Supplement of Atmos. Meas. Tech., 13, 6675–6689, 2020
https://doi.org/10.5194/amt-13-6675-2020-supplement

Atmospheric Measurement Techniques
Open Access

Supplement of

A novel lidar gradient cluster analysis method of nocturnal boundary layer detection during air pollution episodes

Yinchao Zhang et al.

Correspondence to: Siying Chen (csy@bit.edu.cn) and Su Chen (su_chen@bit.edu.cn)

Section 1.

Table 1 Dataset of BIT-lidar

Sample number	Date	Lidar Dataset	Daily Averaged AQI	Daily Averaged $PM_{2.5}$	Daily Averaged PM_{10}	Air quality
1	2016/12/17	18:15-06:10 (next day)	246	196	224	Heavily polluted
2	2016/12/18	18:45-06:15 (next day)	271	221	239	Heavily polluted
3	2016/12/19	18:00-06:29 (next day)	273	223	253	Heavily polluted
4	2016/12/20	17:41-06:26 (next day)	418	376	415	Severely polluted
5	2016/12/21	17:44-06:04 (next day)	431	396	415	Severely polluted
6	2016/12/22	17:20-07:40 (next day)	79	58	77	Severely polluted -> Good
7	2016/12/23	18:08-23:24	62	34	65	Good -> Lightly polluted
8	2016/12/26	19:01-22:10	160	122	152	Moderately polluted
9	2016/12/28	18:39-22:20	87	64	84	Severely polluted -> Good
10	2016/12/29	19:29-22:07	88	65	94	Good -> Severely polluted
11	2016/12/30	17:45-09:00 (next day)	262	212	255	Heavily polluted
12	2016/12/31	17:30-07:00(next day)	351	301	356	Severely polluted
13	2017/1/1	18:00-09:00 (next day)	470	454	512	Severely polluted
14	2017/1/2	17:00-9:00 (next day)	248	198	190	Heavily polluted
15	2017/1/3	18:17-06:25 (next day)	348	311	339	Severely polluted
16	2017/1/4	18:31-06:10 (next day)	389	364	389	Severely polluted
17	2017/1/5	17:30-09:00 (next day)	270	223	242	Heavily polluted
18	2017/1/6	17:20-22:03	248	179	183	Heavily polluted
19	2017/4/6	19:25-22:05	125	98	106	Lightly polluted
20	2017/6/14	19:35-22:25	123	37	60	Lightly polluted
21	2017/6/15	19:23-22:26	155	38	95	Moderately polluted
22	2017/6/19	19:00-22:10	154	37	70	Moderately polluted
23	2017/6/20	19:22-22:30	156	32	64	Moderately polluted
24	2017/6/28	20:02-22:37	189	98	125	Moderately polluted
25	2017/6/29	19:18-22:03	139	52	77	Lightly polluted
26	2017/7/10	19:32-22:33	136	25	55	Lightly polluted
27	2017/7/11	19:45-22:47	208	48	94	Heavily polluted
28	2017/7/12	19:53-22:23	191	74	132	Moderately polluted
29	2017/8/8	20:00-21:26	162	29	69	Moderately polluted
30	2017/8/15	20:05-21:13	106	55	75	Lightly polluted
31	2017/8/21	19:48-21:58	133	72	94	Lightly polluted
32	2017/9/7	19:59-21:59	150	45	74	Lightly polluted
33	2017/9/8	19:53-05:19 (next day)	127	34	72	Lightly polluted

34	2017/9/13	19:51-22:30	123	74	125	Lightly polluted
35	2017/9/14	18:56-22:30	112	85	118	Lightly polluted
36	2017/9/30	22:19-03:16 (next day)	104	71	96	Lightly polluted
37	2017/10/26	18:15-06:42 (next day)	202	123	133	Heavily polluted
38	2017/10/27	18:56-3:57 (next day)	221	120	133	Heavily polluted
39	2017/11/21	18:15-21:42	158	119	158	Moderately polluted

Section 2. The uncertainty of the cluster analysis of the gradient method(CA-GM)

The testing with the real signal are shown below.

Use the RCS(z) signal, and randomly noised $\mathrm{RCS}^{noised}(z)$ by the expression:

$$\mathrm{RCS}^{noised}(z) = \mathrm{RCS}(z) + [\alpha \times \chi(z)] \quad \text{(S2-1)}$$

Where $\chi(z)$ is the random noise function taking values between 0 and 1, z is the height, and α is a varying parameter as introduced in Eq (S2-1) to produces different levels of noise.

