Assessment of Mixed-Layer Height Estimation from Single-wavelength Ceilometer Profiles

Travis N. Knepp[1,2], James J. Szykman[3,4], Russell Long[3], Rachelle M. Duvall[3], Jonathan Krug[3], Melinda Beaver[3], Kevin Cavender[3], Keith Kronmiller[5], Michael Wheeler[5], Ruben Delgado[6], Raymond Hoff[6], Timothy Berkoff[2], Erik Olson[7], Richard Clark[8], Daniel Wolfe[9], David Van Gilst[10], and Doreen Neil[2]

[1]Science Systems and Applications Inc., Hampton, Virginia 23666, USA
[2]NASA Langley Research Center, Hampton, Virginia 23681, USA
[3]US EPA, Research Triangle Park, Durham, North Carolina 27709, USA
[4]Currently assigned to NASA Langley Research Center, Hampton, Virginia 23681, USA
[5]Jacobs Technology Inc., Tullahoma, Tennessee 37388, USA
[6]Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, Maryland 21250, USA
[7]Space Science and Engineering Center, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA
[8]Department of Earth Sciences, Millersville University, Millersville, Pennsylvania 17551, USA
[9]NOAA/ESRL Physical Sciences Division, Boulder, Colorado 80305, USA
[10]National Suborbital Education and Research Center, University of North Dakota, Grand Forks, North Dakota 58202, USA

Correspondence to: T. N. Knepp
(travis.n.knepp@nasa.gov)

Abstract. Differing boundary/mixed-layer height measurement methods were assessed in moderately-polluted and clean environments, with a focus on the Vaisala CL51 ceilometer. This intercomparison was performed as part of ongoing measurements at the Chemistry And Physics of the Atmospheric Boundary Layer Experiment (CAPABLE) site in Hampton, Virginia and during the 2014 Deriving Information on Surface Conditions from Column and Vertically Resolved Observations Relevant to Air Quality (DISCOVER-AQ) field campaign that took place in and around Denver, Colorado. We analyzed CL51 data that were collected via two different methods (BLView software, which applied correction factors, and simple terminal emulation logging) to determine the impact of data collection methodology. Further, we evaluated the STRucture of the ATmosphere (STRAT) algorithm as an open-source alternative to BLView (note that the current work presents an evaluation of the BLView and STRAT algorithms and does not intend to act as a validation of either). Filtering criteria were defined according to the change in mixed-layer height (MLH) distributions for each instrument and algorithm and were applied throughout the analysis to remove high-frequency fluctuations from the MLH retrievals. Of primary interest was determining how the different data-collection methodologies and algorithms compare to each other and to radiosonde-derived boundary-layer heights when deployed as part of a larger instrument network. We determined that data-collection methodology is not as important as the processing algorithm and that much of the algorithm differences might be driven by impacts of local meteorology and precipitation events that pose algorithm difficulties. The results of this study show that a common processing algorithm is necessary for LIght Detection And Ranging (LIDAR)-based MLH intercomparisons, and ceilometer-network operation and that sonde-derived boundary layer heights are higher (10–15% at mid-day) than LIDAR-derived mixed-layer heights. We show that

averaging the retrieved MLH to 1-hour resolution (an appropriate time scale for a priori data model initialization) significantly improved correlation between differing instruments and differing algorithms.

1 Introduction

The atmospheric boundary layer (ABL) is the lowermost portion of the troposphere that is directly influenced by the Earth's surface and responds to surface forcing of heat, moisture, pollutant emissions, and momentum on a timescale of 1 hour or less (Stull, 1988). The ABL can be defined by a number of criteria depending on the particular interest (e.g. thermodynamic boundary layer, chemical boundary layer (CBL), aerosol mixed layer). The ABL is typically defined by thermodynamic data (i.e., potential temperature and/or skew-T plot) obtained from meteorological sondes. While meteorological sondes have excellent vertical resolution, the temporal resolution is generally poor, ongoing regular sonde launches are labor intensive, and coverage is limited. Conversely, mixed-layer heights (MLH), as calculated from backscatter LIght Detection And Ranging (LIDAR) instruments, provide both excellent vertical and temporal resolution. Typical analysis of LIDAR data involves identification of gradients within the aerosol profile (Brooks, 2003), which is generally considered to be a marker for the MLH. With respect to air quality, the top of the ABL often acts like a lid on the lowest layer of the atmosphere and temporarily traps the majority of near-surface anthropogenic and biogenic emissions. As a result, the vertical distribution of ambient air pollutants, and associated precursors, within the ABL and lower-troposphere are strongly influenced by the height of, and vertical mixing within, the ABL.

ABL variability complicates quantitative determination of surface trace-gas levels from a remote-sensing platform (Coen et al., 2014; Herman et al., 2009; Knepp et al., 2015; Lamsal et al., 2008, 2014; Petritoli et al., 2004; Piters et al., 2012). Therefore, properly accounting for ABL variability from a continuous measurement system such as LIDAR will provide invaluable information to policy, health, modeling, and remote-sensing communities for applications sensitive to the vertical profiles of tracers (Compton et al., 2013; Martin, 2008; Scarino et al., 2014). In 2009, the United States National Research Council highlighted ABL height as a high priority observation needed to improve meso-scale predictions of air quality, short-range severe-weather forecasting, and regional climate modeling (NRC, 2009). More recently, the National Plan for Civil Earth Observation called for improved observation density and sampling of the boundary layer (NSTC, 2014). In 2015, as part of the revisions to the ozone (O_3) National Ambient Air Quality Standards, the U.S. Environmental Protection Agency (EPA) finalized a new requirement under the Photochemical Assessment Monitoring Stations (PAMS) program for the collection of continuous MLH observations. By 2019, the PAMS program will involve implementation of approximately fifty air-quality sites in the United States that provide continuous MLH.

Kotthaus et al. (2016) showed that intercomparison of ceilometer data is not a straight-forward endeavor. An intercomparison of ceilometer instrumentation was carried out in support of upcoming PAMS monitoring requirements. Results from an intercomparison of three backscatter LIDAR instruments from the 2014 DISCOVER-AQ field campaign in Colorado (low aerosol load) and the Chemistry and Physics of the Atmospheric Boundary Layer Experiment (CAPABLE) site at NASA's Langley Research Center (LaRC; moderate aerosol load) in Hampton, Virginia are presented herein.

2 Instrumentation

2.1 CL51

The Vaisala (Vantaa, Finland) CL51 ceilometer is a single-wavelength (eye safe Class 1M InGaAs diode laser emitting at 910 $\pm$ 10 nm, pulsed at 6.5 kHz with a 110 ns pulse width with average pulse power of 19.5 mW, and an avalanche photodiode
5 detector centered at 915 nm), single-lens, LIDAR system originally designed to report cloud-base heights and visibility. More recently, ceilometers have been used to estimate MLH (Emeis and Schäfer, 2006; Emeis et al., 2008a, b; Haeffelin et al., 2012; Morille et al., 2007; Schäfer et al., 2012, 2013; Schween et al., 2014; Sokol et al., 2014; Wiegner et al., 2014). These ceilometers have 10 m vertical resolution (with 10 m overlap) to a maximum altitude of 15.4 km ($\pm$ greater of 1% or 5 m precision, all altitudes are with respect to ground level) and up to 2 s temporal resolution (depending on the control software),
10 though profiles are generally averaged over 16–36 s to improve the signal-to-noise ratio (see section 3.1 for more details). An example backscatter plot that includes increased signal at 3 km due to transport of smoke from a Canadian forest fire is presented in Fig. 1.

The CL51 was designed to operate continuously, regardless of meteorological conditions, in an autonomous manner with minimal user support. Due to the emission wavelength's proximity to the near-infrared water vapor bands, ceilometers operating
15 at the stated wavelengths experience water vapor interference, thereby lessening their utility in retrieval of aerosol optical properties. However, the interference on aerosol profile and MLH estimation is negligible (Wiegner et al., 2014).

Two CL51s were deployed as part of the 2014 DISCOVER-AQ mission in Colorado (Golden, and Erie, Colorado). Before and after deployment, these ceilometers were set up to continually collect data at the CAPABLE site and the EPA Ambient Air Innovative Research Site (AIRS) in Durham, North Carolina. The ceilometers were collocated with meteorological sonde
20 (met-sonde) launch sites during the DISCOVER-AQ campaign and at the CAPABLE site, allowing a direct intercomparison of the sonde and LIDAR ABL/MLH methodologies. Furthermore, during the DISCOVER-AQ campaign the ceilometers were collocated with other LIDAR instruments. Intercomparisons are presented in section 5.

2.1.1 Full-profile Collection

The Vaisala standard MLH retrieval is based on a proprietary wavelet/gradient technique built into the logging/analysis software
25 BLView. The BLView software provides not only logging and data analysis (e.g. MLH and cloud-height estimates) but also archiving capability. While the CL51 reports backscatter up to 15.4 km, BLView truncates the data-collection at 4.5 km, precluding ability to monitor upper-troposphere/lower-stratosphere transport of aerosol, smoke, or ash from major events. Therefore, a full-profile collection method that can run side-by-side with the standard data-collection software was developed and implemented.
30 Data transmission from the ceilometer to the logging computer was achieved by splitting an RS-232 connection into two ports on the logging computer: one port logging to BLView and the other logging to a custom script (e.g. as written in Python). The primary drawback of using a secondary script to log the full profile (as opposed to logging in BLView) is the inability
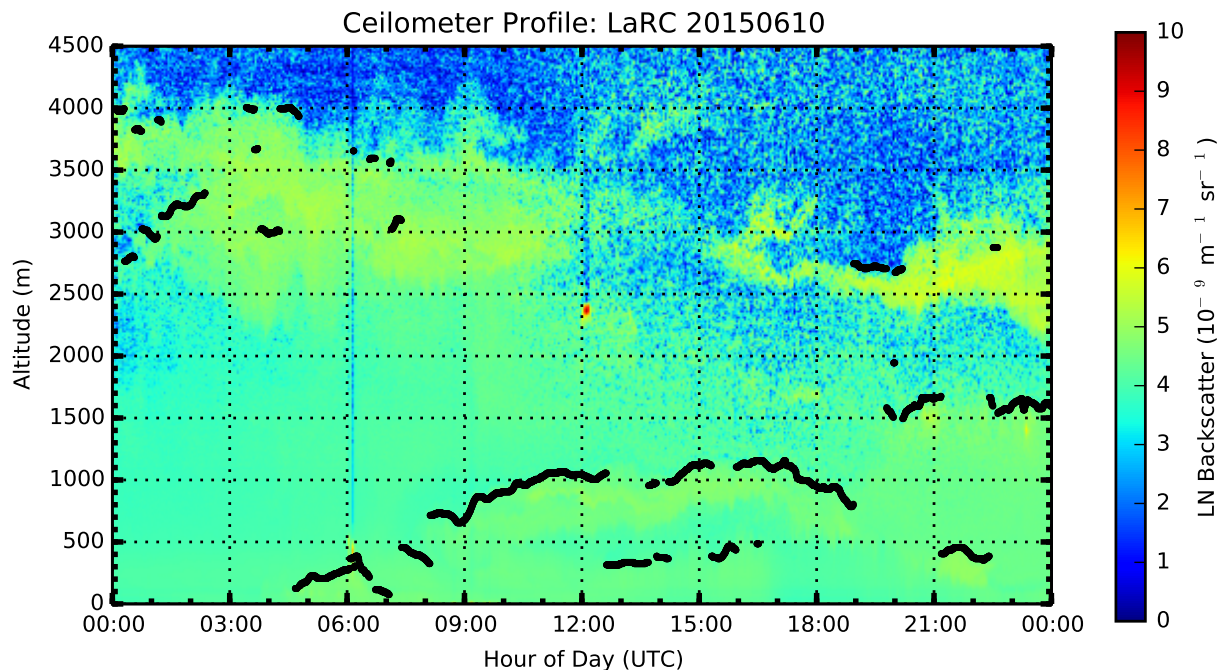
Figure 1. Backscatter curtain plot collected on 10-June 2015 when smoke from a Canadian forest fire was transported over the CAPABLE site. The smoke is observed by increased backscatter in the 2500 – 4000 m range.

to apply proprietary calibration coefficients that are built into the BLView software to the logged data. However, as shown in subsequent sections, this impacts neither the MLH estimates nor the general profile shape substantially.

2.2 Micropulse LIDAR

Elastic LIDAR observations were performed using a Sigma Space (Lanham, Maryland) Micropulse LIDAR (MPL), previously
5 described by Spinhirne (1993) and Welton et al. (2000). Briefly, the MPL transmitter consists of an eye-safe Nd:YLF laser emitting at 527 nm and pulsed at 2.5 kHz with a pulse power of $6 - 10$ μJ. It has a software programmable vertical resolution, with possible values of 15, 30, and 75 m (up to 25 km), and temporal resolutions ranging from 1 s to 15-min. The receiver consists of a 178 mm telescope that collects the backscattered light, which is then focused onto a photon counting silicon avalanche photo-diode (APD). The APD output is recorded by a field programmable gate array data system that enables display
10 and storage of range dependent average count rates on a laptop computer. The raw data are converted to aerosol attenuated backscatter, correcting for instrumental factors such as detector dead time, geometrical overlap, background subtraction, and range-squared normalization. Recorded LIDAR profiles have temporal and vertical resolution of 1 min and 30 m, respectively, as set by the UMBC team for the DISCOVER-AQ campaign. MPL is used for continuous recording of aerosol profiles and optical properties, and calculating MLH values.

