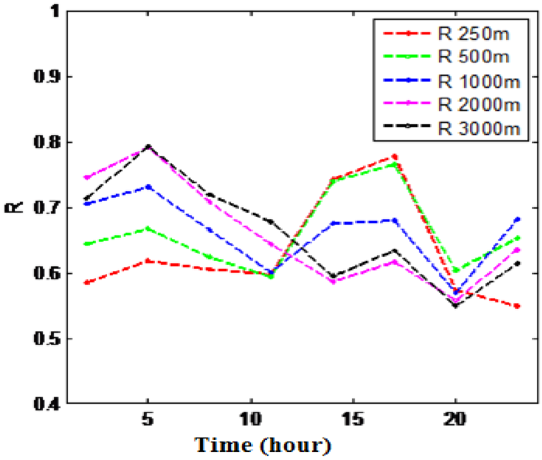


Figure 5.14: Linear correlation coefficient binned by hour between the TEOM PM2.5 Summer 2007 measurements and the path averaged PM2.5 mass from the CMAQ model for different vertical height levels. (R: correlation coefficient; 250 m: path average from the surface to 250 m altitude; 500 m: path average from the surface to 500 m altitude and etc)

It is important to remember that when we compared the vertical structure of the CMAQ against the ceilometers in Fig. 5.13 and 5.14, we found that the strong surface emission behavior in the diurnal pattern predicted by CMAQ within the first 500 meters from the surface, were not seen in the ceilometer observation. However, it may be argued that we were matching PM2.5 to optical backscatter which may be significant problem in the comparison. We mentioned that this should not be a significant issue because the TEOM-ceilometer regression has high correlations in the first 500 meters from the surface; it is useful to consider a more direct matchup such as CMAQ output extinction. This CMAQ extinction variable is based on semi-empirical parameterization which connects the CMAQ component masses to optical extinction (Malm et al., 1994). The performance of this approach against a more direct Mie Scattering approach is considered in Chapter 6. Fig. 5.15 (a) shows the averaged optical extinction data and Fig. 5.15 (b) shows the accompanying seasonally averaged relative humidity (RH). The first point is that the CMAQ extinction parameter has the same near surface behavior with the PM2.5 mass so that it is reasonable to diagnose the CMAQ PM2.5 with the lidar and ceilometer backscatter. In this direction, it is clear that CMAQ primary emissions are not properly being distributed vertically. On the other hand, enhanced extinction is seen in the CMAQ retrievals in the upper atmosphere. When compared to the WRF RH profile, it seems like this is a direct consequence of the enhanced scattering due to hygroscopic humidification. In fact, this enhanced backscatter does exist in our lidar images where increased humidity at the top of the PBL is often accompanied by enhanced RH due to temperature inversion and increase of hazy layers capping the PBL in the afternoon.

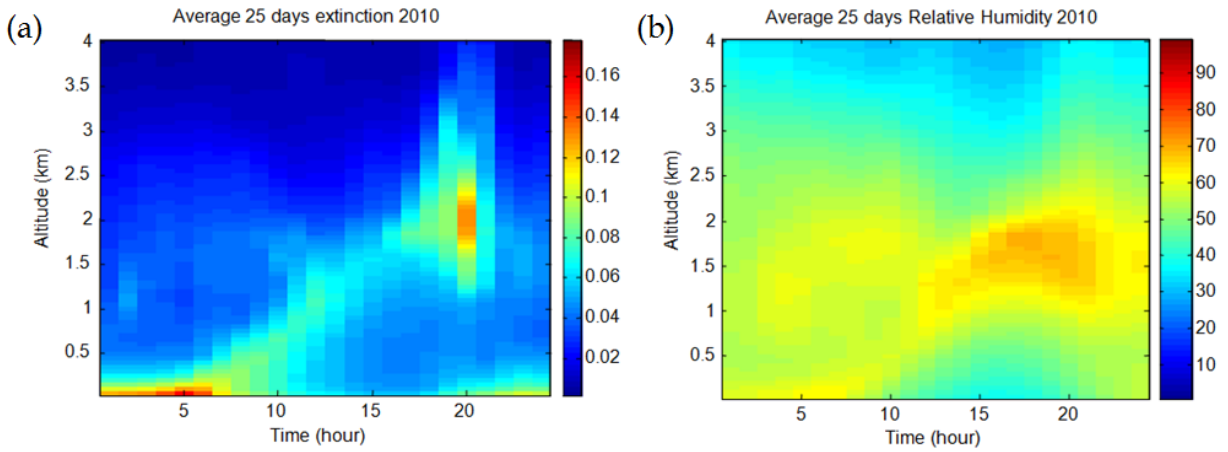


Figure 5.15: (a) Optical extinction (km^{-1}) at wavelength 550 nm based on the MALM parameterization within the CMAQ product (b) Simultaneous WRF RH (percent) retrievals.

In addition, we explored the temperature and water vapor mixing ratios from the WRF model for the summer 2010 forecasts through comparisons to retrieved profiles using a MP-3000A MWR (Knupp et al., 2009) (see Section 2.4 for full instrumentation description). Validation of the MWR upper air temperature profiles by radiosonde measurement is difficult because the distance to the nearest radiosonde station is very far (≈ 60 km). However, we have excellent matchups (< 3%) of the total water vapor integrated which were made with nearby NOAA GPS MET Stations shown in Fig. 5.16.

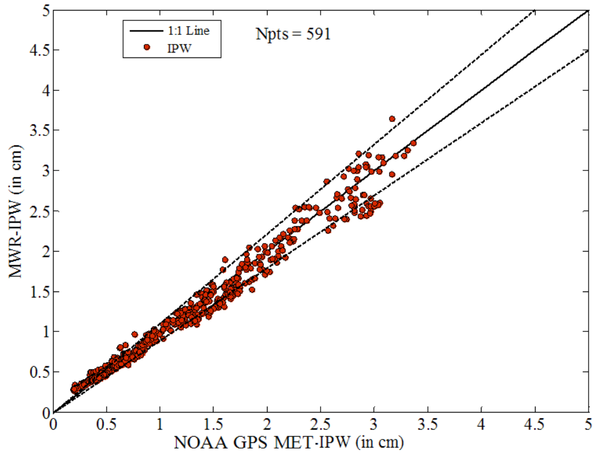


Figure 5.16: Integrated Precipitation Water (IPW) of MWR versus IPW of NOAA GPS MET Station (Sept 2010 – Jan 2011).

In order to better observe the diurnal behavior of WRF model and to try to understand why the PBL height retrievals are fairly good, the temperature and the water vapor mixing ratio from WRF data are used to compare with MWR measurements. The measurement on September 1, 2010 is selected to demonstrate this comparison. Note in this case, the PBL behavior from WRF and lidar are similar as shown in Fig. 5.17. In Fig. 5.18 (a-d), the temperature profile at different heights is made. We note that even though they are compared at different heights, they still have similar diurnal cycle and agreement is pretty good to 1000 meters. Similar comparisons are done on the water vapor mixing ratio (see Fig. 5.19 (a-b)) and similar results are observed. Carrying the comparisons to higher altitude is difficult since we do not have a very good estimation of accuracy at higher altitudes.

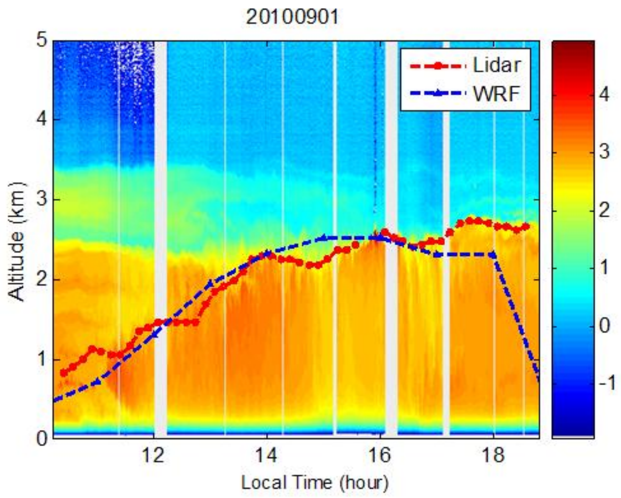


Figure 5.17: Lidar Range-Corrected Measurements on Sep 1, 2010 with PBL height marked (red=lidar and blue=WRF).

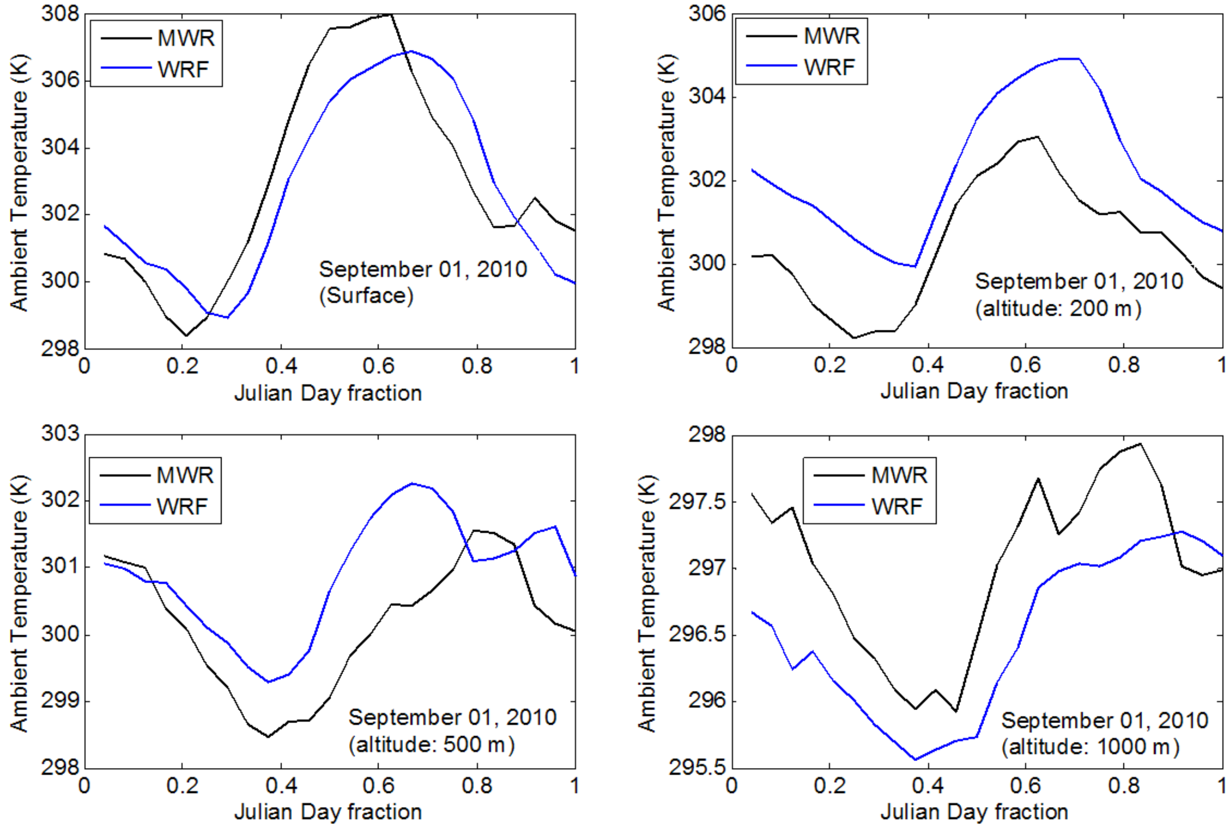


Figure 5.18: Comparison Ambient Temperature Profile between WRF and MWR on Sep 1, 2010 at different height (a) Surface, (b) 200 m, (c) 500 m, and (d) 1000 m.

