Doppler Lidar System Comparison using Discover AQ Data

Qin Liu[1]*, Thomas Rieutord[2], Alan Brewer[3], Aditya Choukulkar[3], Ruben Delgado[4]

*[1]University of Maryland, Baltimore County, Baltimore, MD 21250, USA, *Email: qinliu1@umbc.edu*

[2]Météo-France, Toulouse, France

[3]National Oceanic and Atmospheric Administration, CO 80305, USA

[4]Joint Center for Earth Systems Technology, Baltimore, MD, 21250, USA

INTRODUCTION

The purpose of this study is to evaluate the planetary boundary layer height retrievals from Doppler wind lidars. Analysis was applied to data collected from the two lidar systems during the July-August 2014 Discover AQ and LUMEX campaigns. This comparison aids applications in air quality and wind energy forecasting.

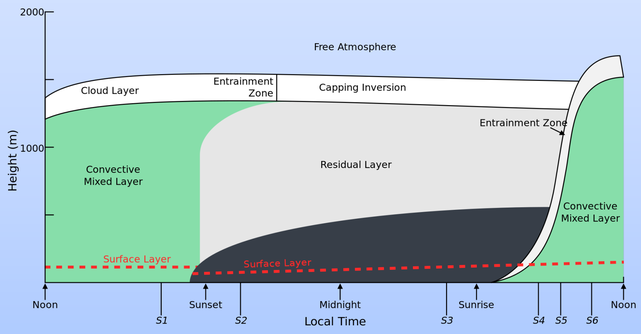

Figure 1: Change of The Planetary Boundary Layer Height

The planetary boundary layer (PBL) of the atmosphere is directly influenced by the presence of the earth's surface, making the height of the PBL change over time due to factors such as temperature, humidity, and aerosols concentration. Doppler Lidar data provide information such as wind speed, and direction, and back-scattered intensity in space and time. The Lidar scans include VAD (velocity azimuth display) scans with wind speed and direction, bowtie scans and vertical staring with velocity variance and range-corrected intensity. Three numerical analysis methods will be tested on the collected Lidar data.

METHODOLOGY

Peak Detection Method

Using Haar Wavelet Transform

$$\psi_H\left(\frac{z-b}{a}\right) = \begin{cases} -1, & \text{if} \quad b-\frac{a}{2} \leqslant z \leqslant b \\ 1, & \text{if} \quad b \leqslant z \leqslant b+\frac{a}{2} \\ 0, & \text{elsewhere} \end{cases}$$

$$W_f(a,b) = \frac{1}{a}\int_{-\infty}^{\infty} f(z)\psi_H\left(\frac{z-b}{a}\right)dz$$

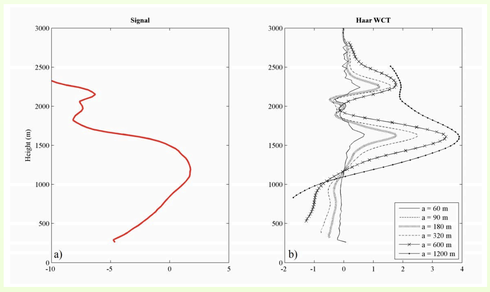

Figure 2: Influence of the dilatation a on the WCT

Mainly used for bt and vs range-corrected intensity profiles (Rci) and horizontal wind speed and direction
Parameters to consider: value of a, b and the choice of thread for continuity test (longest, lowest, and strongest)

Peak-based Thresholding

The BLH is defined as the highest point connected to the ground in the profile

$$\tau_p = \min\left(0.9\,\sigma_w^B + 0.1\,\sigma_w{}^p\ ,\ \sigma_w^0\right)$$

$$Z_{CtG} = \max\left\{z\ /\ \sigma_w(z) > \tau_{p_{CtG}}\right\}$$

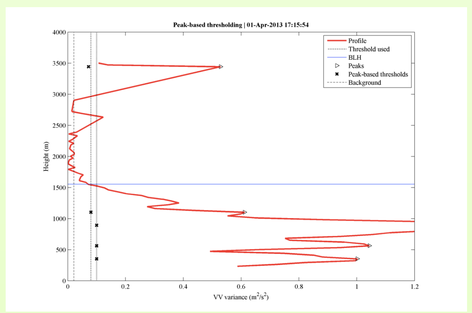

Figure 3: Example of peak-based thresholding

Cluster Analysis

Profiles chosen to apply Cluster Analysis are based on the time range of morning, day, and night.
Morning (bt Rci and var), day (vs Rci and var), and night (bt Rci and var)

$$\text{Classic } K\text{-means} \quad w_k(\mathbf{P}^i) = \begin{cases} 1, & \textit{if} \quad \delta(\mathbf{P}^i, \mathbf{C}^k) = \min\limits_{\ell=1\ldots K}\{\delta(\mathbf{P}^i, \mathbf{C}^\ell)\} \\ 0, & \textit{else} \end{cases}$$

K-means Algorithm

Initialize the seeds (clusters) Calculate the distance from each point to each cluster Assign each point to the closest cluster Redefine the clusters as the centroid of points assigned Repeat the process until the intra-cluster variance no longer decreases

Initial Conditions

Two clusters used, assign top half of the profile to one cluster and lower half to the other cluster
Pros: faster than random seeding, results are consistent every time
Cons: result is not accurate if there's missing data in one single profile