2.3 Meteorological/Ozone Sondes

A meteorological sonde (herein referred to as sonde/radiosonde) is the conventional method for measuring temperature, pressure, and humidity throughout the atmosphere, and characterizing the ABL. Radiosondes were used to identify steep gradients within the potential temperature (theta) profile (Fig. 2 A) as identified by the Heffter criteria shown in Eqs. (1) and (2) where Θ is potential temperature in Kelvin, Z is altitude in meters, and Θ_{top} and Θ_{base} refer to the potential temperature at the top and bottom of the proposed inversion layer as described in (Heffter, 1980; Marsik et al., 1995). This thermodynamic ABL is a product of atmospheric turbulent kinetic energy and lapse rate. Similar gradients can be seen in chemical and aerosol profiles as well (Fig. 2 B-C). For the current study, radiosondes from International Met Systems (iMet; Grand Rapids, Michigan) and ozone sondes from Droplet Measurement Technologies (DMT, now En-Sci; Boulder, Colorado) were used. iMet sondes require no preparation and were used as received from the manufacturer, while ozone sondes were conditioned according to the procedure defined by the World Meteorological Organization recommendations (Smit, 2013).

$$\frac{\Delta\theta}{\Delta Z} \geq 0.005 K * m^{-1} \tag{1}$$

$$\Theta_{top} - \Theta_{base} \geq 2K \tag{2}$$

Results of numerous analyses have been published to illustrate differences between the various chemical and meteorological sensors, and how differing meteorological sensors influence secondary chemical measurements such as ozone (Deshler et al., 2008; Dirksen et al., 2014; Johnson et al., 2002; Miloshevich et al., 2004; Nash et al., 2006, 2011; Smit, 2013; Stauffer et al., 2014). While these influences can impact the derived CBL, the ABL and MLH remain unperturbed. Therefore, the remainder of the current work focuses on the MLH and ABL, with CBL variability regarded as outside the current scope.

3 Algorithms

3.1 BLView

BLView makes use of variable time and altitude averaging when calculating the MLH. Typical averaging time ranges from 14 min at night to 52 min during clear-sky, daytime conditions, and is automatically adjusted within the software according to signal-to-noise ratio. Altitude averaging varies with altitude and ranges from 80 m near the surface to 360 m above 1.5 km. Further, BLView selectively removes false-positive MLH identifications by requiring a minimum number of similar MLH values (±140 m) within the last several minutes, and has the ability to discriminate between MLH inversions and changes in backscatter intensity induced by clouds, precipitation, and fog.

Advantages of the BLView software are the standardization of retrieval parameters and a user interface that provides flexibility in setting user-specified sensitivities. These come at the cost of a database system that makes access to raw data difficult and the inability to batch process archived data, posing a severe limitation on reprocessing datasets with a long record history.
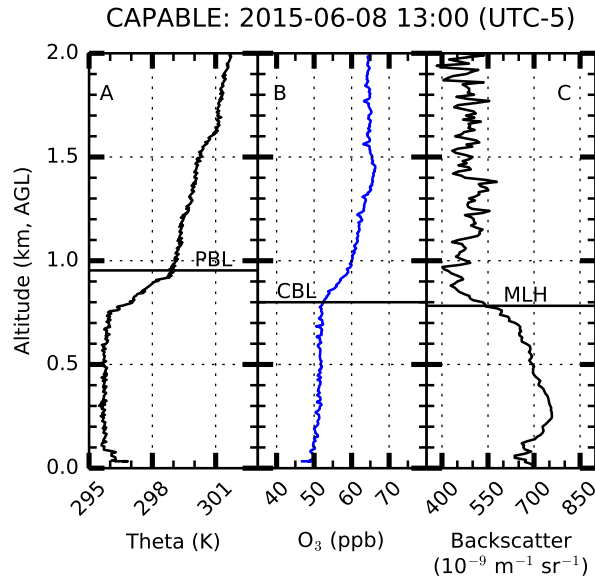
Figure 2. Potential temperature, ozone, and backscatter profiles recorded on 8-June 2015. The horizontal lines indicate the ABL, CBL, and MLH at 13:00 local time.

3.2 STRAT

The STRucture of the ATmosphere (STRAT v1.04) algorithm was developed under a GNU General Public License to analyze aerosol vertical profiles as measured by LIDAR and estimate cloud heights and aerosol MLH from a variety of LIDAR instruments. It is currently in use by the European Aerosol Research Lidar NETwork (EARLINET) (Haeffelin et al., 2012; Hirsikko
5 et al., 2014; Morille et al., 2007; Pappalardo et al., 2014). STRAT uses a covariance wavelet technique (CWT), of which the full details can be found in Morille et al. (2007) and Haeffelin et al. (2012). STRAT can be run exclusively in MATLAB, or a combination of MATLAB and Python. Due to its wide use throughout the European network it is considered here as a viable open-source alternative to BLView.

While BLView provides limited user control of the retrieval process, which is beneficial with regard to standardizing the
10 retrieval process across a network, STRAT provides a significantly greater amount of user control. Such control is desirable since retrieval parameters in a heavily polluted region will likely be different from those in a clean environment. Further, STRAT is provided as raw scripts as opposed to BLView's compiled executable, making the STRAT platform independent and highly user-configurable. STRAT also can run batch jobs, which is useful when reprocessing data from instruments that have a long record history.
15 The STRAT algorithm implements a user-defined normally-distributed weighting function in both the temporal and vertical domains to smooth the data, similar to BLView. In the current study, the STRAT averaging time and vertical resolution were

set to match the BLView settings as much as possible for intercomparison. An analysis of how well the two MLH algorithms agree is presented below.

3.3 UMBC Algorithm

The University of Maryland Baltimore County (UMBC) algorithm was developed independently for estimating MLH from LIDAR backscatter profiles using a CWT similar to STRAT. The STRAT software was designed specifically for single-channel LIDARs (primarily ceilometers) and is not readily customizable to other LIDAR systems, such as the MPL. The UMBC algorithm was designed to be more flexible than STRAT in that regard and uses a CWT to identify the sharp gradient changes indicative of the MLH (Davis et al. 2000; Brooks 2003). A detailed description of the UMBC algorithm has been published in Compton et al. (2013).

4 Locations

4.1 CAPABLE Site

The CAPABLE site was established at LaRC, in the greater Hampton Roads region (a group of cities in coastal Virginia, also known as Tidewater Virginia: Virginia Beach, Norfolk, Chesapeake, Newport News, Hampton, Portsmouth, Suffolk, Poquoson, Williamsburg), for continuous monitoring of air-quality and meteorological parameters to bridge the gap between satellite observations and ground conditions (i.e., where pollutants directly impact living organisms), improve applicability of satellite data to the air-quality user community, and act as a long-term satellite validation site. CAPABLE has a suite of in situ and remote-sensing instruments, including a CL51 ceilometer and sounding station. These instruments allow thorough sampling of the atmosphere to provide valuable in situ and profile information within the lower troposphere in a highly complex (due to bay-breeze events; see Martins et al. (2012)) and moderately polluted (NO_x, SO_2, aerosols) environment yielding valuable satellite ground-truthing and model a priori estimates.

CAPABLE (37.103º N, 76.387º W, 5 m ASL) is located on a peninsula between the James River to the southwest, the Chesapeake Bay to the north, and the Atlantic Ocean to the east. The Hampton Roads region can be described as moderately polluted. Aerosol statistics ($PM_{2.5}$ and aerosol optical thickness (AOT) as recorded by a sun photometer within the AERosol Robotic NETwork (AERONET) as described by Holben et al. (1998)) are presented in Table 1. The data show AOT loads at CAPABLE are significantly higher than at the corresponding Colorado sites, particularly in the lower size distributions (i.e., lower wavelengths in Table 1).

4.2 DISCOVER-AQ/FRAPPE Sites

From 2011 through 2014 the National Aeronautics and Space Administration (NASA) conducted the Deriving Information on Surface Conditions from Column and Vertically Resolved Observations Relevant to Air Quality (DISCOVER-AQ) Earth Venture Suborbital Mission with four field deployments. A primary objective of DISCOVER-AQ was to investigate the ability

Site	λ (nm)/PM Size	Mean	Q_1	Q_2	Q_3
BAO-Tower	380	0.23	0.13	0.19	0.32
BAO-Tower	500	0.15	0.09	0.13	0.22
BAO-Tower	675	0.09	0.05	0.08	0.13
BAO-Tower	870	0.06	0.04	0.06	0.08
BAO-Tower	1020	0.05	0.03	0.04	0.06
Golden, CO	380	0.20	0.10	0.16	0.27
Golden, CO	500	0.13	0.06	0.10	0.18
Golden, CO	675	0.08	0.04	0.06	0.11
Golden, CO	870	0.05	0.03	0.04	0.07
Golden, CO	1020	0.04	0.02	0.03	0.05
CAPABLE	380	0.34	0.21	0.33	0.45
CAPABLE	500	0.23	0.13	0.23	0.32
CAPABLE	675	0.14	0.08	0.14	0.20
CAPABLE	870	0.09	0.05	0.09	0.13
CAPABLE	1020	0.07	0.03	0.06	0.10
CAPABLE	$PM_{2.5}$	5.80	2.59	5.00	8.44

Table 1. Aerosol optical thickness statistics at the three sites under study. Here, Q_1, Q_2, and Q_3 represent the 25th, 50th, and 75th percentiles, respectively. Data have been filtered to show only data collected during the DISCOVER-AQ 2014 field campaign period (July – August 2014).

of satellite remote sensing to inform surface air quality. Since the ABL limits vertical exchange of primary pollutants and directly influences near-surface pollutant concentrations, the ABL height directly influences air quality and chemistry. Therefore, measurements during these missions focused on the vertical distribution of trace gases and aerosols within the ABL and lower troposphere, as well as the diurnal variability of these distributions in conjunction with the ABL. The final DISCOVER-AQ
5   field mission was conducted over Denver and the Front Range region of Colorado in July and August 2014, and was conducted jointly with the Front Range Air Pollution and Photochemistry Experiment (FRAPPE).

4.2.1 Erie, Colorado/BAO-Tower

Data were collected at the Erie, Colorado site (40.045º N, 105.005º W, 1500 m ASL), which is considered to be a clean environment as compared to CAPABLE (see Table 1) from 14-July to 12-August 2014 as part of the DISCOVER-AQ field
10   mission. The Erie site (rural community surrounded by agricultural activity) was located at NOAA's Earth System Research Laboratory's (ESRL) Boulder Atmospheric Observatory (BAO) and served as a combined DISCOVER-AQ/FRAPPE ground site. The site is often referred to as BAO-Tower because of the site's primary feature: a 300 m tower. BAO-Tower provided a unique profiling ability for in situ samplers by mounting them on the tower for static sampling, or on the carriage to collect "active" profiles.

As part of FRAPPE, the University of Wisconsin's (UW) Space Science and Engineering Center trailer, which housed a high spectral resolution lidar and from which regular sonde launches were performed, was stationed at the site. The UW trailer temporarily housed a CL51 during the mission. Due to the proximity of the UW trailer, both ceilometers experienced the same chemical, aerosol, and meteorological conditions.

5   4.2.2   Golden, Colorado

CL51 data were collected at the Golden, Colorado site (39.750° N, 105.183° W, 1850 m ASL) (considered to be a clean environment as compared to CAPABLE, see Table 1) from 14-July to12-August 2014 as part of the DISCOVER-AQ field mission. The Golden site was located next to the National Renewable Energy Laboratory (NREL) on Table Mountain mesa (a flat-topped geographic structure). Due to the site's elevation on the mesa, and its limited emissions sources, conditions at
10   the Golden site were generally clean from an aerosol perspective and did not typically experience a well-developed ABL/ML. This is demonstrated in Fig. 3 by the lack of structure in the diurnal MLH profile. While both the BAO and CAPABLE sites demonstrate the expected nocturnal low/daytime high MLH, the Golden diurnal variability is not as well defined, consistent with ABL development in mountainous terrain (Banta, 1984; Tripoli and Cotton, 1989; Bossert et al., 1989; Bossert and Cotton, 1994).
15   The Golden site housed the U.S. EPA trailer, the LaRC ozone LIDAR, MPL and LEOSPHERE ALS-450 LIDAR operated by UMBC, a SOnic Detection and Ranging (SODAR) instrument operated by Millersville University (MU), and regular met-sonde launches from the MU group.

5   Analysis

LIDAR data collected during the DISCOVER-AQ campaign had sampling times that ranged from 36 to 60 s, while sonde-
20   profile data had average measurement times of 1 s. Due to the nature of sounding data sets, sonde-based ABL's were not averaged to 5-min resolution. To harmonize LIDAR datasets to a common time frame the data were averaged to 5-min resolution unless otherwise specified. Further, it is well known that the atmosphere changes throughout the day due to surface heating, etc. (hence, driving ABL variability). Therefore, some of the analyses were broken into 4-hour segments to remove biases caused by time-of-day influences. Since the primary objective of this assessment was to understand how the CL51 MLH
25   compared with other instruments/methods, all analytical results are presented in relation to the CL51.

The analysis was performed using several ceilometer MLH products to do a thorough comparison of instruments (CL51, MPL, and met-sondes), collection method (allowing BLView to collect profile data with application of calibration factors vs. logging raw data with a custom Python script), and data processing algorithm (BLView vs. STRAT and custom MLH scripts from UMBC). Assessment of data-acquisition methodology is presented first, followed by a comparison of MLH retrieval
30   algorithms applied to data collected by a single instrument, and then a comparison of the various instrumentation. As MLH variability follows a distinct diurnal cycle as shown in Fig. 3, all dates/times are presented in local standard time.
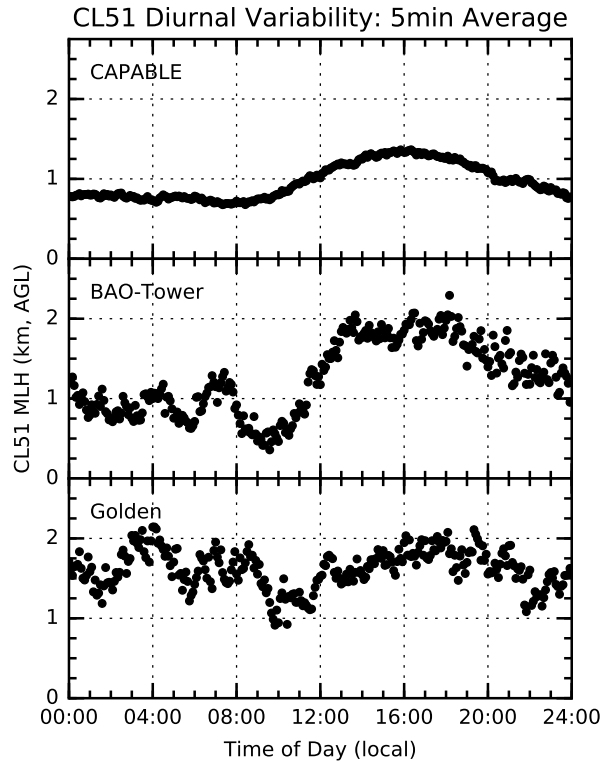
Figure 3. Diurnal variability of the MLH at the three sites. Data were resampled to 5-min averages and filtered.