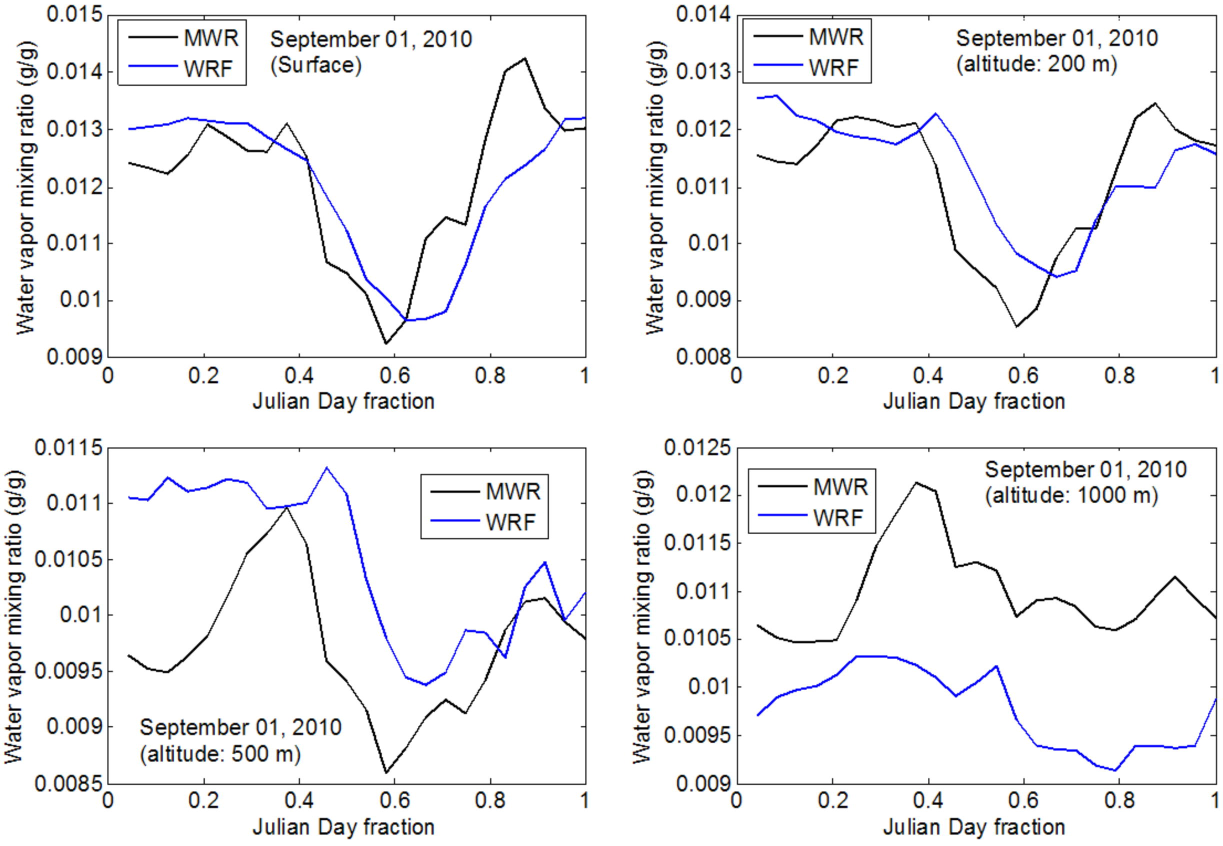


Figure 5.19: Comparison Water Vapor Mixing Ratio Profile between WRF and MWR on Sep 1, 2010 at different heights (a) Surface, (b) 200 m, (c) 500 m, and (d) 1000 m.

5.4 Relation between Diurnal PM Mass and WRF PBL Height

It is interesting to address that the PM2.5 spike is less intense in the observation of summer 2010. Although direct observations using the ceilometers are no possible since the instrument is no longer in operation, we get some insight by inter-comparing the summer 2007 and summer 2010 CMAQ predictions. In Fig. 5.20, the 3D structure of the CMAQ outputs are compared. For convenience, the different PBL height retrievals are superimposed. The most dramatic observation is that the summer 2010 PBL heights are significantly larger on average for summer 2010 during the sunrise / sunset periods, in comparison to the more compressed PBL height of summer 2007. This seems to trap the primary emissions at surface and not letting them vent upwards. On the other hand, for summer 2010, the expansion of the PBL height seems to allow at least partial venting of the pollutants, which is more in line with ceilometers based observations. This is reinforcing the general inverse relationship between overbiases in the surface PM2.5 and the PBL height during the predawn / post sunset periods.

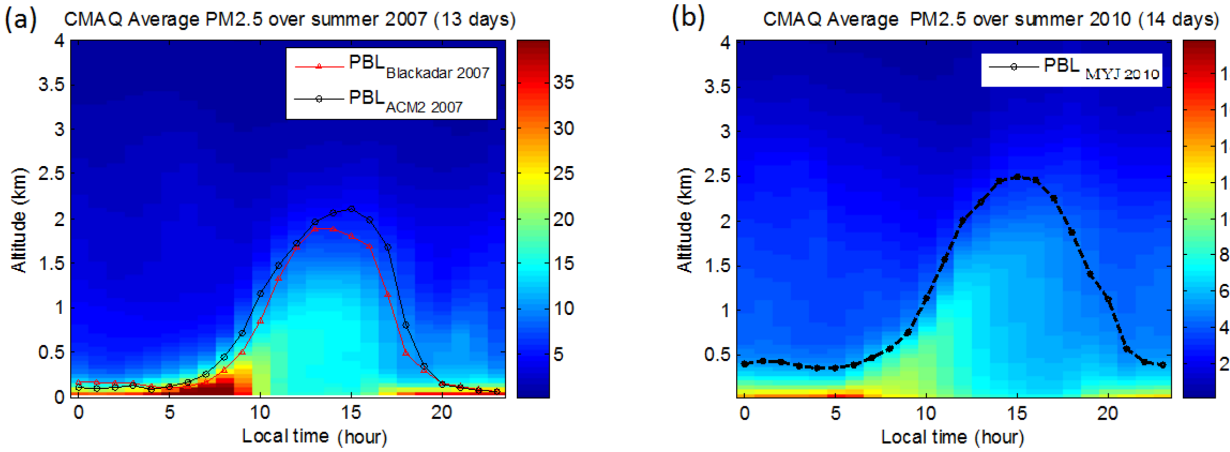


Figure 5.20: The 3-Dimensional structure of the CMAQ predicted PM2.5 mass concentration (a) Summer 2007 (b) Summer 2010.

CHAPTER 6

CMAQ Aerosol Optical Properties Analysis

In this chapter, we mainly focus on the study of aerosol optical properties and modeling. We attempt to construct a general physical – optical converter, which would ultimately provide the means to connect physical properties of the particulates into optical properties. These results can be then used to test the accuracy of the model by using remote sensing instruments. Preliminary efforts show reasonable agreement among the column AOD (extensive property) but less agreement is found for intensive properties such as the spectral ratio of extinction (e.g. angstrom coefficient). In addition, the backscatter coefficients obtained from the calculation are compared against the lidar observation to better observe the vertical discrepancies in the CMAQ model.

6.1 Overview

The CMAQ model uses a multimode aerosol model (see Chapter 1) to quantify aerosol dynamics in the atmosphere. The CMAQ model approach calculates three integral properties of the size distribution for each mode (e.g. i-mode is Aitken, j-mode is accumulation and c-mode is coarse mode): the total particle number concentration, the total surface area concentration and the total mass concentration of the individual chemical components which is sufficient to quantify the log-normal description of the model.

One of the CMAQ outputs is the aerosol extinction coefficient (500 nm) which is approximated in the same manner as that adopted by the AIRNow program (Malm, 1994). In this regression based approach, scattering properties are connected to the speciation of the particulates according to the following empirical relation based on nephelometer studies connecting particulate speciation to extinction which is given by

$$\alpha_{ext} = 0.003 * f(RH) * (\text{ammonium sulfate+ ammonium nitrate}) + 0.004*OM + 0.01*EC + 0.001*FS + 0.0006*CM; \quad (6.1)$$

where ammonium sulfate and ammonium nitrate can be taken as the sum of ammonium, plus sulfate, plus nitrate $\left(\left[NH_4^+\right]+\left[SO_4^{2-}\right]+\left[NO_3\right]\right)$ within the CMAQ model. This is represented by ASO4I_PM25, ASO4J_PM25, ANO3I_PM25, ANO3J_PM25, ANH4I_PM25, and ANH4J_PM25. The light absorbing carbon (EC) was taken as elemental carbon, available in the CMAQ output as AECI_PM25. Organic mass (OM) was taken as the sum of all organic species and is available in the sized resolved CMAQ quantities as the sum of AORGPAI_PM25 and AORGPAJ_PM25. The fine soil (FS) was taken as the unspecified anthropogenic mass A25J_PM25 and A25I_PM25. CM is represented by the coarse mode. In addition, to account for RH effects which bloat the aerosol and enhance the scattering, empirical growth curves as a function of RH are used (see example Fig. 6.1) (Malm et al., 2001). In our model, only the sulfate f(RH) function is used.

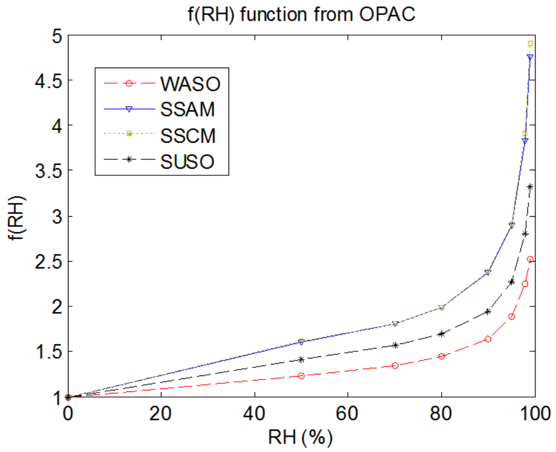


Figure 6.1: f(RH) from OPAC data is a function of relative humidity for different type of aerosol. (WASO=water soluble; SSAM=sea salt; SSCM=sea salt coarse; SUSO=sulfate)

It is expected that such an empirical based method may not perform well under a wide variety of conditions. Furthermore, to explore more complex spectral signatures such as extinction color ratio or the lidar backscatter, a full Mie scattering based treatment is needed. The procedures of this approach are briefly outlined in the following section and more detail can be found in (Binkowski & Rosselle, 2003; Jiang et al., 2006).

6.2 Calculating Aerosol Extinction Procedures

First, let $[M^s{}_m]$ be the PM mass concentration in μg/m^3 for size mode (*m*) and chemical species (*s*). We need to convert mass to volume using equation below

$$[V^s{}_m] = \frac{10^{-12}}{\rho_s}[M^s{}_m] \tag{6.2}$$

where ρ_s is the mass density of the *s*th species in g/cm^3 (Table 6.1). The unit of particulate volume ($V^s{}_m$) is m^3 per unit volume (m^3 / m^3).

Step 2, given the log-normal size distribution,

$$n(\ln D) = \frac{N}{\sqrt{2\pi}\ln\sigma_g}\exp\left[-\frac{1}{2}\left(\frac{\ln D - \ln D_g}{\ln\sigma_g}\right)^2\right] \tag{6.3}$$

where N is the particle number concentration (M_0) within the mode, D is the particle diameter and D_g and σ_g are the geometric mean diameter and geometric standard deviation of the modal distribution, respectively. The *k*th moment of the distribution is defined as

$$M_k = ND_g^k \exp[\frac{k^2}{2}\ln^2\sigma_g]. \tag{6.4}$$

M_0 is the number N of aerosol particles within the mode, suspended in a unit volume of air. For $k = 2$, the moment is proportional to the total particulate surface area within the mode, per unit volume of air. For $k = 3$, the moment is proportional to the total particulate volume within the mode, per unit volume of air. The constant of proportionality between M_2 and surface area is π; the constant of proportionality between M_3 and volume is $\pi/6$. Keep in mind that the geometric standard deviation is the same regardless which moment is selected. M_3 is determined as follows from the fine aerosol species (see Table 6.1):

$$M_{3i} = \sum_{n=1}^{n_{\max}} \frac{\varphi_{i,n}}{\frac{\pi}{6}\rho_n} \qquad or \qquad M_{3j} = \sum_{n=1}^{n_{\max}} \frac{\varphi_{j,n}}{\frac{\pi}{6}\rho_n} \tag{6.5}$$

where $\varphi_{i,n}$ and $\varphi_{j,n}$ are the species mass concentrations [μg m^{-3}], of the nth species in each mode; ρ_n is the average bulk density of the nth species. The third moment for the coarse mode is obtained in a similar way. After obtain the values of number, second and third moments concentrations, the geometric mean diameter (D_g) and geometric mean standard deviation (σ_g) can be calculated from

$$\ln^2 \sigma_g = \frac{1}{3}\ln(M_0) - \ln(M_2) + \frac{2}{3}\ln(M_3) \tag{6.6}$$