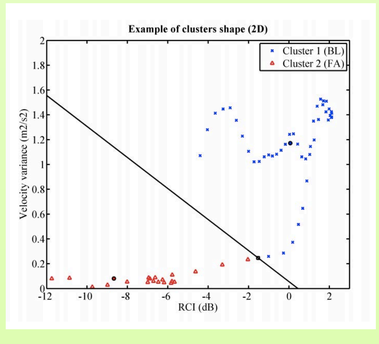

Figure 4: Example of real profile clustering

Convergent Test

Calculate Euclidean distance between each point to the cluster and intra-cluster variance
The algorithm stops when the intra-cluster variance are no longer decreasing
K-means algorithm is an iterative algorithm to create clusters
The BLH is defined as the height where the cluster transitions

RESULTS

A "Typical Day"

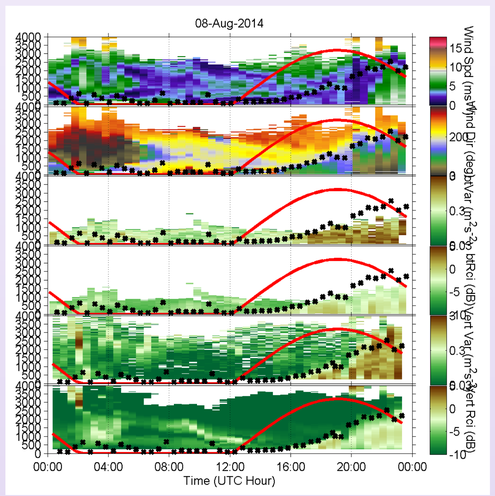

Figure 5: Example of A "Typical Day"

A Bad Day

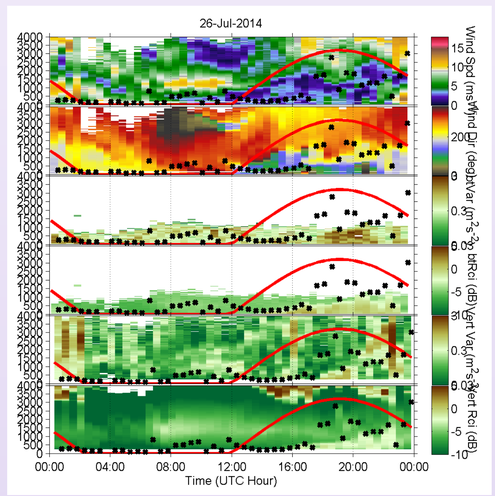

Figure 6: Example of A Bad Day

- Divide the results for each day into six chunks..
- Visually examine each result takes up how many chunks in the BLH estimate graph.
- Successful data is defined when the result exist in the BLH Estimate graph and it fits the background data graph.
- Missing data is defined as the result doesn't exist in the BLH Estimate graph.
- Unsuccessful data is defined when the result exist in the BLH Estimate graph, but doesn't fit the background data graph.

Results of The Visual Examination I

	Estimator	% of S	% of D	% of U	S/(S+U)	(S+U)/(S+D+U)	Methods
wpSpd	Z(W)	0.0637	0.652	0.2843	0.183	0.348	peakdetection
wpTot	Z(W,theta)	0.03432	0.09804	0.8676	0.03805	0.902	Haar wavelet
btVar	Z(sigma h)	0.2037	0.7839	0.01234	0.9429	0.2161	peak threshold
btRci	Z(beta BT)	0.5556	0.09877	0.34567	0.6165	0.9012	Haar wavelet
vsVar	Z(sigma w)	0.04903	0.92646	0.02451	0.6667	0.0735	peak threshold
vsRci	Z(beta VS)	0.4069	0.07353	0.5196	0.4392	0.9265	Haar wavelet
K-means	Zk(beta, sigma)	0.6593	0.1397	0.201	0.7664	0.8603	cluster

* S is for Successful, D is for Data Missing, U is for Unsuccessful. Blue cells highlight the best estimator in the column

- For the Results of the Visual Examination Table, S/(S+U) is successful rate, (S+U)/(S+D+U) is the availability rate
- Cluster Analysis K-means result has the highest successful percentage and very high successful rate and availability rate.
- Peak threshold methods has highest successful rate but low availability rate.
- Haar Wavelet Transform method has highest availability rate but low successful rate.

Even though there are large percentage of missing data on the bt and vs variance profiles, they have relatively high successful rate. K-means (Cluster Analysis) result has the highest successful percentage since it combines the best result from each data scans based on time range of a day. wind data sets are not good estimates for boundary layer, but good for wind shear at the inversion level (LLJ). There is no wind structure that could identify the boundary layer.

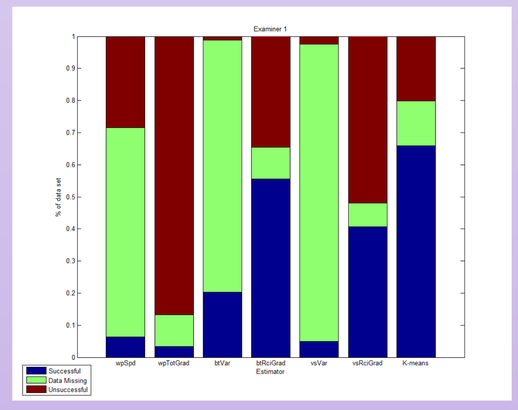

Figure 7: Example of A Visual Examination

FUTURE WORK

After analyzing the data collected by the High Resolution Doppler Lidar (HRDL) during the Discover AQ campaign, the results show that the Cluster Analysis is an effective method for determining the planetary boundary layer height compared to the Peak Detection method and Peak-based Thresholding. These methods will also be applied to a case study on a "typical day" and statistical analysis of all available days for Leosphere data during the LUMEX campaign. A visual examination will again be used to assess and compare the successful rate and availability rate.

ACKNOWLEDGEMENTS

NOAA Office of Education/Educational Partnership Program
NOAA ESRL, NOAA-CREST