5.1 Data Acquisition

Data-acquisition methods were analyzed to determine whether the CL51 data-logging methodology influenced the MLH estimate. As described above, CL51 profile data were logged using two methodologies: BLView and a custom Python routine. The BLView software has the advantage of applying the ceilometer's calibration factors and preconditioning the profiles (here referred to as BLView; note, however, that this refers to the backscatter-profile that is logged by BLView and not the BLView-calculated MLH), while the Python script logged the raw incoming data stream up to the full profile (FP) height (i.e., 15.4 km). The question was, does application of the LIDAR calibration factor influence the MLH estimate? This question is addressed in section 5.1.2, but first, viable filtering criteria to remove spurious MLH fluctuations from the data set were developed prior to analysis, as discussed in section 5.1.1.

5.1.1 Filtering Criteria

Regardless of the data-acquisition method (i.e., BLView or Python), pragmatic data-selection criteria were needed for quality control. Since ABL and MLH variations occur in a generally smooth manner, it is expected that the variance within a short
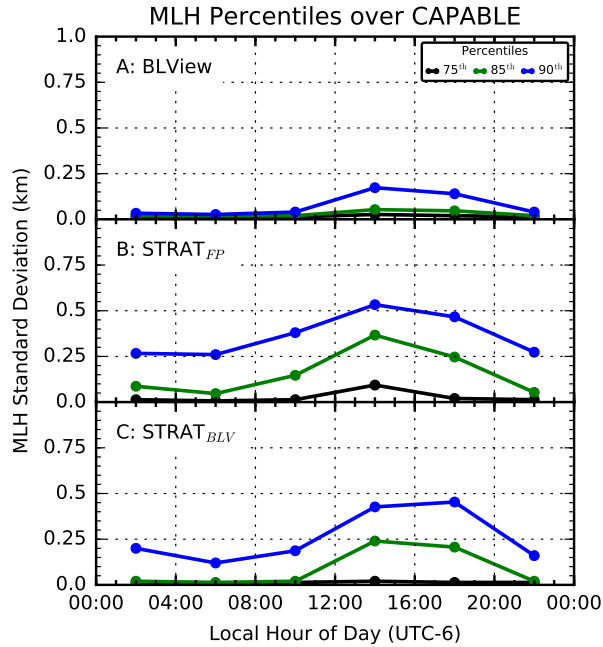
Figure 4. Percentiles for MLH standard deviation throughout the day from the CAPABLE site. Data in panel A were collected and processed in BLView, data in panel B collected with the Python script and processed in STRAT, data in panel C were collected in BLView and processed in STRAT. It is observed that variability was maximum during the afternoon regardless of collection method or processing algorithm.

time interval will be minimal, and that any larger variance is indicative of other events (e.g. precipitation, frontal systems, window contamination). Therefore, cutoff criteria for implementing data filtering were identified. This portion of the analysis was conducted first because application of these cutoff criteria will influence the data acquisition comparison (i.e., BLView-corrected data vs. raw data collected via the Python script).

5   Despite the atmosphere's smooth variation in ABL and MLH, these parameters do change substantially over long periods of time (e.g. an hour or day), with standard deviations significantly increasing over the longer time periods and during rapid transition events. Therefore, the current analysis was performed on short-time-series data (i.e., MLH resampled to 5-min resolution) to eliminate bias caused by natural low-frequency changes. Figure 4 shows a series of percentile plots for data collected at LaRC (N > 30E5), where the standard deviation of MLH was calculated over 5-min intervals and subsequently averaged

10   to provide mean standard deviation every four hours. This figure elucidates the variability of the MLH standard deviation for both collection methods and algorithms. Except for the afternoon period (12:00 – 19:00, local time) when the variability is slightly increased, 85% of the data fall within one standard deviation (≈ 0.20 km) regardless of time of day. Therefore, data with a 5-min standard deviation greater than 0.20 km were removed from subsequent analysis (labeled "filtered"). Data with a relative standard deviation greater than or equal to 20% were also removed. Implementation of these filter criteria removed up

15   to 10% of the data at each site.
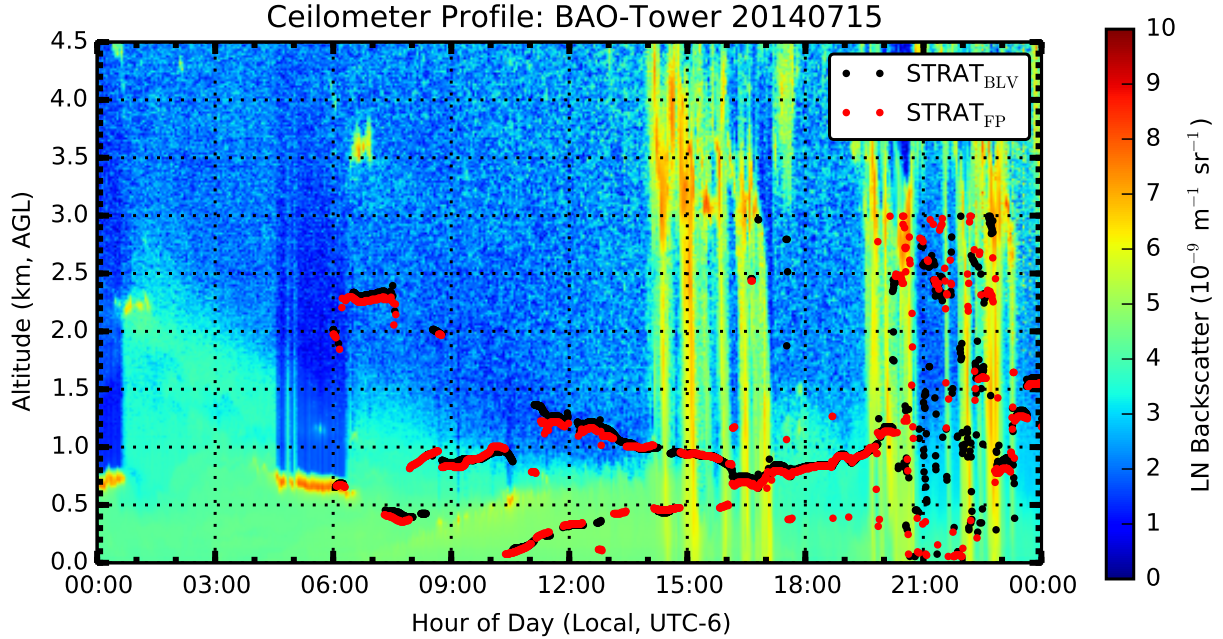
Figure 5. Backscatter curtain plot with STRAT-derived MLH values (5-min mean) from the BLView (BLV) and Python (FP) collection methods.

This filtering method is further supported by observing the variability in the BLView and Python-collected datasets (both processed in STRAT) in relation to backscatter curtains (Fig. 5) where it is observed that much of the difference between the BLView and Python-collected data occurs during times of high variability or precipitation (e.g. 19:00 – 24:00 in Fig. 5). During such events, neither collection method is expected to provide valid MLH estimates; rather, to overcome such discrepancies, if possible, the MLH algorithms must be adjusted accordingly.

5.1.2   Collection Method Dependence

To determine whether the data-collection method influenced MLH estimates, both BLView and Python-collected backscatter profiles were processed on a common algorithm (STRAT) using identical input configuration files. Both the BLView and FP profiles were processed using the STRAT algorithm as described in section 3.2, followed by a 5-min block average.

The data were replotted as correlation plots with the z-axis being representative of the immediate data density (a dimensionless value that has been scaled to 1). The data density was calculated by implementing a Gaussian-based kernel-density estimation (Scott, 1992; Silverman, 1986) as supplied in Python's scipy.stats.kde module, represented mathematically in Eqs. 3–5 where **X** is the 2 x n vector of the **x** and **y** vectors (i.e., flattened and stacked atop one another), n represents the number of points within each dataset (assuming datasets are of equal length), f is the Scott's factor ($n^{\frac{-1}{d+4}}$), d is the number of independent datasets analyzed, and Eq. 5 is evaluated over the range 1 to n. As these density values are used as weights in subsequent

calculations, the output vector is labeled **w** here. It is observed that the majority of MLH estimates fall along the 1:1 line (center column in Fig. 6), though there is significant scatter along both axes.

$$\Delta\mathbf{X} = \mathbf{X} - \mathbf{X}[:, i] \tag{3}$$

$$\mathbf{E} = \sum_{j=1} \Delta\mathbf{X}_j \cdot \frac{cov(\mathbf{X})^{-1}}{f^{-2}} \bullet \Delta\mathbf{X}_j \tag{4}$$

$$5 \quad \mathbf{w}[i] = \left. \frac{\sum\limits_{k=1} e^{-\mathbf{E}_k}}{\sqrt{det\left[2\pi \cdot cov(\mathbf{X}) \cdot f^2\right]}} \right|_i^n \qquad\qquad \{i \in \mathbb{N} : i \leq n\} \tag{5}$$

Figure 6 was divided into 4-hour blocks to identify any time-of-day dependence. The figure shows that most of the data continued to fall along the 1:1 line regardless of time of day, as indicated in the CAPABLE and BAO-Tower density plots. The Golden site displays some disruption in the $16:00-19:59$ panel, but the source of this discrepancy is currently unknown. It has become clear, however, that the meteorology at the Golden site is different from that observed at CAPABLE and BAO-Tower.

10 It is suggested that this difference is primarily driven by orographic perturbations as well as the Golden site's location atop a mesa, both of which can inhibit formation of stable ABL and ML (Bossert et al., 1989; Bossert and Cotton, 1994; Tripoli and Cotton, 1989).

For regulatory and modeling applications, 1-hour averages are standard, requiring the data be averaged down to 1-hour resolution. The impact of the filtering criteria and re-sampling to 1-hour resolution throughout the day can be seen in Fig. 7.

15 Note, the density of data around the 1:1 line is readily apparent in Fig. 7, therefore the z-axis has been converted to relative standard deviation to show the relative variability within each 1-h time block, after application of filtering criteria. The intention is to provide some understanding of how much the MLH will change within the model and regulatory applications' time frame. Table 2 presents statistics on the aggregate analysis. While the aggregate coefficients of correlation and line-of-best-fit (LOBF) equations do not change substantially after re-sampling to 1-hour blocks, the scatter is dramatically reduced. This is likely due

20 to the scatter being evenly distributed around the 1:1 line and the majority of data points falling along the 1:1 line, as observed in the data-density panels of Fig. 7.

It can be concluded from the current analysis that the majority of variability was driven by local atmospheric fluctuations and events that cannot be readily accounted for within the algorithms. In addition, no significant difference is observed between the BLView- and Python-collected data sets on the timescales relevant to model inputs and atmospheric variations, when processed

25 on a common algorithm. Findings presented in section 5.1.3 further support this conclusion.

5.1.3 MLH Algorithm Dependence

In the previous section, the data collection method (i.e., Python vs. BLView) was shown to have little impact on the derived MLH values when the two datasets were processed using a common algorithm (STRAT). The question remains of how the two
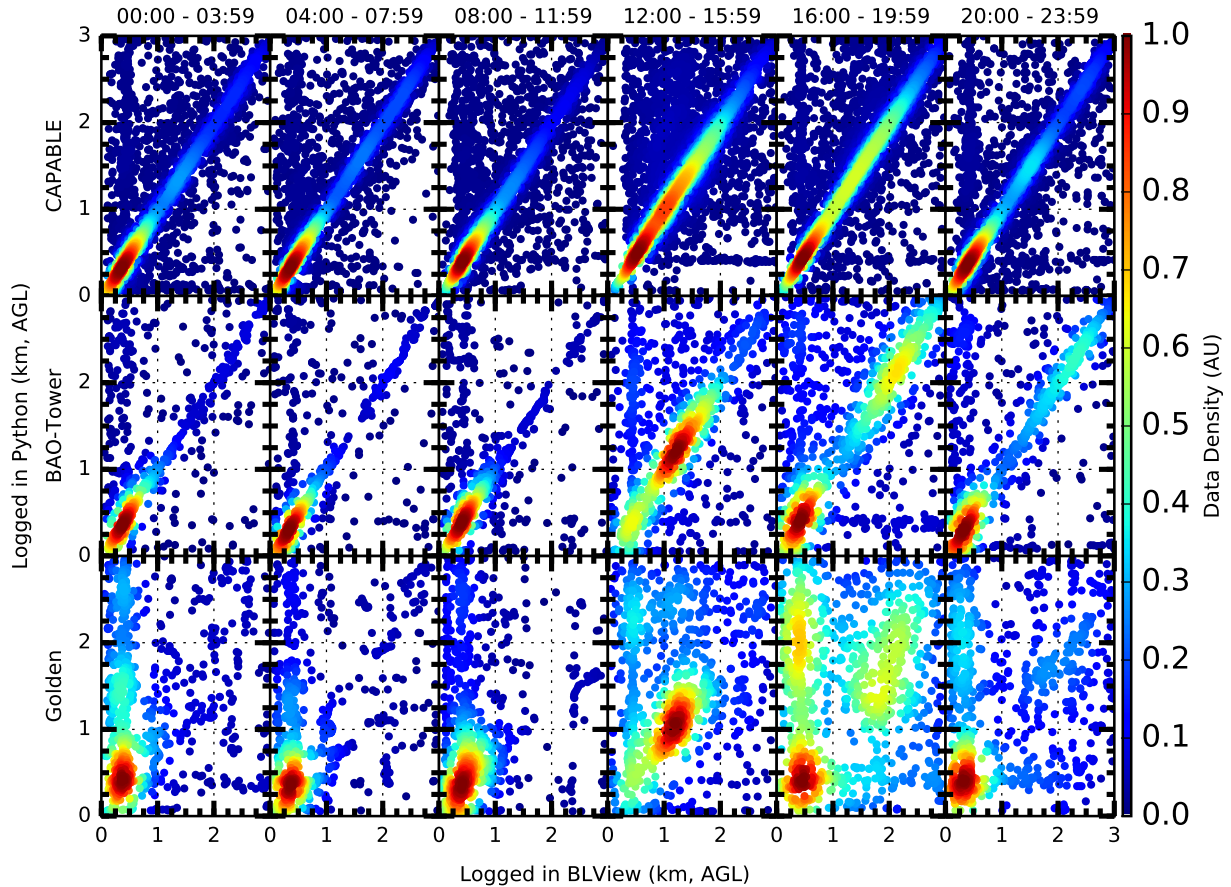
Figure 6. Correlation plots for data collected at the three sites under study. Data density is presented to better understand the distribution within the scatter plots. Data were averaged to 5-min resolution, without application of filtering criteria.