$$D_g^3 = \frac{M_3}{N \exp[\frac{9}{2}\ln^2(\sigma_g)]} \tag{6.7}$$

where $M_0 = N$, $M_2 = N\ D_g^2 \exp[2 \ln^2(\sigma_g)]$, and $M_3 = N\ D_g^3\ exp[9/2\ ln^2(\sigma_g)]$. The standard deviation for the coarse mode is constant at 2.2 μm (CMAQ internal assumption). The moments are calculated by

$$M_{0k} = N_k \tag{6.8a}$$

$$M_{2k} = \frac{S_k}{\pi} \tag{6.8b}$$

$$M_{3k} = \frac{v_k}{\frac{\pi}{6}} \tag{6.8c}$$

where for the *k*th mode particles, S_k is the surface area concentration, N_k is the particle number concentration and V_k is the volume concentration which they are available in the CMAQ outputs. If the number distribution $n_N(D)$ is log-normal, the volume distribution $n_V(D)$ is also log-normal with the same geometric standard deviation σ_g (Seinfeld and Pandis, 1998). Given the dry state log-normal size distribution parameters for number density, D_g and $ln\sigma_g$, we convert from number to volume as

$$\ln D_{gv} = \ln D_g + 3\ln^2\sigma_g \rightarrow D_{gv} = D_g \exp[3\ln^2\sigma_g] \tag{6.9a}$$

$$\ln\sigma_{gv} = \ln\sigma_g \tag{6.9b}$$

Once we have the individual volume diameter and width, the volume for each mode and class is obtained by simply summing the volume over all species in each aerosol class (o) (e.g. soluble, insoluble, sea salt, and soil) as V_m^o. In addition, the reflective index for each species is given in OPAC database (see Table 6.2) but the reflective index in the calculation is weighted by the volume concentration of each aerosol class. The effective refractive index is defined as

$$n_{eff,l} = \frac{\sum_o V_{l,o} n_{l,o}}{\sum_o V_{l,o}} \tag{6.10}$$

where V_i is the volume concentration and n_i is the refractive index for different classes. For example, the standard “Mie Extinction’ product in CMAQ attempts to create effective refractive index based on volume weighting over the different aerosol classes where water particles are also counted as a class. In using the volume (e.g. 3rd moment), water volume is also included and is propagated into calculating the effective Mie parameters (geometric volume diameter, log-sigma). In this treatment, no growing factor is applied since the volume is corrected directly and the moments adjusted.

The aerosol extinction coefficient β_{sp} [km^{-1}] for each aerosol class *(o)* and each size mode *(l)* must be calculated from ambient aerosol characteristics as index of refraction *(m = n – i k)*, volume concentration and size distribution; at wavelength λ, β_{sp} can be expressed as

$$\beta_{sp}^{l,o} = \frac{3\pi}{2\lambda}\int_{-\infty}^{\infty}\frac{Q_{ext}^{o}}{\alpha}\frac{dV_{l,o}}{d\ln\alpha}d\ln\alpha \qquad (6.11)$$

where the lognormal form of particle distribution is given as

$$\frac{dV}{d\ln\alpha} = V_T(\frac{A}{\pi})^{1/2}\exp[-A\ln^2(\frac{\alpha}{\alpha_v})] \qquad (6.12)$$

and $\alpha = \frac{\pi D}{\lambda}$, $\alpha_v = \frac{\pi D_g}{\lambda}$, and $A = \frac{1}{2\ln^2\sigma_g}$. Here, V_T is the total particle volume concentration, and Q_{ext}, the Mie extinction efficiency factor, is a function of α and the index of refraction of the particles. Q_{ext} can be calculated using Mie theorem approximation. The extinction coefficient is a function of the geometric mean diameter for the volume distribution and the complex refractive index. Now, we can obtain aerosol extinction in each vertical layer by summing all size modes *(l)* and aerosol classes *(o)* as the following

$$\beta_{sp} = \sum_{l}\sum_{o}\beta_{sp}^{l,o} \qquad (6.13)$$

Lastly, we integrate all layers to get total aerosol optical depth (AOD). Moreover, we assess the results with the angstrom coefficients which are obtained from AERONET. Since we have the extinction coefficients from our model, the angstrom coefficient can be calculated as

$$\gamma = \frac{-\log\left(\frac{\alpha_{\lambda 2}}{\alpha_{\lambda 1}}\right)}{\log\left(\frac{\lambda 1}{\lambda 2}\right)} \qquad (6.14)$$

where a_{λ} is the extinction coefficient which is a function of selected wavelength (λ).

Table 6.1 Sized-resolved CMAQ PMx Quantities

	Variable	**Description**
Geometric mean diameter (µm)	D_{gi}	i-mode geometric mean diameter
	D_{gj}	j-mode geometric mean diameter
	Dgc	c-mode geometric mean diameter
Geometric standard deviation	σ_{gi}	i-mode geometric standard deviation
	σ_{gj}	j-mode geometric standard deviation
	σ_{gc}	c-mode geometric standard deviation
PMx mass concentration (µg m^{-3})	MassConc	Total PMx mass concentration of all species in all modes
	ASO4I	i-mode sulphate mass concentration
	ANO3I	i-mode nitrate mass concentration
	ANH4I	i-mode ammonium mass concentration
	AECI	i-mode elemental carbon mass concentration
	AORGPAI	i-mode primary organic aerosols mass concentration
	AORGAI	i-mode anthropogenic secondary organic aerosols mass concentration
	AORGBI	i-mode biogenic secondary organic aerosol mass concentration
	AH2OI	i-mode water mass concentration
	A25I	i-mode other primary PM2.5 mass concentration
	ASO4J	j-mode sulphate mass concentration
	ANO3J	j-mode nitrate mass concentration
	ANH4J	j-mode ammonium mass concentration
	AECJ	j-mode elemental carbon mass concentration
	AORGPAJ	j-mode primary organic aerosols mass concentration
	AORGAJ	j-mode anthropogenic secondary organic aerosols mass concentration
	AORGBJ	j-mode biogenic secondary organic aerosol mass concentration
	AH2OJ	j-mode water mass concentration
	A25J	j-mode other primary PM2.5 mass concentration
	ASEAS	marine particle mass concentration
	ASOIL	soil-derived particle mass concentration
	ACORS	other coarse particle mass concentration

PMx number concentration (particle/m^{-3})	NumConc NUMATKN NUMACC NUMCOR	Total PMx number concentration of particle in all modes i-mode particle number concentration j-mode particle number concentration c-mode particle number concentration
PMx surface area concentration (µm^2/m^3)	SRFATKN SRFACC	i-mode particle surface area concentration j-mode particle surface area concentration

Table 6.2 CMAQ Aerosol species and physical properties

Species	**Classes (OPAC)**	**Mass Density**	**Dry state refractive index λ=500 nm**	
		g/cm^3	Real	imaginary
'ASO4J_PM2.5'	Water soluble	1.8	1.430E+00	-1.000E-08
'ASO4I_PM2.5'	Water soluble	1.8	1.430E+00	-1.000E-08
'ANH4J_PM2.5'	Water soluble	1.8	1.530E+00	-5.000E-03
'ANH4I_PM2.5'	Water soluble	1.8	1.530E+00	-5.000E-03
'ANO3J_PM2.5'	Water soluble	1.8	1.530E+00	-5.000E-03
'ANO3I_PM2.5'	Water soluble	1.8	1.530E+00	-5.000E-03
'AORGAJ_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03
'AORGAI_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03
'AORGPAJ_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03
'AORGPAI_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03
'AORGBJ_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03
'AORGBI_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03
'AECJ_PM2.5'	soot	2.2	1.750E+00	-4.500E-01
'AECI_PM2.5'	soot	2.2	1.750E+00	-4.500E-01
'A25J_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03
'A25I_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03
'ACORS_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03
'ASEAS_PM2.5'	Seasalt	2.2	1.500E+00	-1.550E-08
'ASOIL_PM2.5'	Insoluble	2	1.530E+00	-8.000E-03

6.3 Result and Discussion

After calculating the extinction based on the CMAQ outputs using the Mie Scattering approach, we compare the results of Mie Scattering approach against the CMAQ extinctions (Malm approximation) and AERONET AOD measurements (see Section 2.3 for full instrumentation description). Two cases (Aug 31 and Sep 1 of 2010) are selected to present the finding of this comparison. In particular, we find that the calculated AOD match better with the AERONET measurements than CMAQ generated AOD using the Malm approach in both cases (see Fig. 6.2). At the same time, the angstrom coefficients ($\gamma \approx 1$) obtained based on our calculation is significantly smaller than the angstrom coefficients ($\gamma \approx 1.9$) from AERONET (see Fig. 6.3). The model underestimating the angstrom coefficient seems to indicate that the ultrafine component may be significantly underestimated. This would of course bias the total AOD but less significantly. Also, the model refractive indices can be a significant issue.

In addition, we plot in Fig. 6.4 (a-b), the calculated backscatter (km^{-1}) as 3D image so that we can better observe the derived simulated backscatter fields in both time and vertical direction. We first note that the absolute backscatter magnitudes between the model and the calibrated lidar measurements are generally in good agreement. In addition, we note similar qualitative signatures between these measurements such as enhancement of aerosol backscatter in morning that degrades by afternoon. On the other hand, the CMAQ backscatter (and associated PBL height) is weaker than the lidar backscatter at the end of the afternoon (16:00-18:00 EST) which is probably due to the model not accounting for urban heating effects properly. Note that the lidar backscatter is not well calibrated during 15:00-17:00 EST due to the cloud at low altitude (≈2.5 km).

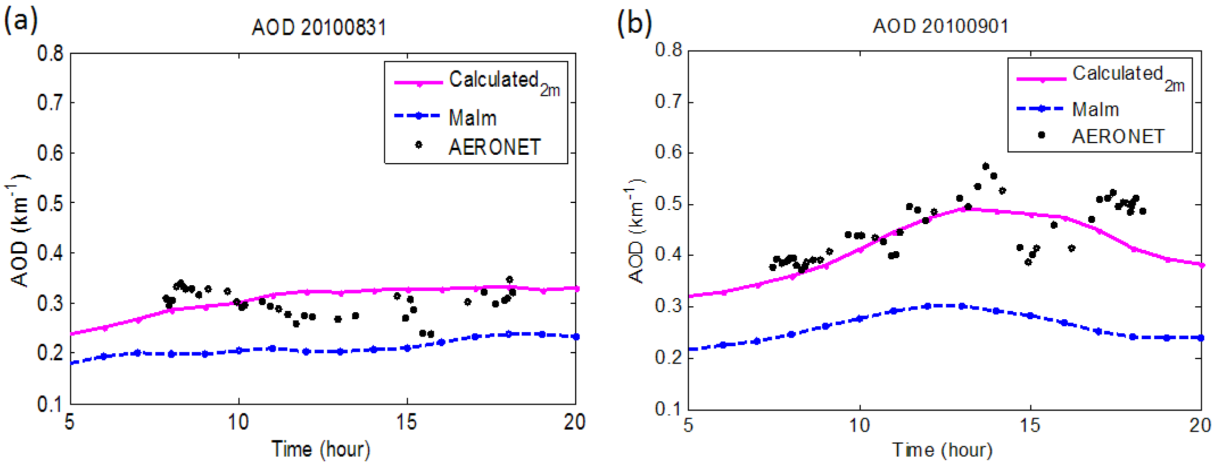


Figure 6.2: AOD (wavelength = 500 nm) comparisons for calculations (2m represent sum of i- and j-mode), model outputs and AERONET data on (a) Aug 31, 2010 and (b) Sep 1, 2010.