	R	LOBF	$\langle FP - BL \rangle$ (km)
CAPABLE 5-min	0.87	$y = 0.913 \cdot x + 0.11$	-0.02 (1.4)
CAPABLE 1-h	0.87	$y = 0.925 \cdot x + 0.11$	-0.03 (2.7)
BAO 5-min	0.76	$y = 0.817 \cdot x + 0.25$	-0.08 (9.1)
BAO 1-h	0.77	$y = 0.814 \cdot x + 0.32$	-0.14 (15.1)
Golden 5-min	0.72	$y = 0.777 \cdot x + 0.30$	-0.08 (8.1)
Golden 1-h	0.77	$y = 0.792 \cdot x + 0.35$	-0.14 (13.0)

Table 2. Summary of aggregate statistics for the Python-collected (FP)/STRAT-processed and the BLView-collected (BLV)/STRAT-processed MLH estimates (y and x, respectively). Data were resampled to 5-min resolution followed by application of filtering criteria to both datasets (lines labeled 1-h present statistics after data were filtered and subsequently resampled by a 1-h block average). Values in parentheses indicate percent of the difference value with respect to the BLView-derived MLH.
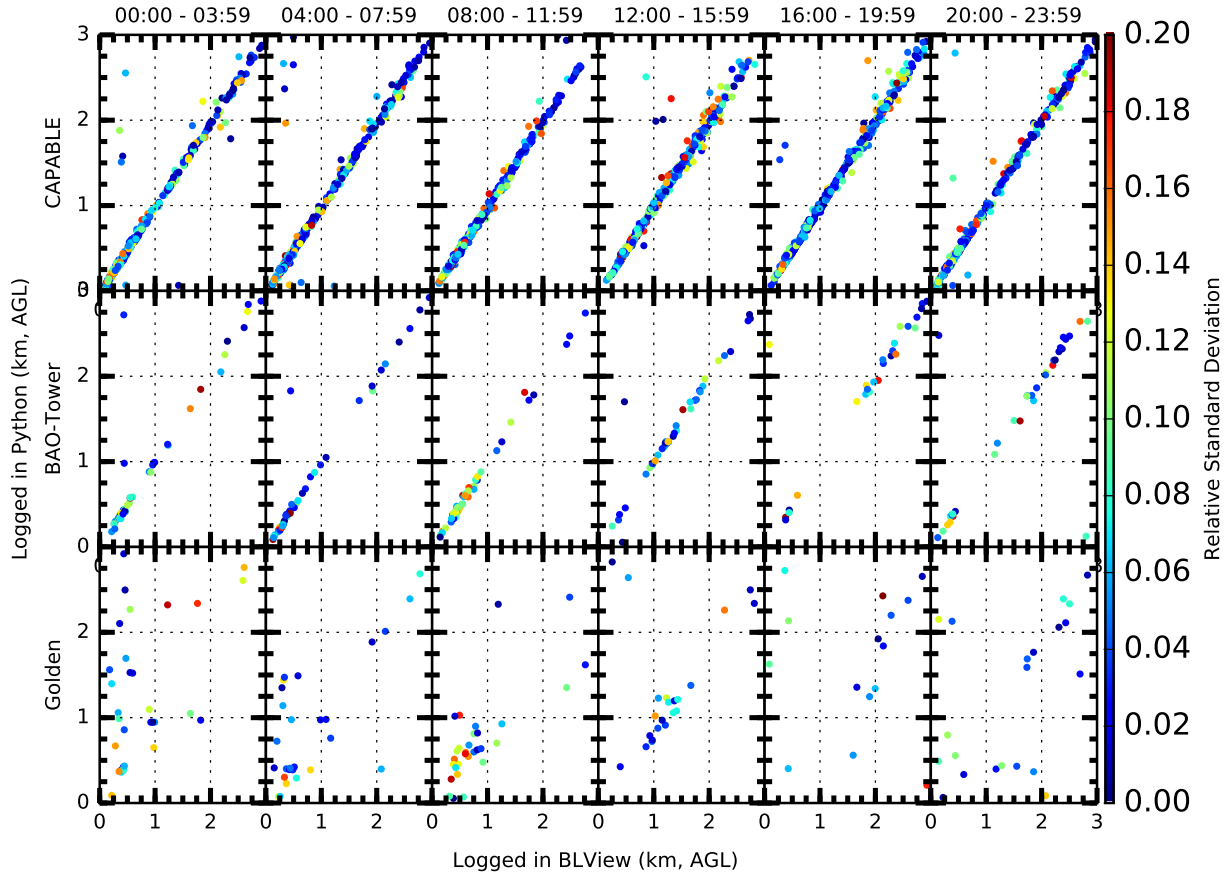
Figure 7. Same data set as in Figure 6, but with the data resampled to 1-h means after application of filtering criteria. Due to the sparseness of the data, as compared to Fig 6, there is not need to present data density on the z-axis. Here, the z-axis represents relative standard deviation to show the relative variability within each 1-h block after filtering.

data sets compare when processed in different algorithms. To answer this question, data collected with the Python script were processed using the STRAT algorithm and were compared with data collected and processed with BLView.

Figure 8 presents scatter plots similar to those in Fig. 6, but with data collected and processed using the two different methods. Most data continued to fall along the 1:1 line, as shown in the density plots, and much of the scatter is caused by short-term variability. However, in contrast to Fig. 6, the scatter is neither as evenly distributed nor as tightly grouped around the 1:1 line. The STRAT-derived MLHs were generally lower than those calculated in BLView (given by the slopes) at all sites, while the aggregate mean difference shows the opposite for the Colorado sites (Table 3), which is likely driven by outliers.

The agreement between the two datasets is less than when a common algorithm was employed (Table 3). Despite the increased scatter, a significant subset of data remains along the 1:1 line. As a test for how well the data fit the 1:1 line, the R and LOBF values were re-calculated using Eq. 5 with weights applied according to data density. Therefore, points that had a

	R	Line of best fit	$\langle FP - BLV \rangle$ (km)	R_w	$LOBF_w$	$\langle FP - BLV \rangle_w$ (km)
CAPABLE 5-min	0.47	$y = 0.499 {\cdot} x + 0.70$	-0.24 (26.8)	0.836	$y = 0.986 {\cdot} x + 0.70$	-0.02 (3.8)
CAPABLE 1-h	0.48	$y = 0.467 {\cdot} x + 0.74$	-0.23 (24.4)	0.799	$y = 0.997 {\cdot} x + 0.74$	-0.08 (12.5)
BAO 5-min	0.43	$y = 0.374 {\cdot} x + 0.65$	0.04 (3.4)	0.789	$y = 0.905 {\cdot} x + 0.65$	-0.01 (0.9)
BAO 1-h	0.39	$y = 0.305 {\cdot} x + 0.72$	0.17 (13.1)	0.740	$y = 0.881 {\cdot} x + 0.72$	-0.01 (1.2)
Golden 5-min	0.24	$y = 0.193 {\cdot} x + 0.90$	0.25 (17.7)	0.541	$y = 0.629 {\cdot} x + 0.90$	0.09 (10.3)
Golden 1-h	0.12	$y = 0.086 {\cdot} x + 1.12$	0.39 (23.6)	0.316	$y = 0.361 {\cdot} x + 1.12$	0.20 (16.1)

Table 3. Summary of statistics for the Python-collected/STRAT-processed and the BLView-collected/BLView-processed MLH estimates. Values in parentheses indicate percent of the difference value with respect to the BLView-derived MLH, and the w subscript indicates a weighting function was applied. Data were resampled to 5-min resolution followed by application of filtering criteria to both datasets (lines labeled 1-h present statistics after data were filtered and subsequently resampled by a 1-h block average).

	R	Line of best fit	$\langle FP - BLV \rangle$ (km)	R_w	$LOBF_w$	$\langle FP - BLV \rangle_w$ (km)
CAPABLE 5-min	0.54	$y = 0.553 {\cdot} x + 0.61$	-0.19 (21.2)	0.91	$y = 0.975 {\cdot} x + 0.61$	-0.03 (4.2)
CAPABLE 1-h	0.54	$y = 0.519 {\cdot} x + 0.64$	-0.18 (18.5)	0.87	$y = 0.973 {\cdot} x + 0.64$	-0.06 (9.8)
BAO 5-min	0.41	$y = 0.326 {\cdot} x + 0.58$	0.14 (13.1)	0.78	$y = 0.843 {\cdot} x + 0.58$	-0.01 (1.4)
BAO 1-h	0.31	$y = 0.232 {\cdot} x + 0.65$	0.32 (25.5)	0.58	$y = 0.573 {\cdot} x + 0.65$	-0.09 (13.0)
Golden 5-min	0.25	$y = 0.184 {\cdot} x + 0.83$	0.36 (24.5)	0.49	$y = 0.484 {\cdot} x + 0.83$	0.16 (16.6)
Golden 1-h	0.14	$y = 0.101 {\cdot} x + 0.96$	0.55 (32.8)	0.27	$y = 0.229 {\cdot} x + 0.96$	0.37 (29.1)

Table 4. Summary of statistics for the BLView-collected/STRAT-processed and the BLView-collected/BLView-processed MLH estimates. Values in parentheses indicate percent of the difference value with respect to the BLView-derived MLH, and the w subscript indicates a weighting function was applied. Data were resampled to 5-min resolution followed by application of filtering criteria to both datasets (lines labeled 1-h present statistics after data were filtered and subsequently resampled by a 1-h block average). Herein, the comparison is limited strictly to the MLH algorithms.
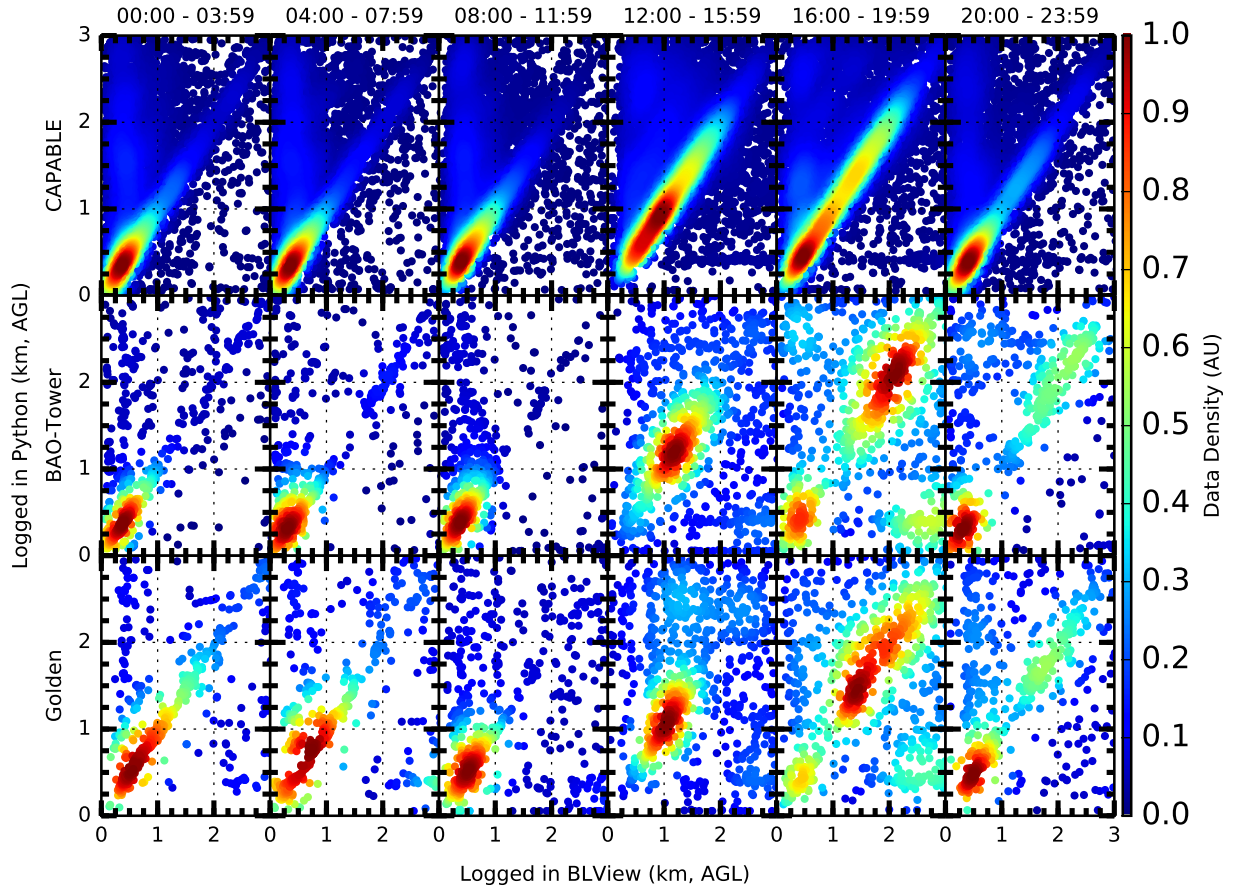
Figure 8. Correlation plots for data collected at the three sites under study. At all sites the data were collected by/processed in, Python/STRAT and BLView/BLView. Plots show the data density to better understand the distribution within the scatter plots. Data were averaged to five-minute resolution, without application of filtering criteria.

greater number of surrounding data points received more weight, while more isolated points received less weight. Weighted coefficients of correlation were calculated using Eq. (6), where variables with a w subscript indicate weighted means. Weighted regressions were performed by simultaneously solving the modified normal equations of regression shown in Eqs. (7) and (8) with weighting factors applied.

5   $$R = \frac{\sum\limits_{i=1}^{N} w_i \cdot (x_i - \bar{x}_w) \cdot (y_i - \bar{y}_w)}{\sqrt{\sum\limits_{i=1}^{N} w_i \cdot (x_i - \bar{x}_w)^2 \cdot \sum\limits_{i=1}^{N} w_i \cdot (y_i - \bar{y}_w)^2}} \tag{6}$$

$$m = \frac{N \sum\limits_{i=1}^{N} w_i x_i y_i - \left(\sum\limits_{i=1}^{N} w_i x_i \right) \left(\sum\limits_{i=1}^{N} w_i y_i \right)}{N \sum\limits_{i=1}^{N} w_i x_i^2 - \left(\sum\limits_{i=1}^{N} w_i x_i \right)^2} \tag{7}$$

$$b = \frac{\sum\limits_{i=1}^{N} w_i y_i - m \sum\limits_{i=1}^{N} w_i x_i}{N} \tag{8}$$

These weighted statistics are not included to suggest that the agreement has actually improved (R), nor do they suggest improved predictability (LOBF). Rather, the improved R values and slopes reflect the degree to which the data are predominantly
5   distributed around the 1:1 line to the exclusion of other regions. As an example, the improvement in the Golden regressions, despite weighting, is notably less than the other two sites. This is likely due to more spread in the data, which mitigates the influence of the points along the 1:1 line in the regression analyses. Therefore, the preponderance of the data collected at the CAPABLE and BAO-Tower sites fall nearer the 1:1 line when processed using the different algorithms as compared to the data collected at the Golden site. Further, despite most data falling nearer the 1:1 line for these two sites, influences remain that
10 neither the STRAT configuration nor the current filter methodology can account for, which is likely driving the poor correlation as compared to Table 2. This is possibly a product of how the differing algorithms handle atmospheric interferential events (e.g. precipitation, fog). Application of a filtering methodology to account for and remove these events will be the subject of future study.

Finally, the analysis was repeated by using STRAT to process backscatter data collected by BLView for comparison with
15 the BLView-collected/processed product. As concluded in section 5.1.2, the data collection method had little influence on the MLH estimation when both datasets were processed using a common algorithm (STRAT). Based on that conclusion, it would be expected that the current comparison would be similar to the previous comparison as summarized in Table 3. This is, in fact, what was observed. The aggregate statistics for the BLView-collected, STRAT-processed vs. BLView-collected/processed intercomparison are presented in Table 5, wherein we see similarity with Table 3. These findings further support the conclusion
20 that data collection methods (including application of calibration factors) play much less of a role in identifying a qualitative gradient within the profile than the choice of MLH algorithm. Indeed, it can be concluded that choice and configuration of the algorithm is critical and that, for network intercomparisons, all networked LIDAR systems should have their data processed by a common algorithm.

5.2   Sonde Intercomparison

25 Meteorological soundings have been a staple for profiling the atmosphere and deriving ABL heights for decades. These ABL heights are typically derived using potential temperature (e.g. using the Heffter criteria) or through analyzing skew-T, log-P plots that implement potential temperature, both of which are different from the gradient-based MLH algorithms implemented

Site	Algorithm	Not Filtered	Filtered
CAPABLE	BLView	0.79 (26)	0.82 (23)
CAPABLE	STRAT	0.14 (14)	0.82 (11)
BAO-Tower	BLView	0.63 (16)	0.58 (14)
BAO-Tower	STRAT	0.34 (16)	0.79 (14)
Golden	BLView	-0.28 (12)	0.74 (10)
Golden	STRAT	0.70 (13)	0.70 (13)
Composite	BLView	0.55 (54)	0.80 (47)
Composite	STRAT	0.44 (43)	0.72 (38)

Table 5. Summary of statistics for the CL51/sonde MLH/ABL intercomparison, corresponding to Fig. 9. Numbers in parentheses indicate sample size. Composite statistics were generated by looking at all sites as a single dataset. In this table only, the filtering method for the STRAT-based MLH is based on visual identification of false MLH values due to clouds/precipitation events and unusually clean atmospheres as described in text.

here. As ABL data are typically used in chemical transport models, it is necessary to determine how these MLH data compare to the sonde-derived ABL data collected at the three measurement locations.

Intercomparison of sonde-based ABL and ceilometer-based MLH can be complicated due to the fundamentally different nature of the two observations. Sondes provide a direct measurement of the atmosphere, while ceilometers provide an indirect
5 (i.e. remotely-sensed) measurement. Therefore, care must be taken when comparing the two sets of observations. Further, the aerosol profile can be impacted by aerosol layers transported aloft, thereby offsetting the MLH estimate. Since the sondes capture an ephemeral snapshot of the atmosphere's current conditions and traverse several kilometers in the horizontal direction due to winds, the ceilometer data were averaged over 30-min for comparison. Additionally, each measurement can be impacted by atmospheric phenomena that can affect the measurements in different ways and can in turn affect the comparison of the
10 measurements. Met-sondes can be impacted by local updrafts/downdrafts and result in ABL estimates that are higher or lower than the time- or space-averaged MLH. The response time of the sensors is less than one second, thereby minimizing offset in vertical structure. The CL51 MLH is calculated based on identification of a sufficiently steep, vertically-averaged, backscatter gradient, so if there are additional aerosol layers just above the MLH, the contrast between the aerosol layers might not be strong enough for the CL51 to identify each layer or the correct altitude of the MLH.
15 Correlation plots for the CL51 MLH calculated via BLView compared to sonde ABL are shown in Fig 9 panels A-C. For all coincidence times, the CAPABLE site showed the best correlations between the CL51 and sondes. The correlation for the CL51 versus all the sondes (N = 25) at the CAPABLE site was R=0.79, with a similar correlation R = 0.82 (N = 22) when the filtering criteria were implemented. For daytime data, the CAPABLE site contained two early morning sondes (before 10:00 local time), with all other sondes launched between 10:00 and 16:00 local time. By late morning, ≈10:00 local time,
20 the vertical dispersion of aerosols due to turbulent mixing has likely resulted in a well-mixed boundary layer, so the ABL and MLH coincide in elevation, which is evident in Fig. 9 A where many of the data points fall close to the 1:1 line.
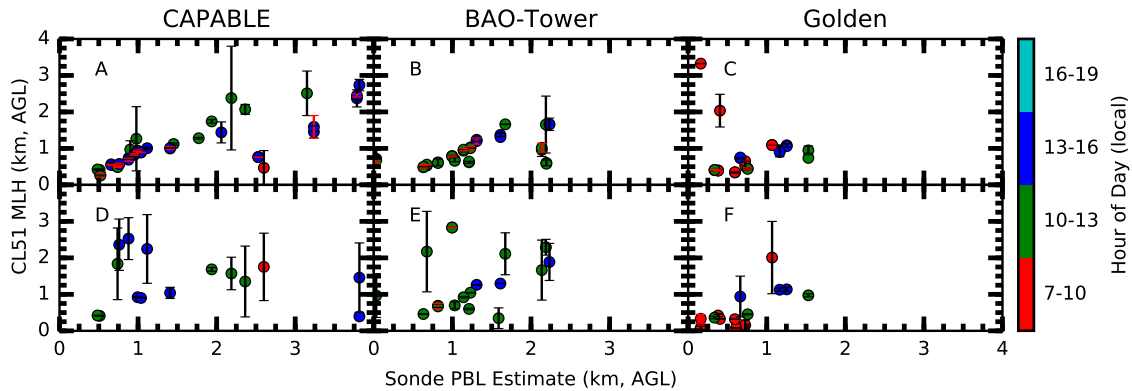
Figure 9. Correlation plots for CL51 MLH and sonde-derived ABL estimates. Black error bars represent the spread in unfiltered data, while the red error bars represent the filtered dataset. MLH values (30-min average, centered on sonde-launch time) were calculated in BLView (panels A-C) and STRAT (panels D-F) and resampled to 30-min resolution. Error bars indicate standard deviation of the CL51-derived MLH within the 30-min period.

Met-sonde data collected at the BAO-Tower site showed lower correlations than the CAPABLE site (unfiltered R = 0.63, N = 16; filtered R = 0.58, N = 14), while the Golden site correlations (unfiltered R = -0.28, N=12) appear to be strongly impacted by two morning sonde launches, which occurred during a transition period when the boundary layer was experiencing rapid growth. Upon applying the filtering criteria, the two early morning data points were removed, resulting in a much improved correlation (filtered R = 0.74, N = 10) for the Golden site. These results indicate the CL51 might have difficultly capturing an accurate MLH during rapidly changing conditions, such as during early morning and late evening transition periods in a clean atmosphere.

It is somewhat surprising that the filtered correlation for the Golden site is better than the filtered result for the BAO-Tower site, given the BAO-Tower site is situated farther to the east of the Rocky Mountains, at the start of the High Plains, which are less influenced by very local geographic perturbations, and that a similar relationship is not observed in the CL51 intercomparisons (Tables 2, 3, and 5). As a check of the met-sonde potential temperate profiles, the potential temperature data from the NASA P-3B aircraft spirals conducted over the Golden and Erie sites is shown in Figs. 12 and 13. These spirals are coincident with the launch of the met-sondes from the sites. The coincident CL51 backscatter profiles are also plotted in Figs. 12 and 13. The agreement between the radiosonde and P-3B aircraft profiles is good, indicating that the potential temperature within the aircraft spiral radius is consistent with that of the radiosonde. These figures show agreement between the potential temperature ABL and CL51 MLH by identifying the same first major gradient in the MLH data on certain days.

The STRAT-derived intercomparison with sonde ABL is presented in Fig 9 panels D-F, where it is observed that the agreement is significantly less than when BLView was used to calculate MLH. This disparity is caused by spurious MLH values from STRAT that are observed under two conditions: 1. during heavy cloud cover/precipitation events STRAT sometimes falsely identified the cloud deck as the MLH and completely ignored the MLH gradient 1-2 km below the cloud; 2. STRAT failed to
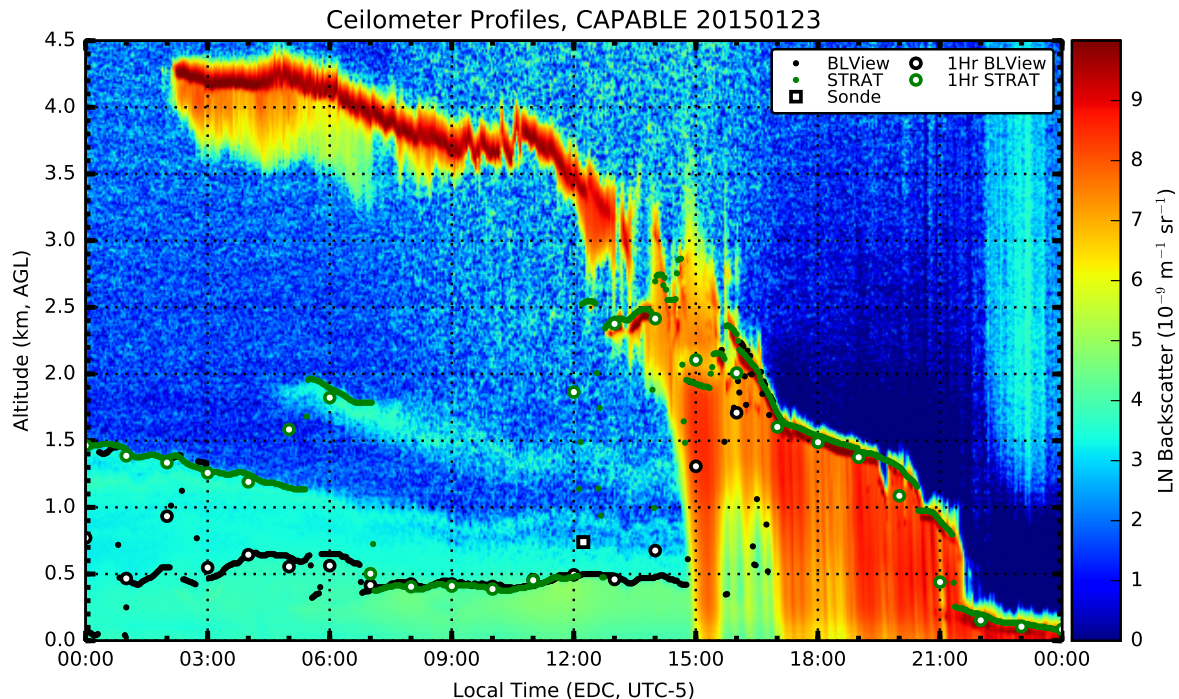
Figure 10. Example plot where STRAT identifies cloud deck as the MLH (12:00).

identify a valid MLH when the atmosphere was exceptionally clean, and instead identified a stronger, spurious, gradient 2-4 km up. An example of the first type is presented in Fig. 10 where STRAT switches from properly identifying the MLH at ≈0.5 km to identifying the cloud deck (≈2.4 km) as the MLH starting around 12:00 local time and an example of the second type is shown in Fig. 11. A corresponding shift was not observed in the BLView-derived MLH for the same day, indicating BLView
5 has been trained to recognize these spurious events and ignore them.