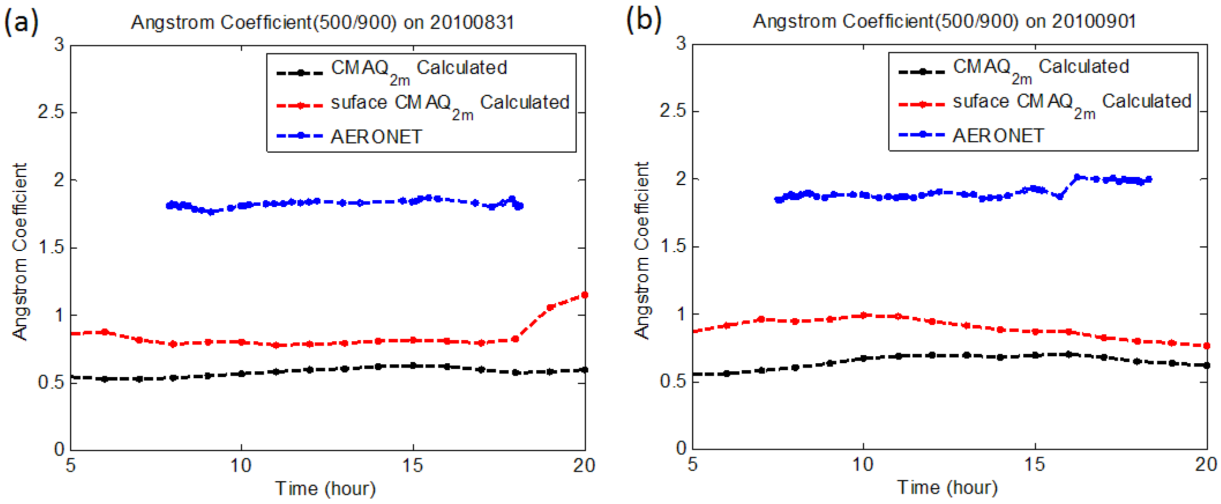


Figure 6.3: Angstrom coefficient (unitless) comparisons for calculated result (CMAQ$_{2m}$ = total column; surface CMAQ$_{2m}$ = ground to 250 m) and AERONET data on (a) Aug 31, 2010 and (b) Sep 1, 2010.

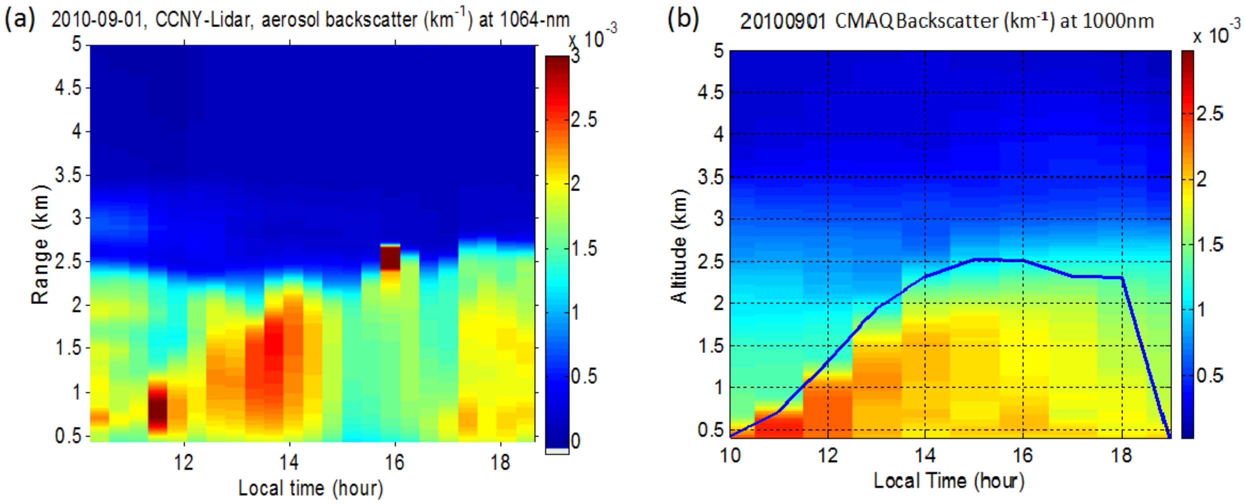


Figure 6.4: September 1, 2010 measurements of (a) Lidar (1064 nm) backscatter and (b) CMAQ calculated backscatter (1000 nm) in unit km^{-1}.

CHAPTER 7

Combining VIS and MIR Lidars to Separate Aerosol Modes in the Troposphere

This chapter discusses and outlines some possible MIR (together with further improvements to be expected in near term) based schemes which may be capable of aerosol detection to 800 meters with present systems. The performance of MIR lidar SNR based on possible components available in the market is shown. The application of separating fine and coarse aerosol modes is discussed at the end of the chapter.

7.1 Overview

Indentifying and quantifying ambient aerosols are important for air quality applications. Unlike trace gases where chemical spectral signatures are sharp and well defined, aerosol spectral signatures are broader and highly overlapping. Thus, separation of aerosols into different size classes requires very broad spectral coverage from the visible (VIS) to mid-infrared (MIR). In this chapter, we investigate the feasibility in using a VIS (0.65 µm) diode laser combined with a suitable pulsed high power Quantum Cascade Laser (QCL) (4.6 µm) to obtain backscatter measurements that can be used to isolate fine and coarse mode aerosol fractions (Gan et al, 2010b). Based on realistic source characteristics, we study the information content in the spectral extinction using different combinations of extinction measurements. We apply least squares minimization to a wide range of aerosol multimode mixtures which is obtained using realistic models obtained from the OPAC database (Hess et al., 1998). This database is very useful as the optical extinction spectra and backscatter spectra are enclosed over a wide range of wavelength from 250 nm to 40 µm. In particular, the latest QCL system is able to achieve SNR values ≈10 with suitable temporal and spatial averaging for aerosol layers (at distance of ≈1.5 km), which make it possible for PBL layer studies.

7.2 Signal to Noise Ratio (SNR)

Before calculating the SNR, we need to find receive signal, detector noise, pre-amplifier noise and digitizer quantization noise for the MIR system (Corrigan et al., 2009). First, the received power can be determined by using lidar equation as summarized below:

$$P_{rec} = P_L \beta_{target} T dR \frac{A}{R^2} \tag{7.1}$$

where P_L is the average power emitted from the laser source (W), β_{target} is the backscattering coefficient (m^{-1}), T is the two way atmospheric transmission, dR is the effective pulse length (m), A is the receive area (m^2) and R is the round trip distance (m) between target and system (assuming the transmitter and receiver are located at the same place). Therefore, the signal voltage can be found as:

$$V_{pulse} = P_{rec} K_{\det} G \tag{7.2}$$

where K_{det} is the Responsivity (V/W) and G is the gain of the pre-amplifier. The system noise is determined by adding the noise voltage of each component in the detection scheme. The detector noise and pre-amplifier noise can be estimated by using Eq. (7.3) and (7.4).

$$V_{N,\det} = NEP\, K_{\det}\, G \tag{7.3}$$

$$V_{N,TIA} = GN\sqrt{BWP} \tag{7.4}$$

where *NEP* is the noise equivalent power of the detector (W), *N* is the pre-amp noise given by the manufacture (V/Hz$^{1/2}$) and *BWP* is the bandwidth of the pre-amp. Because $V_{N,DAQ}$ is very small in the system so it can be neglected in the calculation. Lastly, the *SNR* per pulse is calculated for lidar system by including all noise as below:-

$$SNR_{pulse} = \frac{V_{pulse}}{\sqrt{V_{N,\det}^2 + V_{N,TIA}^2 + V_{N,DAQ}^2}} \tag{7.5}$$

Manufacturer specifications for the possible choices of detector, amplifier and digitizer in commercially available products are listed in Table 7.1 and 7.2. Once the SNR for a single pulse is calculated, the SNR after pulse averaging can be estimated as

$$SNR(n) = \sqrt{n} SNR_{pulse} \tag{7.6}$$

where *n* is the number of samples or pulses. The spatial and time resolution are selected base on minimum *SNR* requirement (>1) for a given range.

7.3 Experimental Method

We propose a lidar transmitter and receiver arrangement as shown in Fig. 7.1. A pulsed laser beam is sent into the atmosphere, hits air molecules and other particles, and is scattered by them. The scattered light is sent into various directions and only a small part is reflected back towards its origin. The backscattered light is then collected by telescopes and conducted into a detection unit. After passing through filters, the signal is received by a detector, converted into electronic signals, and recorded with temporal resolution. Maximum distance that the system can reach is limited by the strength of the laser. However, there is another concern that we have to consider before we build the system. First, the beam divergence of the laser should be smaller than the telescope and detector acceptance angle so that all power is accounted. For instance, the QCL beam divergence is about 0.5 mrad, which should be smaller than the acceptance angle (0.7 mrad) of the detector which is limited by the focal length of the telescope and the detector area (Fig. 7.2). The acceptance angle can be calculated using:

$$\theta = \tan^{-1}\frac{radius}{focal\ length} \tag{7.7}$$

Here, we propose a VIS-MIR lidar system with available components in the market. One of the reasons of using VIS channel is because it is useful for system calibration and alignment. The system parameters are listed in Table 7.1 and 7.2

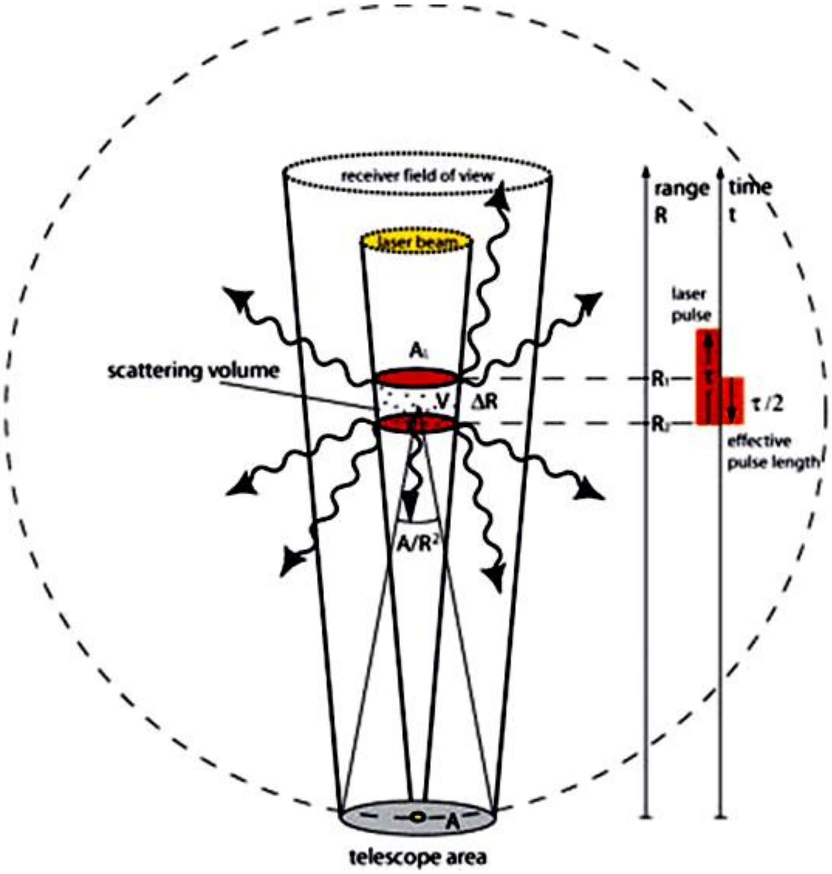


Figure 7.1: Illustration of the lidar geometry

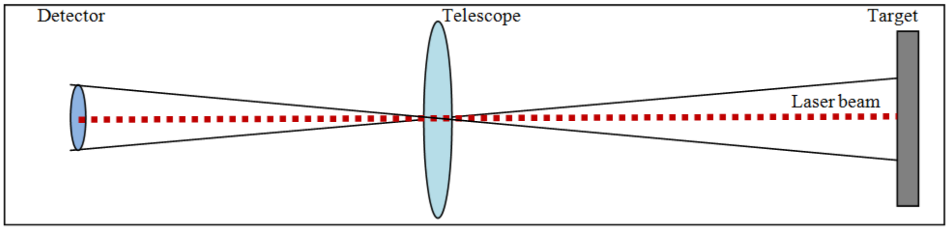


Figure 7.2: Field of view (FOV) of the telescope and detector

Table 7.1 Proposed MIR System Parameters

Component	**Parameter**	**Values**
QCL Laser (center at 4.6 µm) – manufactured by Pranalytica	Pulse Power (peak)	4.0 W
	Pulse length	50 ns – 500 ns
	Pulse rise/fall time	Less than 5 ns
	Spectral width	~150 nm
	Maximum pulse repetition rate	1 MHz
	Beam Divergence	0.5 mrad
Photovoltaic Detector (center at 5 µm) Thermoelectrically cooled – manufactured by Vigo System	Detectivity*	$2.5\mathrm{x}10^9$ $\mathrm{cmHz}^{1/2}/\mathrm{W}$
	Responsivity	>1.3 A/W
	Acceptance angle, F/#	36^o, 1.62
	Diameter	1 mm
	Rise time	~20 ns
	Resistance	11 Ω
Preamps – manufactured by Boston Electronics Corporation	Gain (R_D is detector resistance)	2 K/R_D
	Bandwidth	200 MHz
	Noise	1.8 $\mathrm{nV/Hz}^{1/2}$
	Transimpedance factor	2000

Telescope for both channels	Diameter	25.4 cm
	F/#	3
	FOV	49×10^{-9} sr
Digitizer for both channels – manufactured by Gage	Sampling rate (max)	200 MS/s
	Resolution	12 bits
	DC accuracy	±0.5%

Table 7.2 Proposed VIS System Parameters

Component	**Parameter**	**Values**
AlGalnP Laser Diode (center at 0.65 µm) –manufactured by Opnext	Output power	100 mW
	Maximum pulse repetition rate	20 MHz
	Beam divergence	<1 mrad
PMT with gate function – manufactured by Hamamatsu	Spectral Response	300 nm – 850 nm
	Diameter	8 mm
	Radiant sensitivity	50 mA/W
	Dark current	0.4 nA
	Rise time	0.78 ns
	Transit time	5.4 ns
	TTS	0.23 ns
	Resistance	50 Ω
Preamps – manufactured by Hamamatsu	Current to voltage conversion factor	10^5 V/A
	Frequency bandwidth	0-20 kHz
	Equivalent noise input (rms)	1.5 nA

7.4 Discussion and Result

In order to eveluate the proposed system, we extimate the backscatter response we would get for urban aerosol by utilizing OPAC database (Hess et al., 1998). Based on the magnitude of the urban signal at 50% RH, the backscatter values obtained (see Fig. 7.3) are used to calculate the SNR that can be expected for the MIR lidar systems whose specifications are defined Table 7.1.

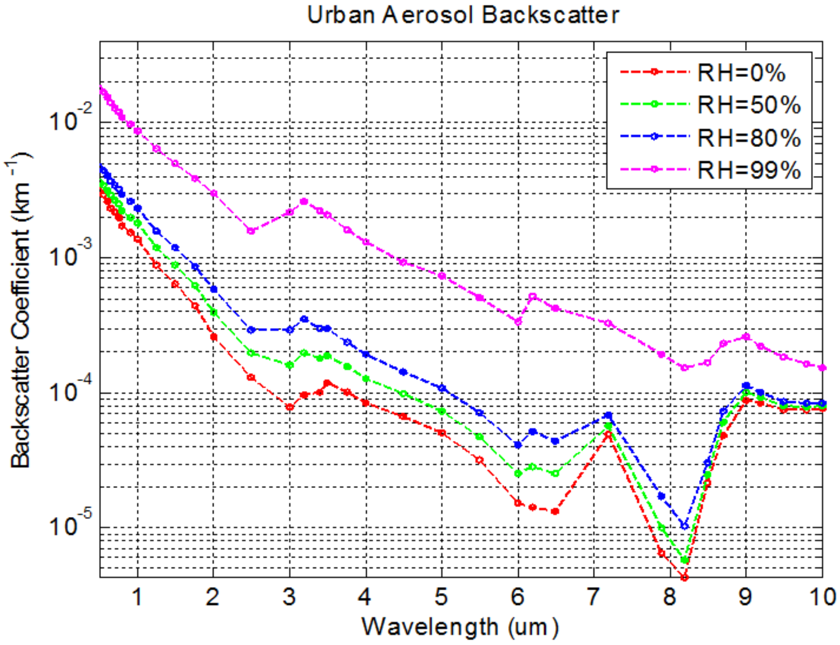


Figure 7.3: Urban aerosol backscatter from OPAC

It must be understood that because of the need to sample at far range so the spatial resolution must be no smaller than the pulse width. Thus, the repetition rate must be limited to 100 KHz which sets a limit of the average power that can be obtained. Currently, we use QCL with average power of 40 mW (pear power = 4 W) as our standard but in the graphs, we allow the power to vary up to 80 mW to account for rapid advances in the QCL technology. In Fig. 7.4

and 7.5, we plot the simulated SNRs of MIR and VIS lidar for different time and spatial averaging scenarios. For the MIR system, Fig. 7.4 (a) is the highest SNR we can achieved with spatial averaging 75 meters and time averaging 15 minutes. Mean while, the VIS lidar is a mutual system and its performance is much better than the MIR lidar as seen in Fig. 7.5 (a-b), so optimizing this system is not a concern. Fig. 7.4 (d) show that the SNR is slightly above 1 which indicates the minimum resolution requirement for the system has to be at least 30 meters and 10 minutes.

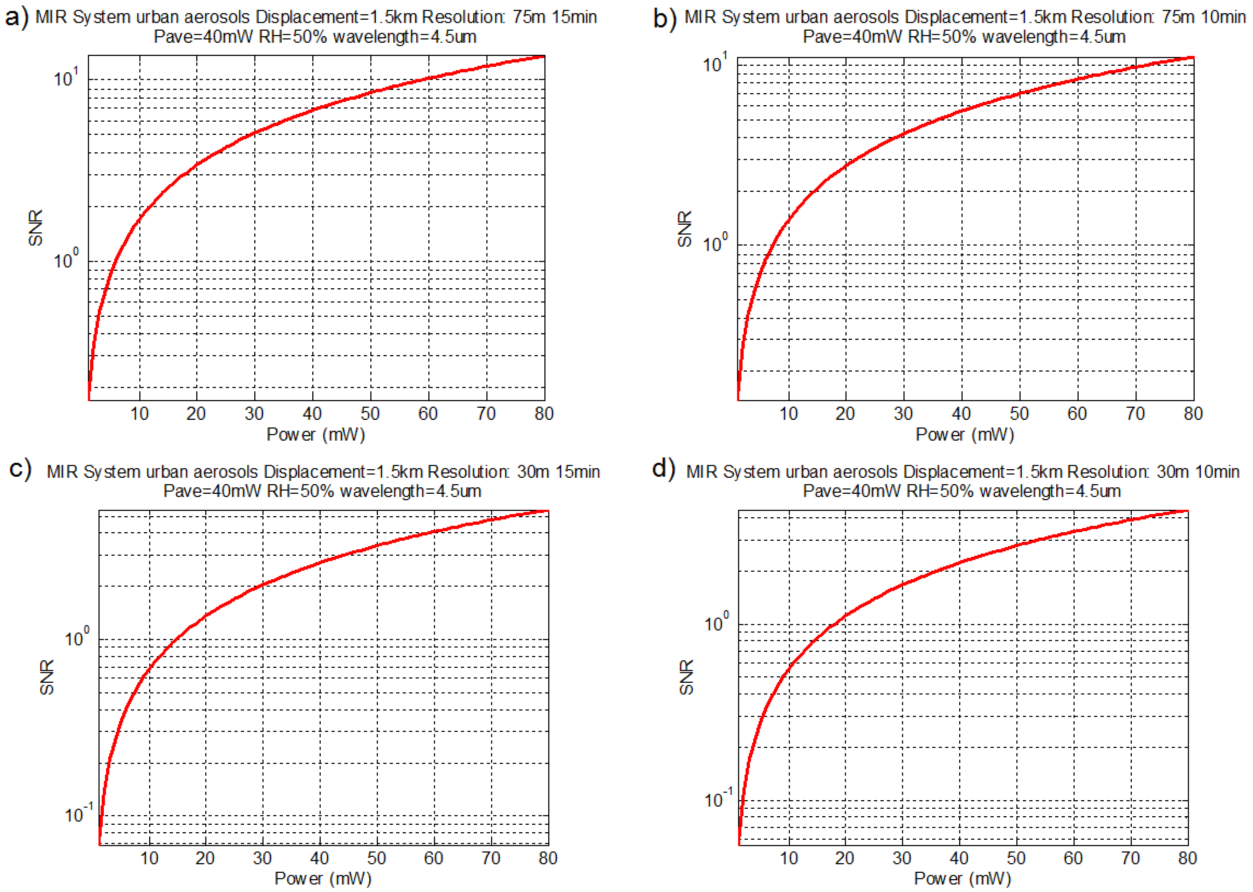


Figure 7.4: SNRs of MIR system for a target displacement 1.5km with different time and spatial averaging.

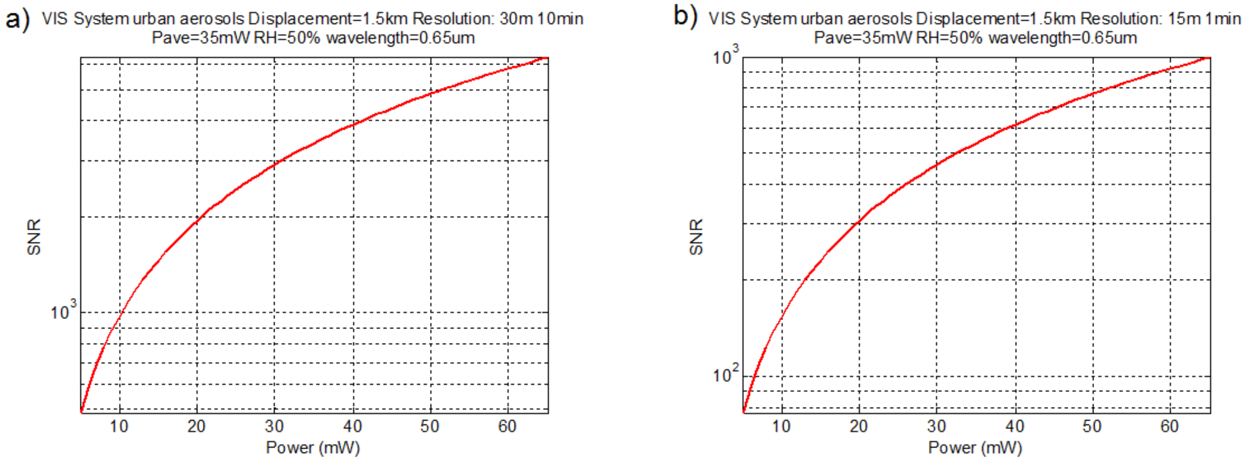


Figure 7.5: SNRs of VIS system for a target displacement 1.5km with different time and spatial averaging.

All these results only accounted for the detector / amplifier (e.g. thermal) noise which means the SNR may go below 1 in realistic condition due to the environment uncertainties and background noise. For a full treatment, we must also consider the possibility of stray skylight as well as thermal re-emission of the atmosphere back to the detector. We make use of a radiative transfer code (SBDART) to calculate the downwelling radiance for a vertically pointing detector (e.g. user zenith = 180^o) for both solar reflected skylight and atmosphere emission to the ground.

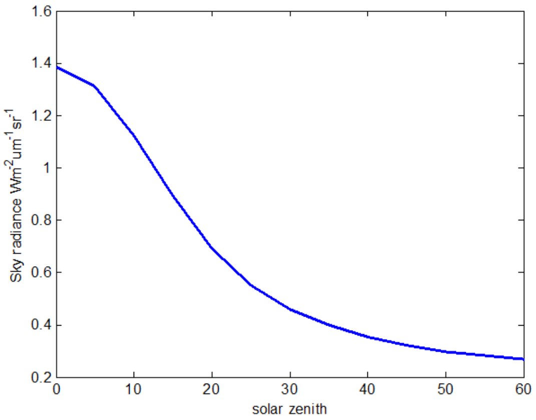


Figure 7.6: The radiance of the solar reflected signal calculated with SBDART

The sky radiance ($Wm^{-2}\mu m^{-1}sr^{-1}$) can be converted to an optical background signal on the receiver plane using

$$N_{sky} = R_{sky} \times A_{receiver} \times BW_f \times \Omega_r \tag{7.8}$$

where N_{sky} is background sky noise, R_{sky} is sky radiance, $A_{receiver}$ is receive surface area, BW_f is filter width and Ω_r is receiver solid angle.

Putting the numbers (assuming a realistic max. solar zenith angle of 30^o) into Eq. (7.8) leads to the result that per pulse power from the transmitter is much larger than the sky noise term and it can be ignored. This calculation of course emphasizes the need for high quality laser profiles (e.g. low divergence angles) which limited the detector FOV and cuts off the sky noise terms. Similarly, calculation using SBDART of the radiance of the thermal emission of the atmosphere leads to the value of 0.34 ($Wm^{-2}\mu m^{-1}sr^{-1}$) which is less than the solar term and can therefore be neglected as well. As we see from the above discussion, SNR values of 5-10 can be expected for the 4.6 µm channel.