After removing these "false" MLH values the coefficient of correlation between STRAT-derived MLH and sonde ABL (pre-filtering) improved for all sites to 0.82, 0.79, and 0.70 for LaRC, BAO-Tower, and Golden, respectively. The results of these correlations is encouraging and is indicative of the importance of properly training the STRAT algorithm to identify and exclude these false-positive events. The down side is that despite having better correlation (after removing spurious events),
10 the variance of STRAT MLH values larger than that of BLView, indicating that defining an MLH filter criteria is dependent on the algorithm in use. However, the positive aspect of this is that the STRAT algorithm, being open source with the source code available, can, in theory, be modified by end users to identify and account for these spurious events.

Overall, all three sites show good correlation between the CL51 and met-sonde data, with MLH and ABL estimates from the sondes being, on average, higher than the CL51 MLH (200 m (13%), 390 m (15%), -240 m (9%) for CAPABLE, BAO-Tower,
15 and Golden respectively) as indicated in the linear regression lines plotted in Fig. 9, with the exception being the unfiltered results for Golden.
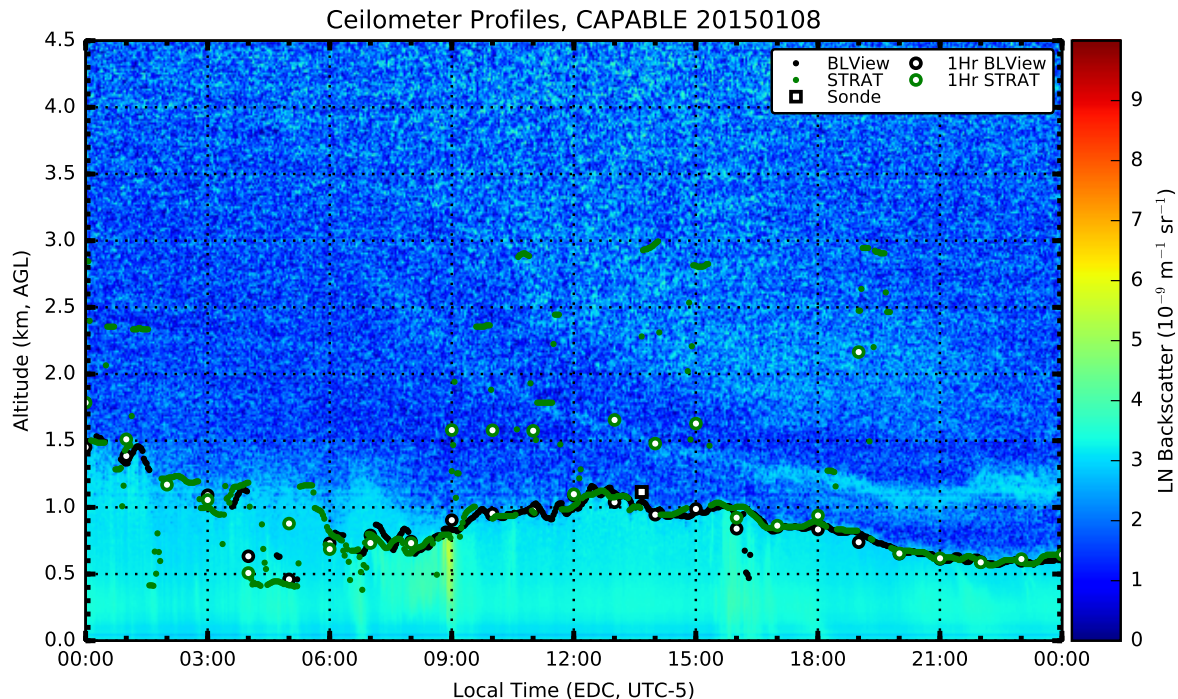
Figure 11. Example plot where STRAT fails to identify a reasonable MLH due to unusually clean conditions.

5.3 MPL Intercomparison

The MPL instrument was collocated with the CL51 stationed at the NREL site in Golden, Colorado. Being a LIDAR instrument, it profiles the atmosphere similarly to the CL51 with the major difference being their hardware. The two instruments emit different wavelengths (CL51:910 nm, MPL:532 nm), causing the instruments to differ in sensitivity with respect to particle

5   size and geometry. Therefore, it is feasible that the two instruments observed "different" atmospheres in a quantitative manner (e.g. AOT). However, if the ML is well mixed, then the general particle distribution and gradient will be the same, making the two inter-comparable.

Figures 14 and 15 shows that the agreement between the two instruments and algorithms (BLView, STRAT for CL51 profiles and UMBC algorithm processing MPL profiles) is poor, even though a significant subset of data fall along the 1:1 line, as

10   indicated by data density (*z*-axis). The low correlation is partly driven by the invariability in one instrument as compared to the other at lower MLH values ($\leq$ 500 m). Removal of MLH below 500 m improved the coefficients of correlation for the 5-min averaged data to 0.467, 0.489, and 0.469 for BLView-derived MLH values (Fig. 14, panels A, B, C respectively) and 0.433, 0.471, and 0.368 for STRAT-derived MLH (Fig. 15, panels A, B, C respectively) values. Similar to the algorithm comparison, much of the variability between the two instruments and algorithms occurs during events that inhibit a reliable estimation (e.g.,

15   fog, precipitation) of MLH (as seen in Fig. 16).

BAO-Tower, CO Potential Temperature Profiles
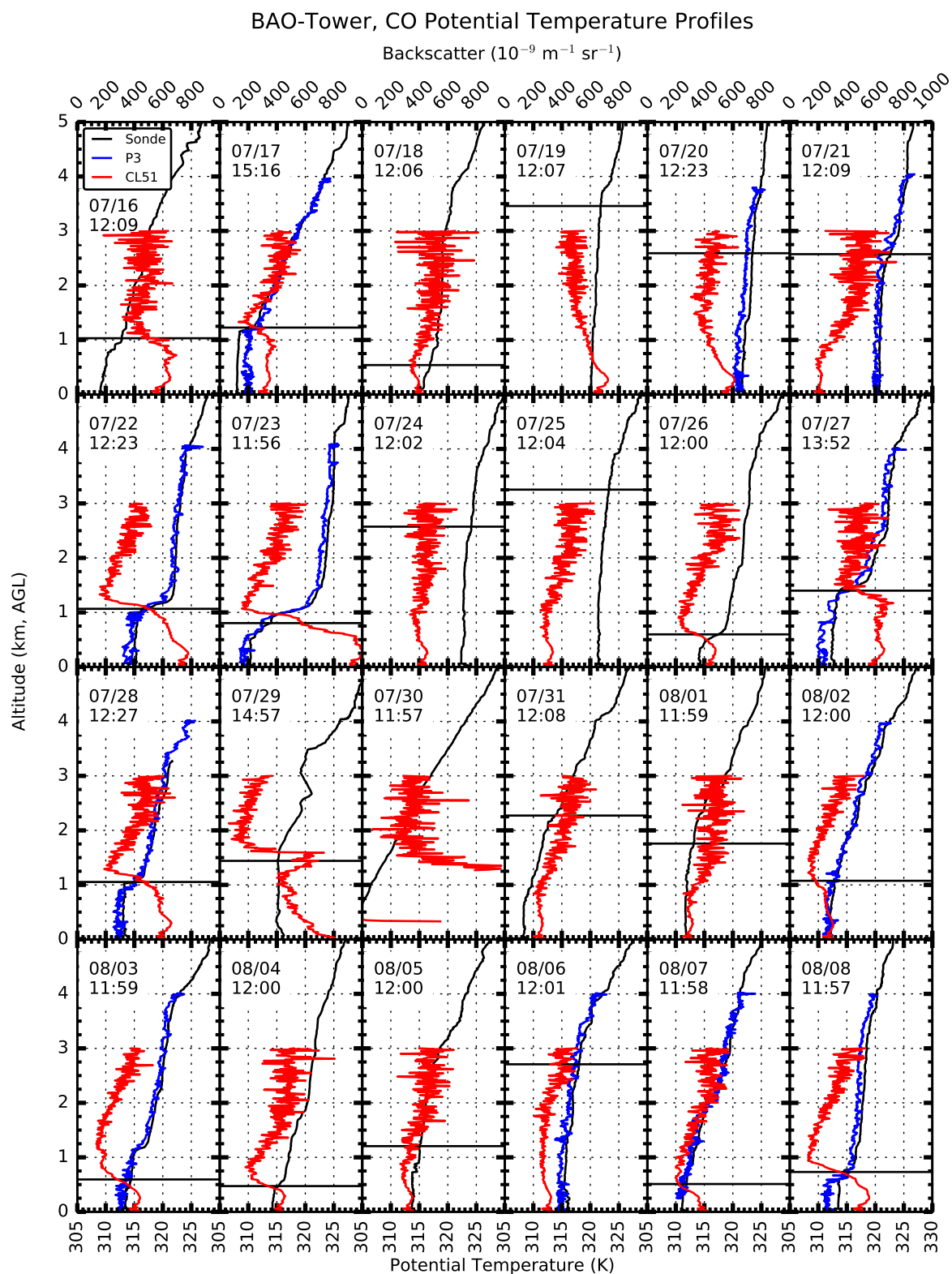
Figure 12. Potential temperature and Cl51 backscatter profiles collected at the BAO-Tower site. Horizontal lines indicate MLH as determined by BLView.
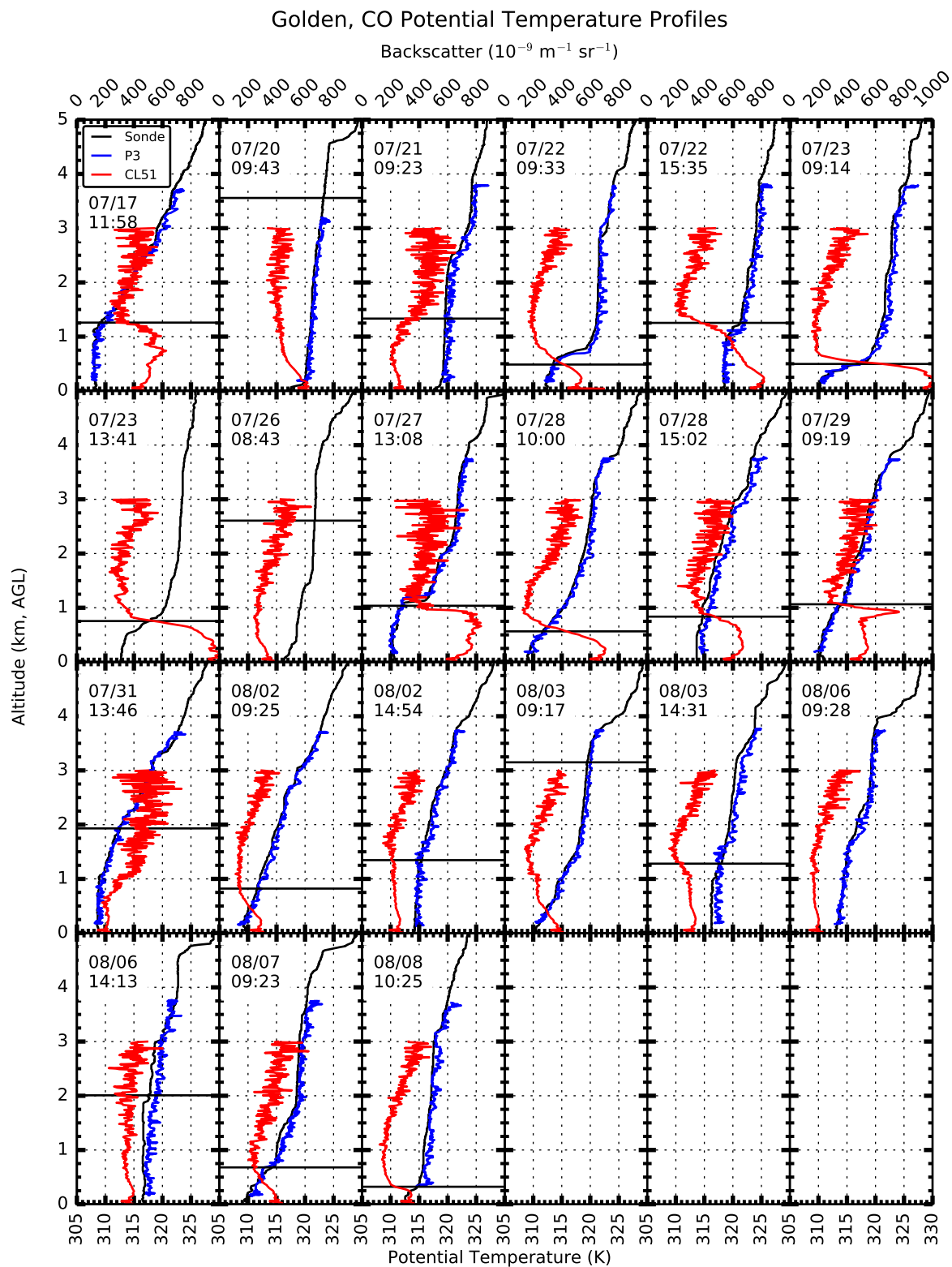
Figure 13. Potential temperature and Cl51 backscatter profiles collected at the Golden NREL site. Horizontal lines indicate MLH as determined by BLView.
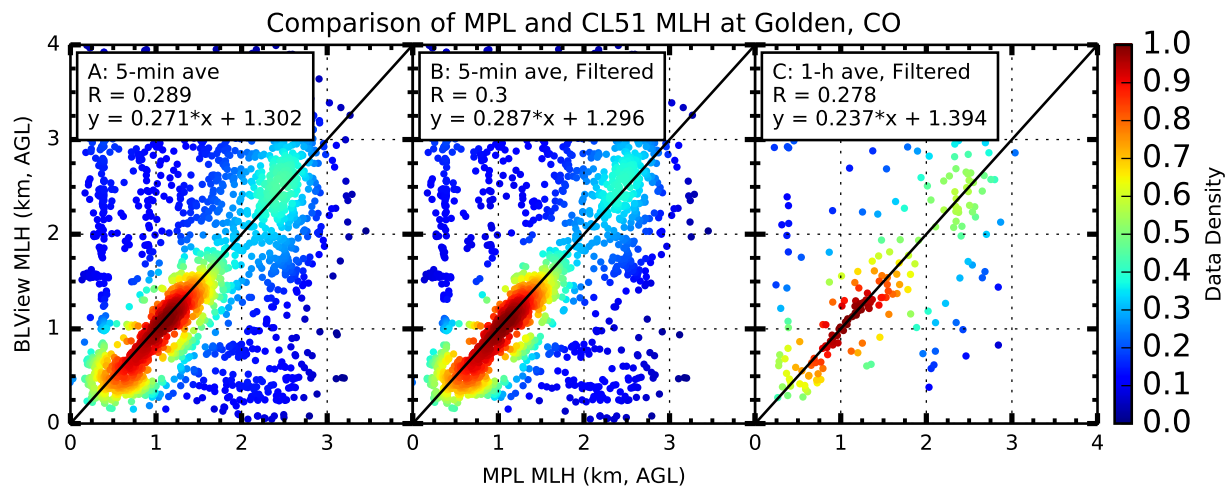
Figure 14. Correlation and data-density plots for the CL51 (BLView processed) and MPL (UMBC algorithm) MLH estimates from Golden, Colorado.
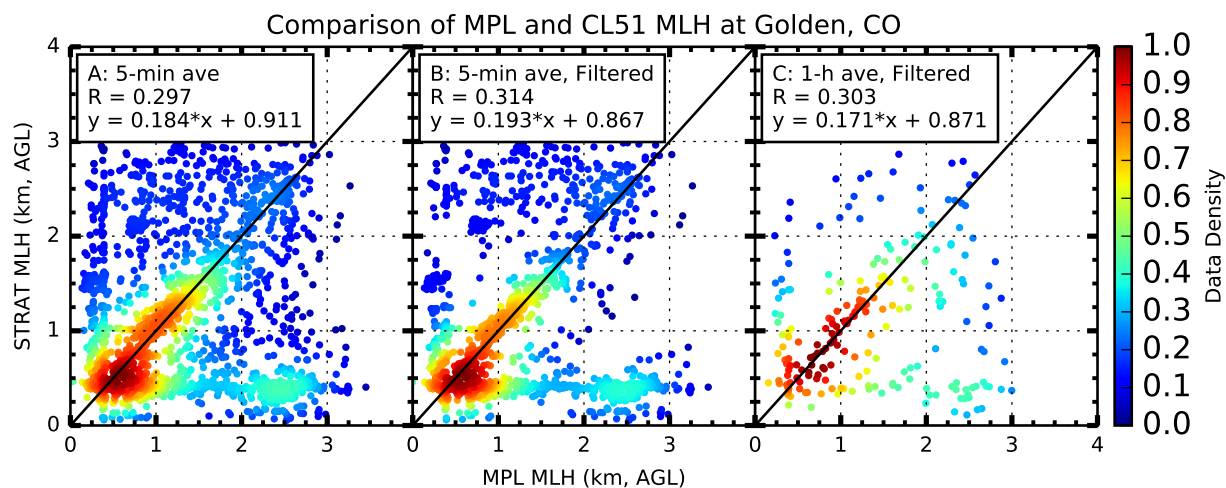
Figure 15. Correlation and data-density plots for the CL51 (STRAT processed) and MPL (UMBC algorithm) MLH estimates from Golden, Colorado.
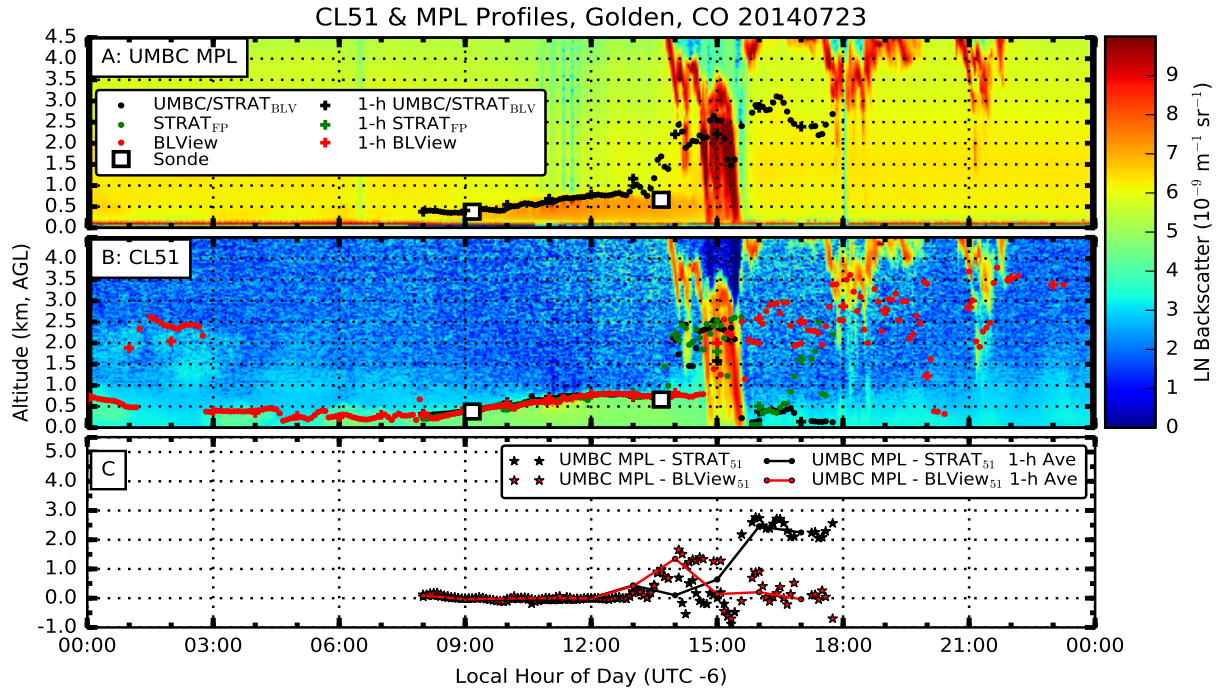
Figure 16. Comparison of the CL51 and UMBC MPL profiles for 23/27-July 2014 at the Golden, Colorado site.