7.5 Application – Multi-wavelength Backscatter Retrieval

Determining the structure of aerosols is important in separating anthropogenic sources (usually identified with small particle sizes) from natural aerosols such as desert dust, sea salt etc. Obviously, the use of multi-wavelengths should provide a natural means in separating fine and coarse mode contributions. To test this approach, we made use of sunphotometer measurements from AERONET which provided representative microphysical size distributions and refractive indices. Unfortunately, the refractive index is only available to 1 µm in AERONET, so we needed to extrapolate to the 4.6 µm channel but we do not expect significant differences in doing this. The procedure for obtaining the backscatter color ratio from the AERONET microphysical parameters is outlined below:

1. Using only the spherical mode, we can obtain the volume (*V*) distribution and the complex refractive index (*m*) from the sunphotometer observation such as radius (*r*).

   $$\left[\frac{dV}{d\log(r)}\right],\ \left(m_r(\lambda), m_i(\lambda)\right) \tag{7.9}$$

2. Then calculate the optical backscatter coefficients for each distribution in the AERONET archive using the Mie theory:

   $$\beta L(\lambda) = \int_{r_{\min}}^{r_{\max}} \left(\frac{3}{4r}\right) Q_\beta\left(q = \frac{2\pi r}{\lambda}, m_r(\lambda), m_i(\lambda)\right)\left[\frac{dV}{d\log(r)}\right] d\left(\log(r)\right) \tag{7.10}$$

   where β is the backscatter, L(λ) is Lidar ratio, Q_β is backscatter coefficient, q is particle size parameter and λ is wavelength.

3. Meanwhile, we can also obtain as a retrieval product, the physical fine mode volume and coarse mode volume.

In this direction, we can observe how the backscatter color ratio correlates to the fine – coarse volumes. To begin, we look at the backscatter ratio using 0.6 and 1.5 µm. Based on Fig. 7.7 (a), the correlation is not impressive and can be explained by the relatively small difference in the wavelengths. On the other hand, if we use the backscatter at 4.6 µm together with 0.6 µm, the correlation coefficient is much improved going from 0.44 to 0.75 as shown in Fig. 7.7 (b). Besides, we also calculate the correlation coefficient for extinction ratio to fine – coarse volumes at same wavelengths. Fig. 7.8 (b) shows that it is better than the correlation coefficient obtained by backscatter color ratio.

The results however must be modified to account for noise uncertainties. Based on the SNR results for the 0.6 and 4.6 μm channels, realistic backscatter uncertainties of 10% at 0.6 μm (due to errors in the calibration using the Rayleigh atmosphere) and 20% at 4.6 μm are used.

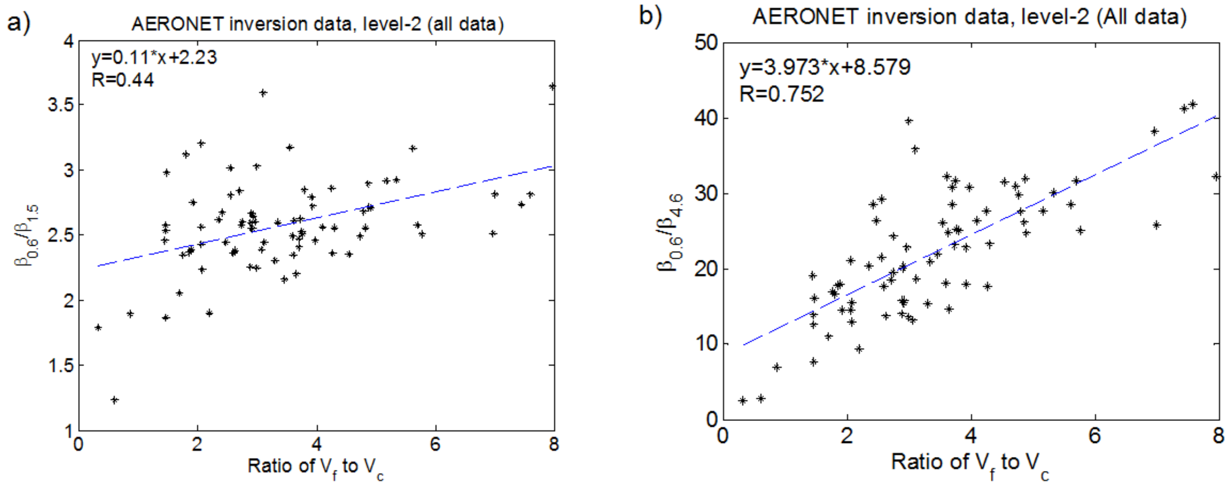


Figure 7.7: Backscatter color ratio and fine-coarse mode volume ratio of aerosol derived from AERONET inversion data.

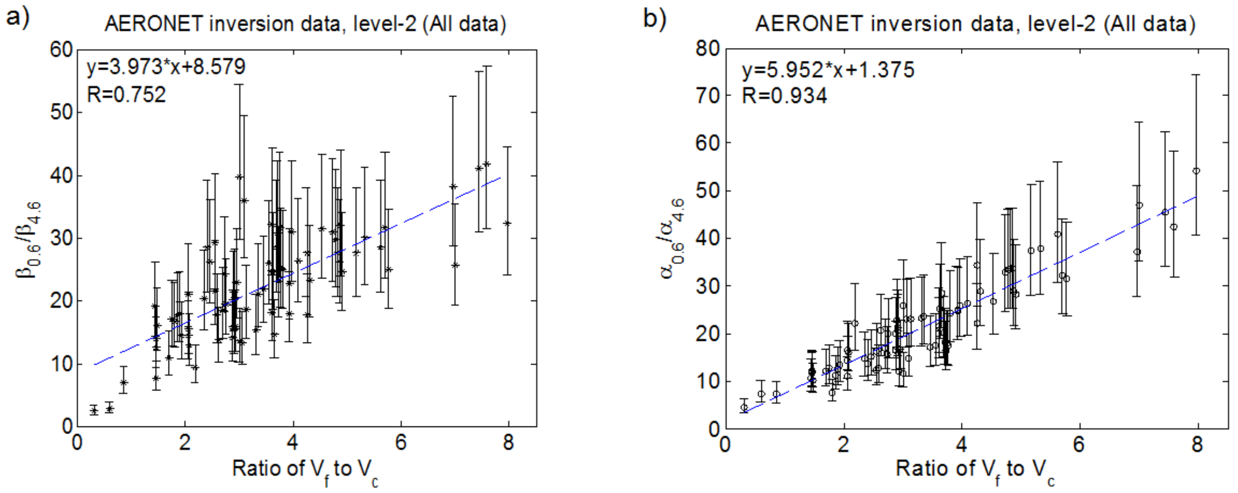
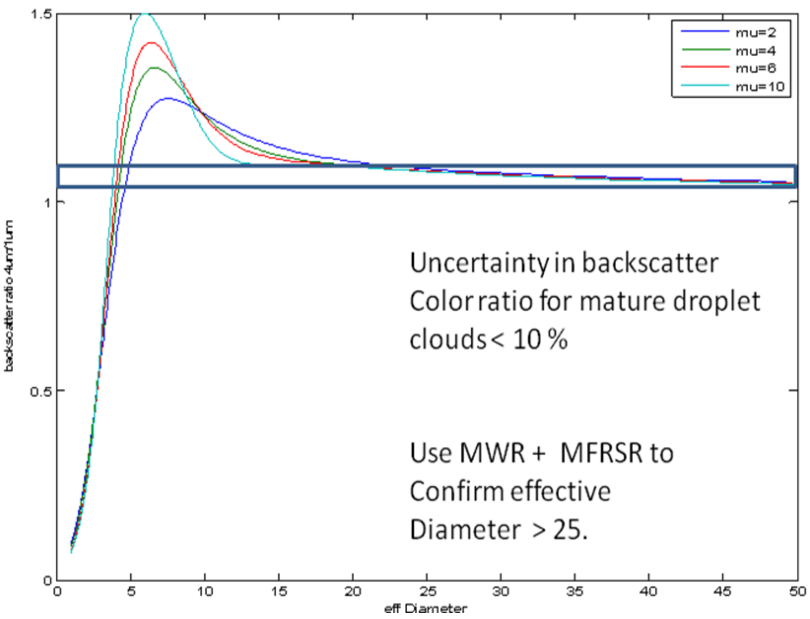


Figure 7.8: Backscatter color ratio and extinction ratio versus fine-coarse mode volume ratio of aerosol.

7.6 Remarks

The results obtained above illustrate the strong sensitivity of this type of measurement since backscatter is much more sensitive to refractive index variability. However, further improvements in the 4.6 µm channel and calibration should improve the results. It is important to point out that if we do the same exercise for the extinction, we see much less sensitivity to uncertainties. Getting the extinction at each channel can be performed using the horizontal path reflection geometry. As an added bonus, in this geometry, the achievable SNR can be expected to be significantly improved although no ranging of the aerosol is obtained. To conclude, the MIR lidar system can, in principle, be used together with either a VIS (or NIR) to better diagnose fine / coarse mode aerosols both in lidar mode and as a path attenuation using retro-reflectors. In processing the lidar, an absolute calibration is needed. For a NIR system, we have demonstrated that using suitable cloud techniques, calibration of the NIR laser source at 1064 nm is possible to errors < 10% (Wu et al., 2010) using low altitude water vapor clouds. In looking to use the backscatter to calibrate at MIR, we plot in Fig. 7.9 the backscatter ratio between 1.064 and 4.6 µm. The important observation is that the ratio uncertainty can be less than 10% if sufficiently mature cloud drops > 25 µm are used. Validating these mature clouds is possible using our combined Multifilter Shadowband Radiometer (MFRSR) / MWR system. In particular, measuring the diffuse transmittance for thick clouds provides an estimate of the cloud optical depth while the MWR gives the liquid water path. These 2 measurements can be combined to give the effective droplet size (Min et al., 2003). An example is given Fig. 7.10 for very recent data and illustrates it should be quite simple to differentiate clouds with sufficient droplet size.



Uncertainty in backscatter
Color ratio for mature droplet
clouds < 10 %

Use MWR + MFRSR to
Confirm effective
Diameter > 25.

Figure 7.9: MIR/MIR Backscatter color ratio showing small deviations for mature cloud drops.

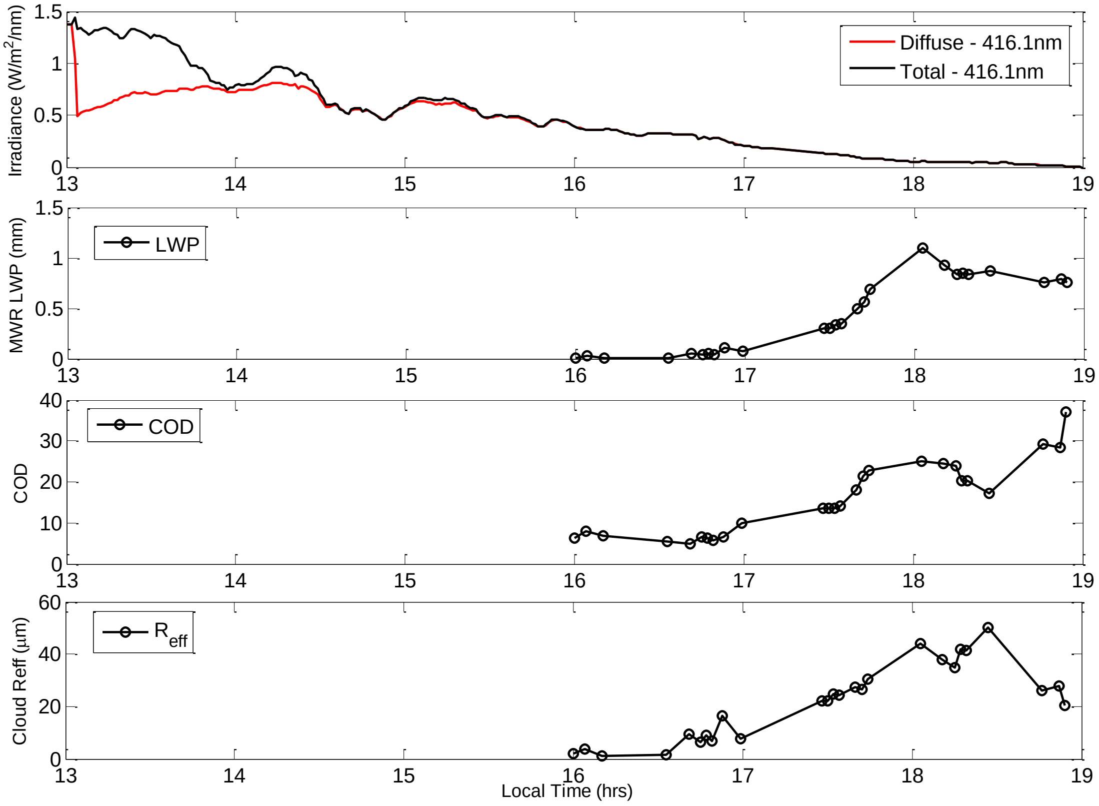


Figure 7.10: MFRSR / MWR retrievals. (a) Total / Diffuse transmittance (b) MWR Liquid Water Path (c) Cloud Optical Depth (COD) (d) Effective Cloud Radius.

Finally, the backscatter off clouds can also be obtained with the QCL system and may have applications in combination with VIS-NIR measurements to obtain droplet size in water phase clouds (Wu et al., 2009)

CHAPTER 8

Summary and Conclusions

The main findings of this thesis are summarized as follows:-

1. In Chapter 3, we proposed simple modifications to the standard wavelet covariance transform (WCT) based technique which allowed us to better isolate the PBL from plume and residual layers. These modifications included clouds screening, utilizing of height-dependent dilation and translation, and checking of continuality of the PBL. This modified algorithm gives us the flexibility to accommodate various atmosphere conditions while remain accurate in retrieving PBL height.

2. The usefulness in this approach (WCT) was illustrated in Chapter 5 by statistical comparisons of lidar derived PBL heights against WRF derived PBL height retrievals. In particular, other methods such as the maximum variance technique showed poor correlations to the PBL height during the period of transition (e.g. growing or decaying processes). This was a symptom of the technique being fooled by residual layer as well as sub-layer structure within the PBL.

3. Using the modified WCT technique, we found that for the hindcast runs, BLK has slightly better correlation (R=0.87) than the ACM2 scheme (R=0.84) (see Table 3.2 and Fig. 5.4). As seen in the histogram analysis of Fig. 5.5, the mean bias of the BLK scheme (m=190 meters) is smaller than that observed in the ACM2 scheme (m=340 meters). On the other hand, the fluctuation of the deviation is higher for BLK scheme which is about σ_{BLK}=300 meters while

the ACM2 schemes standard deviation is about σ_{ACM2}=280 meters. However, based on realistic estimates of systemic biases, the observed biases are reasonable within the errors that might be expected. Still, for ACM2 scheme, the overbias was more significant so we tentatively conclude that the BLK scheme was slightly better. In forecast mode, we applied the MYJ scheme; the correlation was surprisingly high although the dispersion was about 70% higher which is to be expected. The above results show good performance for 12 km grid simulations with simple modeling of the urban land surface. Most important, while we may not have expected good matches to begin with; our results do seem to illustrate quite impressive agreement with high linear correlations (R). However, there are statistically significant $(\Delta H \sim \sigma(H))$ overbiases observed in some of the model outputs (especially ACM2) which still needs further study This finding shows however that the WRF model at 12 km resolution can be used to forecast the main convective features of the PBL.

4. We also provided strong evidence that the anomalous spikes of PM 2.5 obtained in the summer during predawn and post-sunset are due more to the vertical distribution of pollutants within the CMAQ model than to errors in emission inventory. In particular, we found that the vertical structures of the pollutants emissions were strongly confined near the surface, in comparison to the vertical structure seen by the direct observations of ceilometer. These observations imply that the major cause of the overestimation is because of the model confinement of the pollutants emissions near the surface. To support this remark that the vertical structures of the pollutants emissions is the key factor; we showed that the path integrated CMAQ PM2.5 has better correlation to the TEOM PM2.5 than the near surface CMAQ PM2.5 particularly in the pre-dawn period.

5. In addition, we outline the procedures of calculating the extinctions coefficients from CMAQ output based on Mie Scattering technique. This effort was to make it possible to connect CMAQ outputs to optically measureable parameters. We utilized data from AERONET to assess this model and compare it to the semi-empirical approach (Malm et al., 1994). We see reasonably good agreements among the Mie Scattering model output in comparison to the Malm approach for the total column AOD. Calculations for intensive properties such as the angstrom coefficient however were less successful in reproducing the measurements and it is expected that the particle size and the refractive index model have to be more realistic than what is being considered at present.

6. Finally, we reported on some possible methods to integrate MIR remote sensing measurements to improve aerosol retrieval. In particular, we explore the enhance performance of a VIS-MIR lidar system in separating the fine and coarse mode by utilizing a broader wavelength spectral. In the simulated test, we show for realistic uncertainties, the retrieval of fine / coarse fraction was improved.

As discussed in our findings, extensive research work is needed especially in the aerosol optical properties, 24 hours / 7 days PBL dynamics and understanding the root causes for WRF model performance biases. Below is the list of possible future research:-

1. Utilizing the long term lidar PBL retrieval to build a seasonal lookup table for WCT parameter “a” and “b”. This lookup table should provide better picture of the variation of the parameter “a” and “b” due to the seasonal change. Thus this information can help the WCT technique become more automated.

2. Statistical analysis of meteorology measurements such as temperature, water vapor mixing ratio and wind speed should be concentrate on. This information provide thermodynamic of the atmosphere which is directly connected to the PBL behavior. In particular, microwave radiometer can provide high resolution measurement of these parameters so the detail comparison can be made.

3. Consider more turbulent indicators in the model including the acquisition of WRF standard output such as ground heat flux, upward heat and moisture flux and latent heat flux at the surface.

4. For the MYJ scheme, we can explore directly the TKE fields to see if simple redefinitions of the threshold of TKE magnitude will affect the PBL retrievals.

5. In addition, an improved ceilometer will be available for 24 hours / 7 days measurements with better SNR which can hopefully provide better observation of the PBL.

6. Recalibrate the microwave radiometer to get better upper atmosphere water vapor and temperature as well as possible upper atmosphere winds from Doppler lidar / SODAR measurements.

References

Anderson, J. R., Hardy, E. E., Roach, J. T., Witmer, R. E., 1976. A Land Use And Land Cover Classification System For Use With Remote Sensor Data, USGS Geological Survey Professional Paper 964.

Baar, H., Ansmann, A., Engelmann, R. & Althausen, D., 2008. Continuous monitoring of the boundary-layer top with lidar. Atmos. Chem. Phys. Discuss., 8, 10749-10790.

Baker. D. et al., 2009, Sensitivity Testing of WRF Physics Parameterizations for Meteorological Modeling and Protocol in Support of Regional SIP Air Quality Modeling in the OTR Ozone Transport Commission Modeling Committee.

Binkowski, F. S., & Roselle, S. J., 2003. Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component, 1. Model description, *J. Geophys. Res.*, 108(D6), 4183, doi:10.1029/2001JD001409.

Boers, R., Spinhirne, J. D. & Hart, W. D., 1987. Lidar Observations of the Fine-Scale Variability of Marine Stratocumulus Clouds. J. Applied Meteorology, Vol. 27, 797-810.

Brooks, I. M., 2003. Finding Boundary Layer Top: Application of a Wavelet Covariance Transform to Lidar Backscatter Profiles. J. Atmospheric & Oceanic Technology, Vol. 20, 1092-1105.

Ching, J. & Byun, D., Chapter 1 - Introduction To The Models-3 Framework And The Community Multiscale Air Quality Model (CMAQ), EPA/600/R-99/030.

Cohn, S. A. & Angevine, W. M., 2000. Boundary Layer Height and Entrainment Zone Thickness Measured by Lidars and Wind-Profiling Radars. J. Applied Meteorology, Vol. 39, 1233-1247.

Corrigan, P., Lwin, M., Gross, B. & Moshary, F., 2009. Design of an ozone and ammonia open-path sensor with a QCL. Optical Society of America.

Culf, A. D. 1989. Acoustic sounding of the atmospheric boundary layer at Halley, Antarctica, Antarctic Science 1 (4): 363-372.

Currie, C. & Valencia, S., 2004. Evaluation of the Micro-Pulse LIDAR Calibration Constant. in Proceeding *of 22nd International Laser Radar Conference*, Edited by Gelsomina Pappalardo and Aldo Amodeo, ESA SP-561, European Space Agency, pp.279-282.

Davis, K. J., Gamage, N., Hagelberg, C. R., Kiemle, C., Lenschow, D. H. & Sullivan, P. P., 2000. An Objective Method for Deriving Atmospheric Structure from Airborne Lidar Observations. J. Atmospheric & Oceanic Technology, Vol. 17, 1455-1468.

Doraiswamy, P., Hogrefe, C., Hao, W., Civerolo, K., Ku, J.Y. & Sistla, G., 2010, A Retrospective Comparison of Model-Based Forecasted PM2.5 Concentrations with Measurements, J. Air & Waste Manage. Assoc. 60(11):1293-308.

Dubinsky, R. H., Carswell, A. I. & Pal, S. R., 1985. Determination of cloud microphysical properties by laser backscattering and extinction measurements. Appl. Opt. 24, 1614-1622.

Eloranta, E. W., 1998. A practical model for the calculation of multiply scattered lidar returns. Appl. Opt. 37, 2464-2472.

Emeis, S., Schafer, K. & Munkel, C., 2008. Surface-based Remote Sensing of the Mixing-Layer Height – a review", Meteorological Zeitschrift, Vol. 17, No. 5, 621-630.

Eresmaa, N., Karppinen, A., Joffre, S. M., Rasanen, J. & Talvitie, H., 2006. Mixing height determination by ceilometer. Atmos. Chem. Phys., 6, 1485-1493.

EPA (1999a). User Manual for the EPA Third-Generation Air Quality Modeling System (Models-3Version 3.0), EPA-600/R-99/055, pp. 1-1 – 1-18, United States Environmental Protection Agency, June 1999.

EPA (1999b). Science Algorithms of the EPA Models-3Community Multiscale Air Quality (CMAQ) Modeling System, PART II: Chapters 9-18, EPA-600/R-99/030, pp. 10-1 – 10-23, United States Environmental Protection Agency, March 1999.

Flamant, C., Pelon, J., Flamant, P. H. & Durand, P., 1997. Lidar determination of the entrainment zone thickness at the top of the unstable marine atmospheric boundary layer. Bound.-Layer Meteor. 83, 247-284.

Flagg, D. D. & Taylor, P. A., 2008. Sensitivity to morphology in urban boundary layer modeling at the mesoscale. 18th Symposium on Boundary Layers and Turbulence, 7A.3.

Gamage, N. & Hagelberg, C., 1993. Detection and analysis of microfronts and associated coherent events using localized transform. J. Atmos. Sci, 50, 750-756.

Haij, de, M.J., Wauben, W.M.F., Klein Baltink, H. and Apituley, A. (2009). Determination of the mixing layer height by a ceilometer *Proceedings of the 8th International Symposium on Tropospheric Profiling, ISBN 978-90-6960-233-2 Delft, The Netherlands.*

Hayden, K. L., Anlauf, K. G., Hoff, R. M., Strapp, J. W., Bottenheim, J. W., Wiebe, H. A., Froude, F. A., Martin, J. B., Steyn, D. G. & McKendry, I. G., 1997. The Vertical Chemical and Meteorological Structure of the Boundary Layer in the Lower Fraser Valley during Pacific '93. J. Atmos. Environ. 31, 2089-2105.

Hennemuth, B., & Lammert, A., 2006. Determination of the atmospheric boundary layer height from radiosonde and lidar backscatter, Boundary-Layer Meteorology 120: 181–200.

Hess, M., Koepke, P., & Schult, I., 1998. Optical Properties of Aerosols and Clouds: The Software Package OPAC. Bulletin of the American Meteorological Society, Vol. 79 No. 5, 831-844.

Hogan, R. J., 2006. Fast approximate calculation of multiply scattered lidar returns," Appl. Opt. 45, 5984-5992.

Hogrefe, C. et al., 2007. Daily Simulation of Ozone and Fine Particulates over New York State: Findings and Challenges. Journal of Applied Meteorology & Climatology, Vol. 46, 961-979, doi:10.1175/JAM2520.1

Holben, B. N., Eck, T. F., Slutsker, I., Tanre´, D., Buis, J. P., Setzer, K. A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I. & Smirnov, A., 1998. AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, REMOTE SENS. ENVIRON. 66:1–16

Hooper, W. P. & Eloranta, E. W., 1986. Lidar Measurements of Wind in the Planetary Boundary Layer: The Method, Accuracy and Results from Joint Measurements with Radionsonde and Kytoon. J. Climate & Applied Meteorology, Vol. 25, 990-1001.