The most commonly used statistical techniques used for comparing two datasets depended on two key assumptions: data were normally distributed and homoscedastic. The CL51 and MPL MLH 5-min averaged datasets were confirmed to be non-normal via the Kolmogorov-Smirnov test and passed Levene's test for homoscedasticity (p-value 0.39). Therefore, similarity between the two corresponding probability distributions was determined using the two-sample Kolmogorov-Smirnov test. It
5 was determined that the 5-min averaged MPL and CL51 datasets were statistically different ($p \ll 0.01$), regardless of filtering and averaging. However, when considering 1-hour averaged data that were filtered to remove data with large relative standard deviations (≥ 0.20) and MLH ≤ 0.5 km, the two datasets were statistically indistinguishable (p 0.8). While we cannot account for the bias induced by these low-altitude MLH values it is quite clear that they significantly influence the intercomparison. Given that this is the first intercomparison of these two instruments and algorithms, it is not surprising that a significant
10 difference was identified in this regime.

6 Conclusions

A CL51-focused intercomparison of different ABL/MLH methodologies was performed at three different sites that experience different meteorological, aerosol, and emission conditions. The CL51 MLH results were compared with ABL from radiosondes at all three locations; as well as an MPL at the Golden, Colorado site.

Two collection methods and processing algorithms were tested for the CL51 MLH calculation. We demonstrated that the data-collection method played an insignificant role in MLH estimation when the datasets were processed using a common algorithm. Furthermore, the choice of processing algorithm played a significant role in MLH estimation. Therefore, we recommend that, for ceilometer and LIDAR networks, a common MLH processing algorithm be employed. Agreement between the different algorithm products might be dictated, to a large degree, by local atmospheric fluctuations and interferential events (e.g. fog), which should be a topic for future investigation.

A total of 53 potential temperature profiles from radiosondes were used to evaluate the CL51. While the 53 radiosondes were spread across 3 sites, this represents a robust data set of soundings. Overall, the met-sonde-derived ABL was higher than the CL51 MLH. Comparison of MLH from the CL51 versus met-sondes shows the CL51 performed best at the CAPABLE research site (non-filtered R = 0.79, filtered R = 0.82), a moderately polluted coastal site primarily influenced by a combination of sulfate and marine aerosols. Both the Golden and BAO-Tower sites are located in cleaner environments than CAPABLE and show good correlation between the CL51 and met-sondes (Golden filtered R = 0.74, BOA non-filtered R = 0.63, filtered R = 0.58) with two early morning sondes at the Golden site strongly influencing the non-filtered correlation (R = -0.28). These two sondes measured a very shallow boundary layer, < 500 m, while the CL51 identified the MLH above 2 km, which was likely due to residual aerosol layers aloft. The lower correlations at the Colorado sites (Golden and BAO-Tower) were likely due to the sites' proximity to the Rocky Mountains. Complex atmospheric flow patterns, which are driven by the Rocky Mountains to the west of the Front Range area, can induce the formation of distinctive dynamic features such as up and downslope flows (Bossert et al., 1989; Bossert and Cotton, 1994; Sullivan et al., 2016; Tripoli and Cotton, 1989). The Golden site likely experienced greater up- and down-slope flows than the BOA-Tower site because of its location along the slope of the mountains and on a mesa. Such local orographic influences likely impacted or challenged the well-mixed assumption required to compare thermodynamic ABL measured via potential temperature and MLH measured via aerosol backscatter in the current study. These influences should be made a consideration in future intercomparisons.

The results of the CL51 versus the UMBC algorithm that was run on MPL data showed low correlation (R = 0.3).However, the majority of coincident MLH observations from both instruments were clustered around the 1:1 line in the regression plots. When data-filtering criteria were applied, the two data sets were statistically indistinguishable (p > 0.8). Additional analysis is planned to further explore the cause of the low correlation. However, the MLH from the CL51 and MPL agree well when there is a well-defined MLH.

Acknowledgements. Funding for this work was provided by the EPA Air, Climate, and Energy Research Program and GEO-CAPE mission studies. T. N. Knepp was funded through the STARS-III contract. Although this paper has been reviewed by the EPA and approved for publication, it does not necessarily reflect EPA policies or views. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Banta, R.: Daytime Boundary-Layer Evolution over Mountainous Terrain .1. Observations of the Dry Circulations, Monthly Weather Review, 112, 340–356, doi:10.1175/1520-0493(1984)112<0340:DBLEOM>2.0.CO;2, 1984.

Bossert, J. and Cotton, W.: Regional-Scale Flows in Mountainous Terrain 1. A Numerical and Observational Comparison, Monthly Weather Review, 122, 1449–1471, doi:10.1175/1520-0493(1994)122<1449:RSFIMT>2.0.CO;2, 1994.

Bossert, J., Sheaffer, J., and Reiter, E.: Aspects of Regional-Scale Flows in Mountainous Terrain, Journal of Applied Meteorology, 28, 590–601, doi:10.1175/1520-0450(1989)028<0590:AORSFI>2.0.CO;2, 1989.

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

Coen, M. C., Praz, C., Haefele, A., Ruffieux, D., Kaufmann, P., and Calpini, B.: Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model, Atmospheric Chemistry and Physics, 14, 13 205–13 221, doi:10.5194/acp-14-13205-2014, 2014.

Compton, J. C., Delgado, R., Berkoff, T. A., and Hoff, R. M.: Determination of Planetary Boundary Layer Height on Short Spatial and Temporal Scales: A Demonstration of the Covariance Wavelet Transform in Ground-Based Wind Profiler and Lidar Measurements, Journal of Atmospheric and Oceanic Technology, 30, 1566–1575, doi:10.1175/JTECH-D-12-00116.1, 2013.

Deshler, T., Mercer, J. L., Smit, H. G. J., Stubi, R., Levrat, G., Johnson, B. J., Oltmans, S. J., Kivi, R., Thompson, A. M., Witte, J., Davies, J., Schmidlin, F. J., Brothers, G., and Sasaki, T.: Atmospheric comparison of electrochemical cell ozonesondes from different manufacturers, and with different cathode solution strengths: The Balloon Experiment on Standards for Ozonesondes, Journal of Geophysical Research-Atmospheres, 113, doi:10.1029/2007JD008975, 2008.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmospheric Measurement Techniques, 7, 4463–4490, doi:10.5194/amt-7-4463-2014, 2014.

Emeis, S. and Schäfer, K.: Remote sensing methods to investigate boundary-layer structures relevant to air pollution in cities, Boundary-Layer Meteorology, 121, 377–385, doi:10.1007/s10546-006-9068-2, 2006.

Emeis, S., Schäfer, K., and Muenkel, C.: Long-term observations of the urban mixing-layer height with ceilometers, in: 14TH International Symposium for the Advancement of Boundary Layer Remote Sensing, edited by Mann, J and Bingol, F and Courtney, M and Jorgensen, HE and Lindelow, P and Mikkelsen, T and Pena, A and Sjoholm, M and Wagner, R, vol. 1 of *IOP Conference Series-Earth and Environmental Science*, doi:10.1088/1755-1307/1/1/012027, 14th International Symposium for the Advancement of Boundary Layer Remote Sensing, Tech Univ Denmark, Copenhagen, DENMARK, JUN 23-25, 2008, 2008a.