Janjić, Z.I., 1990. The step-mountain coordinate: physical package. *Mon. Wea. Rev.*, 118, 1429-1443.

Janjić, Z.I., 1994. The step-mountain eta coordinate model: further developments of the convection, viscous sublayer and turbulence closure schemes, *Mon. Wea. Rev*., 122, 927-945.

Janjic, Z. I., 2001. Nonsingular Implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model. NCEP Office Note No. 437, 61 pp.

Janjic, Z. I., 2003a. A Nonhydrostatic Model Based on a New Approach. *Meteorology and Atmospheric Physics*, 82, 271-285.

Janjic, Z. I., 2003b. The NCEP WRF Core and Further Development of Its Physical Package. 5th International SRNWP Workshop on Non-Hydrostatic Modeling, Bad Orb, Germany, 27-29 October.

Jiang, W., Smyth, S., Giroux, E., Rothand, H., Yin, D., 2006. Differences between CMAQ fine mode particle and PM2.5 concentrations and their impact on model performance evaluation in the lower Fraser valley, Atmospheric Environment, 40, 4973-4985.

Kaimal, J., Abshire, N., Chadwick, R., Decker, M., Hooke, W., Kropfli, R., Neff, W., Pasqualucci, F. & Hildebrand, P., 1982. Estimating the Depth of the Daytime Convective Boundary Layer. J. Appl. Meteor., 21, 1123-1129.

King, M. D., Tsay, S., Platnick, S. E., Wang, M.H. & Liou, K.N., 1997. Cloud retrieval algorithms for MODIS: optical thickness, effective particle radius, and thermodynamic phase, MODIS Algorithm Theoretical Basis Document. No. ATBD-MOD-05, MOD06–Cloud product (*NASA Goddard Space Flight Center,* version 5).

Knupp, K. R., Coleman, T., Phillips, D., Ware, R., Cimini, D., Vandenberghe, F., Vivekanandan, J., & Westwater, E., 2009, Ground-Based Passive Microwave Profiling during Dynamic Weather Conditions. *J. Atmos. Oceanic Technol.*, 26, 1057–1073.

Kovalev, V.A. & Eichinger, W.E., 2004. Elastic Lidar: Theory, Practice, and Analysis Methods. Wiley-Interscience, USA.

Kunkel, K. & Weinman, J, 1976. Monte Carlo analysis of multiply scattered lidar returns. *J*. Atmos. Sci. 33, 1772–1781.

Lammert, A. & Bosenberg, J., 2006. Determination of the Convective Boundary-Layer Height with Laser Remote Sensing. Boundary-Layer Meteorology, 119: 159-170.

Malm, W.C, Sisler, J.F., Huffman, D., Eldred, R.A. & Cahill, T.A. 1994. Spatial and seasonal trends in particle concentration and optical extinction in the United States. *Journal of Geophysical Research-Atmospheres* 99 (D1), pp. 1347–1370.

Malm, W.C. & Day, D. E., 2001. Estimates of aerosol species scattering characteristics as a function of relative humidity. Atmospheric Environment 35 (2001) 2845-2860.

Martucci, G., Matthey, R., Mitev, V. & Richner, H., 2007. Comparison between Backscatter Lidar and Radiosonde Measurements of the Diurnal and Nocturnal Stratification in the Lower Troposphere. J. Atmospheric & Oceanic Technology, Vol. 24, 1231-1244.

Melfi, S. H., Spinhirne, J. D. & Chou, S-H., 1985. Lidar Observation of Vertically Organized Convection in the Planetary Boundary Layer over the Ocean. J. Climate & Applied Meteorology, Vol. 24, 806-821.

Menut, L., Flamant, C., Pelon, J. & Flamant, P. H., 1999. Urban boundary-layer height determination from lidar measurements over the Paris area. Applied Optics, Vol. 38, No. 6, 945-954.

Mellor, G.L. and T. Yamada, 1982. Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, 20, 851-875.

Miles, N. L., Verlinde, J. & Clothiaux, E. E., 2000. Cloud droplet size distributions in low-level stratiform clouds. J. Atmos. Sci. 57, 295−311.

Min, Q.-L., M. Duan, & R. Marchand (2003), Validation of surface retrieved cloud optical properties with in situ measurements at the Atmospheric Radiation Measurement Program (ARM) South Great Plains site, J. Geophys. Res., 108(D17), 4547, doi:10.1029/2003JD003385.

Münkel, C., Eresmaa, N., Räsänen, J., Karppinen, A., 2007. Retrieval of mixing height and dust concentration with lidar ceilometers, Boundary-Layer Meteorol. 124, 117–128.

Münkel, C. & Roininen, R., 2008. Mixing layer height assessment with a compact lidar ceilometer. Symposium on Recent Developments in Atmospheric Applications of Radar and Lidar, 88th AMS Annual Meeting (New Orleans, LA).

Münkel, C. & Roininen, R., 2010. Investigation of Boundary Layer Structures with Ceilometer using a Novel Robust Algorithm. 15th Symposium on Meteorological Observation and Instrumentation, Session 5, Remote Sensing: Air Quality/Detection of Atmospheric Gases.

Nielsen-Gammon, J. W., Hu, X-M., Zhang, F. & Pleim, J. E., 2010. Evaluation of Planetary Boundary Layer Scheme Sensitivities for the Purpose of Parameter Estimation. American Meteorological Society Monthly Weather Review. Vol. 138, 3400-3417. DOI: 10.1175/2010MWR3292.1

O'Connor, E. J., Illingworth, A. J. & Hogan, R. J., 2004. A technique for autocalibration of cloud lidar**.** J. Atmos. Ocean Tech. 21(5), 777-778.

Olofson, F., 2008. Thesis of lidar studies of tropospheric aerosols and clouds. University of Alaska, Fairbanks, USA

Operational Guidance for the Community Multiscale Air Quality (CMAQ) Modeling System Version 4.7.1, June 2010 Reliease.

Otte, T. L., 2008a. The impact of nudging in the meteorological model for retrospective air quality simulations. Part I: Evaluation against national observation networks, J. Appl. Meteor. Climatol., 47, 1853-1867.

Otte, T. L., 2008b. The impact of nudging in the meteorological model for retrospective air quality simulations. Part II: Evaluating collocated meteorological and air quality observations", J. Appl. Meteor. Climatol., 47, 1868-1887.

Parlange, M. B. & Brutsaert, W. 1990. Are Radionsode Time Scales Appropriate to Characterize Boundary Layer Wind Profiles? Journal of Applied Meteorology, Vol. 29, 249-255.

Piironen, A. K. & Eloranta, E. W., 1995. Convective boundary layer mean depths and cloud geometrical properties obtained from volume imaging lidar data", J. Geophysical Research, Vol. 100, No. D12, 25569-25576.

Pinnick, R. G., Jennings, S. G., Chýlek, P., Ham, C., Grandy Jr., W. T., 1983. Backscatter and extinction in water cloud. J. Geophys. Res. 88 (11), 6787-6796.

Platt, C. M. R., 1981. Remote sounding of high clouds. III. Monte Carlo calculations of multiple-scattered lidar returns. J. Atmos. Sci. 38, 156–167.

Pleim, J. E., 2007a. A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. J. Appl. Meteor. Climatol., 46, 1383–1395.

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. J. Appl. Meteor. Climatol., 46, 1396–1409.

Pleim, J. E. and Gilliam, R., 2009. An indirect data assimilation scheme for deep soil temperature in the Pleim-Xiu land surface model, J. Appl. Meteor. Climatol., 48, 1362-1376.

Seinfeld, J. H. & Pandis, S. N., 1998. *Atmospheric chemistry and Physics: From Air Pollution to Climate Change*. John Wiley and Sons.

Simpson, M., Raman, S., Lundquist, J.K., & Leach, M., 2006. A study of the variation of urban mixed layers heights. Atmospheric Environment, doi:10.1016/j.atmosenv.2006.08.029

Spinhirne, J.D., Boers, R. & Hart, W.D., 1989. Cloud top liquid water from lidar observations of marine stratocumulus. J. Appl. Meteor. 28, 81-90.

Steyn, D. G., Baldi, M. & Hoff, R. M., 1999. The Detection of Mixed Layer Depth and Entrainment Zone Thickness from Lidar Backscatter Profiles. J. Atmospheric & Oceanic Technology, Vol.16, 953-959.

Stull, R.B., 1988. An Introduction to Boundary Layer Meteorology. Kluwer Academic Publishers, Dordrecht/Boston/London.

Tao, Z., McCormick, M. P., Wu, D., Liu, Z. & Vaughan, M. A., 2008. Measurements of cirrus cloud backscatter color ratio with a two-wavelength lidar. Appl. Opt. 47, 1478-1485.

Voss, K. J., Welton, E. J., Quinn, P. K., Frouin, R., Miller, M. & Reynolds, R. M., 2001. Aerosol optical depth measurements during the Aerosols99 experiment. J. Geophys. Res. 106(D18), 20,811–20,819.

Wandinger, U., 1998. Multiple-scattering influence on extinction and backscatter coefficient measurements with Raman and high-spectral-resolution lidars. Appl. Opt. 37, 417-427.

Weather Research and Forecast (WRF) ARW version 3 Modeling System User's Guide January 2011.

Welton, E.J., Voss, K.J., Quinn, P.K. et al., 2002. Measurements of aerosol vertical profiles and optical properties during INDOEX 1999 using micropulse lidars. J. Geophys. Res. 107 (D19), 8019.

Wu, Y., Chaw, S., Gross, B., Moshary, F. & Ahmed, S., 2009. Low and optically thin cloud measurements using a Raman-Mie lidar. Appl. Opt. 48, 1218-1227.

User's Guide for the NMM Core of the Weather Research and Forecast (WRF) Modeling System Version 3, Chapter 5: WRF NMM Model.

Yang, P. & Liou, K. N., 1996. Finite-difference time domain method for light scattering by small ice crystals in three-dimensional space. J. Opt. Soc. Am. A 13, 2072-2085 (1996)

Zhang, Y., Dubey, M. K., Olsen, S. C., Zheng, J. & Zhang, R. 2009. Comparisons of WRF/Chem simulations in Mexico City with ground-based RAMA measurements during the 2006-MILAGRO. Atmos. Chem. Phys., 9, 3777-3798.

Websites:-

URL [1]: http://profmandia.wordpress.com/2010/05/14/vermont-state-climatologist-why-is-that-link-still-there/

URL [2]: http://www.enviroliteracy.org/article.php/1338.html

URL [3]: http://en.wikipedia.org/wiki/Planetary_boundary_layer

URL [4]: http://www.theweatherprediction.com/basic/pbl/

URL [5]: http://www.urbanheatislands.com/home

URL [6]: http://www.actionbioscience.org/environment/voogt.html#primer

List of Papers

This thesis is based on the work presented in the following papers:

- Gan, C.M., Wu, Y.H., Gross, B., and Moshary, F., "Application of active optical sensors to probe the vertical structure of the Urban Boundary Layer and assess anomalies in Air Quality Model PM2.5 Forecasts", Submitted to Atmosphere Environment April 2011.
- Gan, C.M., Wu, Y.H., Gross, B., and Moshary, F., "Application of remote sensing instruments in air quality monitoring", Book Chapter of Air Quality, ISBN: 978-953-307-569-3, June 2011.
- Wu, Y.H., Gan, C.M., Gross, B., Moshary, F., and Ahmed, S., **"**Calibration of lidar at 1064-nm channel using the water-phase and cirrus clouds", Accepted to Applied Optics 2010.
- Gan, C.M., Wu, Y.H., Gross, B., Arend, M. and Moshary, F., "A Comparison of Estimated Mixing Height by Multiple Remote Sensing Instruments and its Influence on Air Quality in Urban Regions", IGARSS Symposium Proceedings, July 2010.
- Gan, C.M., Wu, Y.H., Gross, B., Moshary, F., and Ahmed, S., "Statistical Comparison between Hysplit Sounding and Lidar Observation of Planetary Boundary Layer Characteristics over New York City", SPIE Defense, Security and Sensing, Proceedings Vol. 7684, April 2010.
- Gan, C.M., Corrigan, P., Wu, Y.H., Gross, B., and Moshary, F., "Combining VIS and MIR lidars to separate aerosol modes in the troposphere", SPIE Defense, Security and Sensing, Proceedings Vol. 7673, April 2010.