Emeis, S., Schäfer, K., and Muenkel, C.: Surface-based remote sensing of the mixing-layer height - a review, Meteorologische Zeitschrift, 17, 621–630, doi:10.1127/0941-2948/2008/0312, DACH 2007 Conference, Hamburg, GERMANY, SEP 10-14, 2007, 2008b.

Haeffelin, M., Angelini, F., Morille, Y., Martucci, G., Frey, S., Gobbi, G. P., Lolli, S., O'Dowd, C. D., Sauvage, L., Xueref-Remy, I., Wastine, B., and Feist, D. G.: Evaluation of Mixing-Height Retrievals from Automatic Profiling Lidars and Ceilometers in View of Future Integrated Networks in Europe, Boundary-Layer Meteorology, 143, 49–75, doi:10.1007/s10546-011-9643-z, 15th International Symposium for the Advancement of Boundary-Layer Remote Sensing (ISARS), Univ Versailles St-Quentin-en-Yvelines, Inst Pierre Simon Laplace (IPSL), Paris, FRANCE, JUN 28-30, 2010, 2012.

Heffter, J.: Air Resources Laboratories atmospheric transport and dispersion model (ARL-ATAD), NOAA, 1980.

Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M., and Abuhassan, N.: NO2 column amounts from ground-based Pandora and MF-DOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, Journal of Geophysical Research-Atmospheres, 114, doi:10.1029/2009JD011848, 2009.

Hirsikko, A., O'Connor, E. J., Komppula, M., Korhonen, K., Pfuller, A., Giannakaki, E., Wood, C. R., Bauer-Pfundstein, M., Poikonen, A., Karppinen, T., Lonka, H., Kurri, M., Heinonen, J., Moisseev, D., Asmi, E., Aaltonen, V., Nordbo, A., Rodriguez, E., Lihavainen, H., Laaksonen, A., Lehtinen, K. E. J., Laurila, T., Petaja, T., Kulmala, M., and Viisanen, Y.: Observing wind, aerosol particles, cloud and precipitation: Finland's new ground-based remote-sensing network, Atmospheric Measurement Techniques, 7, 1351–1375, doi:10.5194/amt-7-1351-2014, 2014.

Holben, B., Eck, T., Slutsker, I., Tanre, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET - A federated instrument network and data archive for aerosol characterization, Remote Sensing of Environment, 66, 1–16, doi:10.1016/S0034-4257(98)00031-5, 1998.

Johnson, B., Oltmans, S., Vomel, H., Smit, H., Deshler, T., and Kroger, C.: Electrochemical concentration cell (ECC) ozonesonde pump efficiency measurements and tests on the sensitivity to ozone of buffered and unbuffered ECC sensor cathode solutions, Journal of Geophysical Research-Atmospheres, 107, doi:10.1029/2001JD000557, 2002.

Knepp, T., Pippin, M., Crawford, J., Chen, G., Szykman, J., Long, R., Cowen, L., Cede, A., Abuhassan, N., Herman, J., Delgado, R., Compton, J., Berkoff, T., Fishman, J., Martins, D., Stauffer, R., Thompson, A., Weinheimer, A., Knapp, D., Montzka, D., Lenschow, D., and Neil, D.: Estimating surface NO2 and SO2 mixing ratios from fast-response total column observations and potential application to geostationary missions, Journal of Atmospheric Chemistry, pp. 1–26, doi:10.1007/s10874-013-9257-6, http://dx.doi.org/10.1007/s10874-013-9257-6, 2015.

Kotthaus, S., O'Connor, E., Münkel, C., Charlton-Perez, C., Haeffelin, M., Gabey, A. M., and Grimmond, C. S. B.: Recommendations for processing atmospheric attenuated backscatter profiles from Vaisala CL31 ceilometers, Atmospheric Measurement Techniques, 9, 3769–3791, doi:10.5194/amt-9-3769-2016, http://www.atmos-meas-tech.net/9/3769/2016/, 2016.

Lamsal, L. N., Martin, R. V., van Donkelaar, A., M., S., Celarier, E. A., Bucsela, E., Dunlea, E. J., and Pinto, J. P.: Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument, Journal of Geophysical Research-Atmospheres, 113, doi:10.1029/2007JD009235, 2008.

Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W. H., Pickering, K. E., Bucsela, E. J., Gleason, J. F., Martin, R. V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Szykman, J. J., and Knepp, T. N.: Evaluation of OMI operational standard NO2 column retrievals using in situ and surface-based NO2 observations, Atmospheric Chemistry and Physics, 14, 11 587–11 609, doi:10.5194/acp-14-11587-2014, 2014.

Marsik, F., Fischer, K., McDonald, T., and Samson, P.: Comparison of Methods for Estimating Mixing Height used during the 1992 Atlanta Field Intensive, Journal of Applied Meteorology, 34, 1802–1814, doi:10.1175/1520-0450(1995)034<1802:COMFEM>2.0.CO;2, International Specialty Conference on Regional Photochemical Measurement and Modeling Studies, SAN DIEGO, CA, NOV 08-12, 1993, 1995.

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

Martins, D. K., Stauffer, R. M., Thompson, A. M., Knepp, T. N., and Pippin, M.: Surface ozone at a coastal suburban site in 2009 and 2010: Relationships to chemical and meteorological processes, Journal of Geophysical Research-Atmospheres, 117, doi:10.1029/2011JD016828, 2012.

Miloshevich, L., Paukkunen, A., Vomel, H., and Oltmans, S.: Development and validation of a time-lag correction for Vaisala radiosonde humidity measurements, Journal of Atmospheric and Oceanic Technology, 21, 1305–1327, doi:10.1175/1520-0426(2004)021<1305:DAVOAT>2.0.CO;2, 2004.

Morille, Y., Haeffelin, M., Drobinski, P., and Pelon, J.: STRAT: An automated algorithm to retrieve the vertical structure of the atmosphere from single-channel lidar data, Journal of Atmospheric and Oceanic Technology, 24, 761–775, doi:10.1175/JTECH2008.1, 2007.

Nash, J., Smout, R., Oakley, T., Pathack, B., and Kurnosenko, S.: WMO Intercomparison of Radiosonde Systems Vacoas, Mauritius, 2-25 February 2005, Report WMO/TD-No. 1303, World Meteorol. Organ., Geneva, Switzerland, 2006.

Nash, J., Oakley, T., Vömel, H., and Wei, L.: WMO Intercomparison of High Quality Radiosonde Systems Yangjiang, China, 12 July - 3 August 2010, Report WMO/TD-No. 1580, World Meteorol. Organ., Geneva, Switzerland, 2011.

NRC: Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks, in: Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks, pp. 1–234, Natl Academies Press, 2101 Constitution Ave, Washington, DC 20418 USA, 2009.

NSTC: National Plan for Civil Earth Observations, Tech. rep., Executive Office of the President, https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/national_plan_for_civil_earth_observations_-_july_2014.pdf, 2014.

Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linne, H., Ansmann, A., Boesenberg, J., D'Amico, G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D., and Wiegner, M.: EARLINET: towards an advanced sustainable European aerosol lidar network, Atmospheric Measurement Techniques, 7, 2389–2409, doi:10.5194/amt-7-2389-2014, 2014.

Petritoli, A., Bonasoni, P., Giovanelli, G., Ravegnani, F., Kostadinov, I., Bortoli, D., Weiss, A., Schaub, D., Richter, A., and Fortezza, F.: First comparison between ground-based and satellite-borne measurements of tropospheric nitrogen dioxide in the Po basin, Journal of Geophysical Research-Atmospheres, 109, doi:10.1029/2004JD004547, 2004.

Piters, A. J. M., Boersma, K. F., Kroon, M., Hains, J. C., Van Roozendael, M., Wittrock, F., Abuhassan, N., Adams, C., Akrami, M., Allaart, M. A. F., Apituley, A., Beirle, S., Bergwerff, J. B., Berkhout, A. J. C., Brunner, D., Cede, A., Chong, J., Clemer, K., Fayt, C., Frieß, U., Gast, L. F. L., Gil-Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Großmann, K., Hemerijckx, G., Hendrick, F., Henzing, B., Herman, J., Hermans, C., Hoexum, M., van der Hoff, G. R., Irie, H., Johnston, P. V., Kanaya, Y., Kim, Y. J., Baltink, H. K., Kreher, K., de Leeuw, G., Leigh, R., Merlaud, A., Moerman, M. M., Monks, P. S., Mount, G. H., Navarro-Comas, M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., du Piesanie, A., Pinardi, G., Puentedura, O., Richter, A., Roscoe, H. K., Schönhardt, A., Schwarzenbach, B., Shaiganfar, R., Sluis, W., Spinei, E., Stolk, A. P., Strong, K., Swart, D. P. J., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Whyte, C., Wilson, K. M., Yela, M., Yilmaz, S., Zieger, P., and Zhou, Y.: The Cabauw Intercomparison campaign for Nitrogen Dioxide measuring Instruments (CINDI): design, execution, and early results, Atmospheric Measurement Techniques, 5, 457–485, doi:10.5194/amt-5-457-2012, 2012.

Scarino, A. J., Obland, M. D., Fast, J. D., Burton, S. P., Ferrare, R. A., Hostetler, C. A., Berg, L. K., Lefer, B., Haman, C., Hair, J. W., Rogers, R. R., Butler, C., Cook, A. L., and Harper, D. B.: Comparison of mixed layer heights from airborne high spectral resolution lidar, ground-based measurements, and the WRF-Chem model during CalNex and CARES, Atmospheric Chemistry and Physics, 14, 5547–5560, doi:10.5194/acp-14-5547-2014, 2014.

Schäfer, K., Wagner, P., Emeis, S., Jahn, C., Münkel, C., and Suppan, P.: Mixing layer height and air pollution levels in urban area, in: Remote Sensing of Clouds and the Atmosphere XVII; and LIDAR Technologies, Techniques, and Measurements for Atmospheric Remote Sensing VIII, edited by Kassianov, EI and Comeron, A and Picard, RH and Schafer, K and Singh, UN and Pappalardo, G, vol. 8534 of *Proceedings of SPIE*, SPIE; SELEX GALILEO; THALES, doi:10.1117/12.974328, 2012.

Schäfer, K., Ling, H., Münkel, C., and Emeis, S.: Long-term study of air urban quality together with mixing layer height, in: Remote Sensing of Clouds and the Atmosphere XVIII; and Optics in Atmospheric Propagation and Adaptive Systems XVI, edited by Comeron, A and Kassianov, EI and Schafer, K and Stein, K and Gonglewski, JD, vol. 8890 of *Proceedings of SPIE*, SPIE, doi:10.1117/12.2027045, 2013.

Schween, J. H., Hirsikko, A., Löhnert, U., and Crewell, S.: Mixing-layer height retrieval with ceilometer and Doppler lidar: from case studies to long-term assessment, Atmospheric Measurement Techniques, 7, 3685–3704, doi:10.5194/amt-7-3685-2014, 2014.

Scott, D.: Multivariate Density Estimation: Theory, Practice, and Visualization, A Wiley-interscience publication, John Wiley & Sons, 1992.

Silverman, B.: Density Estimation for Statistics and Data Analysis, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 1986.

Smit, H.: Quality Assurance and Quality Control for Ozonesonde Measurements in GAW, Report, World Meteorol. Organ., Geneva, Switzerland, 2013.

Sokol, P., Stachlewska, I. S., Ungureanu, I., and Stefan, S.: Evaluation of the boundary layer morning transition using the CL-31 ceilometer signals, Acta Geophysica, 62, 367–380, doi:10.2478/s11600-013-0158-5, 2014.

Spinhirne, J.: Micropulse LIDAR, IEEE Transactions on Geoscience and Remote Sensing, 31, 48–55, doi:10.1109/36.210443, 1993.

Stauffer, R. M., Morris, G. A., Thompson, A. M., Joseph, E., Coetzee, G. J. R., and Nalli, N. R.: Propagation of radiosonde pressure sensor errors to ozonesonde measurements, Atmospheric Measurement Techniques, 7, 65–79, doi:10.5194/amt-7-65-2014, 2014.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Atmospheric Sciences Library, Springer, 1988.

Sullivan, J. T., McGee, T. J., Langford, A. O., Alvarez, II, R. J., Senff, C. J., Reddy, P. J., Thompson, A. M., Twigg, L. W., Sumnicht, G. K., Lee, P., Weinheimer, A., Knote, C., Long, R. W., and Hoff, R. M.: Quantifying the contribution of thermally driven recirculation to a high-ozone event along the Colorado Front Range using lidar, Journal of Geophysical Research-Atmospheres, 121, 10 377–10 390, doi:10.1002/2016JD025229, 2016.

Tripoli, G. and Cotton, W.: Numerical Study of an Observed Orogenic Mesoscale Convective System 2. Analysis of Governing Dynamics, Monthly Weather Review, 117, 305–328, doi:10.1175/1520-0493(1989)117<0305:NSOAOO>2.0.CO;2, 1989.

Welton, E., Voss, K., Gordon, H., Maring, H., Smirnov, A., Holben, B., Schmid, B., Livingston, J., Russell, P., Durkee, P., Formenti, P., and Andreae, M.: Ground-based lidar measurements of aerosols during ACE-2: instrument description, results, and comparisons with other ground-based and airborne measurements, Tellus Series B-Chemical and Physical Meteorology, 52, 636–651, doi:10.1034/j.1600-0889.2000.00025.x, 2000.

Wiegner, M., Madonna, F., Binietoglou, I., Forkel, R., Gasteiger, J., Geiß, A., Pappalardo, G., Schäfer, K., and Thomas, W.: What is the benefit of ceilometers for aerosol remote sensing? An answer from EARLINET, Atmospheric Measurement Techniques, 7, 1979–1997, doi:10.5194/amt-7-1979-2014, 2014.