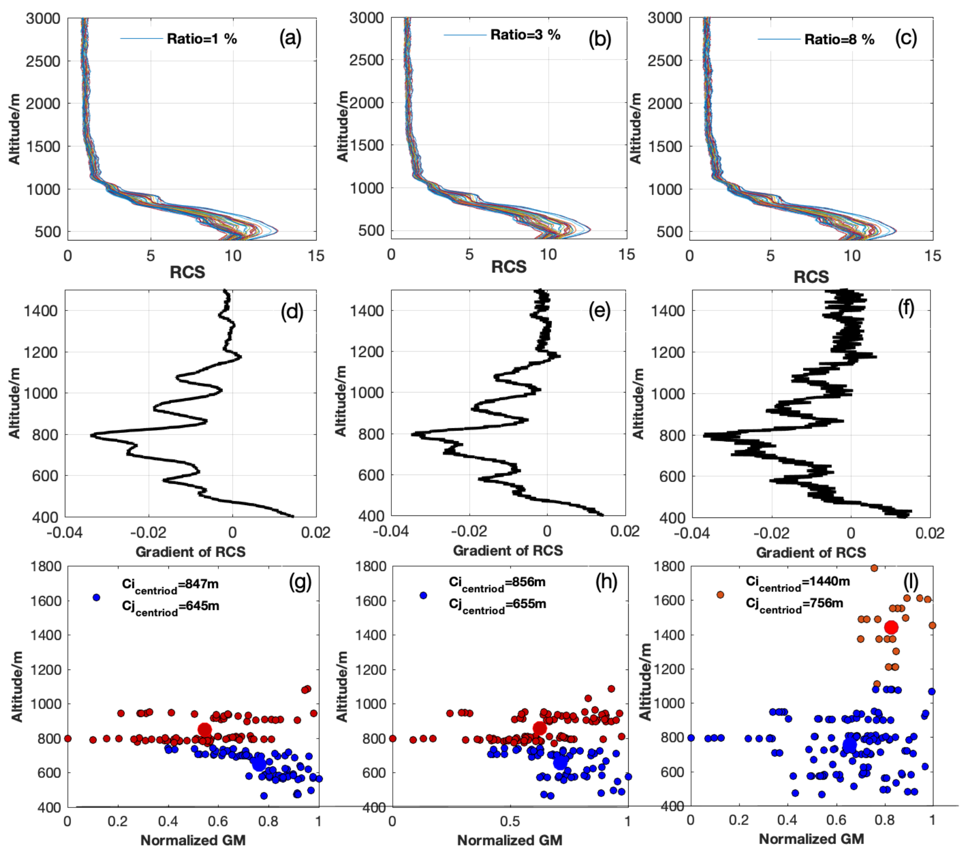

Figure S2-1. The real lidar RCS for the heavily polluted case (17 Dec 2016 20:00-21:00 LST). (a-c) three noise level cases. (d-f) the gradient of RCS. (g-i) the first weighted k-means clustering with the distribution of clusters in blue and red, respectively, The results are shown by red and blue solid points, and their the centroids are represented by larger points of the same colour. The corresponding centroid is marked at top of figure.

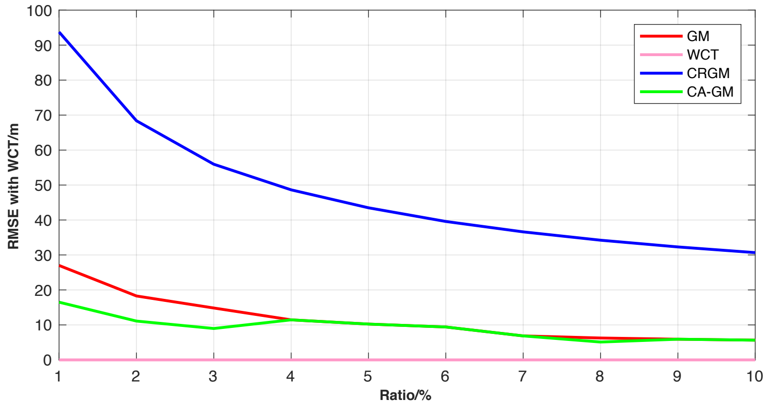

Figure S2-2 RMSE between the WCT and the other three algorithms (GM,CRGM and CA-GM)

As a result of the figure S2-2 shows, the CA-GM has less RMSE than GM at the ratio of 1%-4%. The figure S2-1 (g-h) shows similar groups in different range of noise affection. However, the clustering changes at the results of S2-1(i). Due to the noise distribution of the signal, the centriod of the cluster will get higher and lose the ability to restrict the changes of GM.

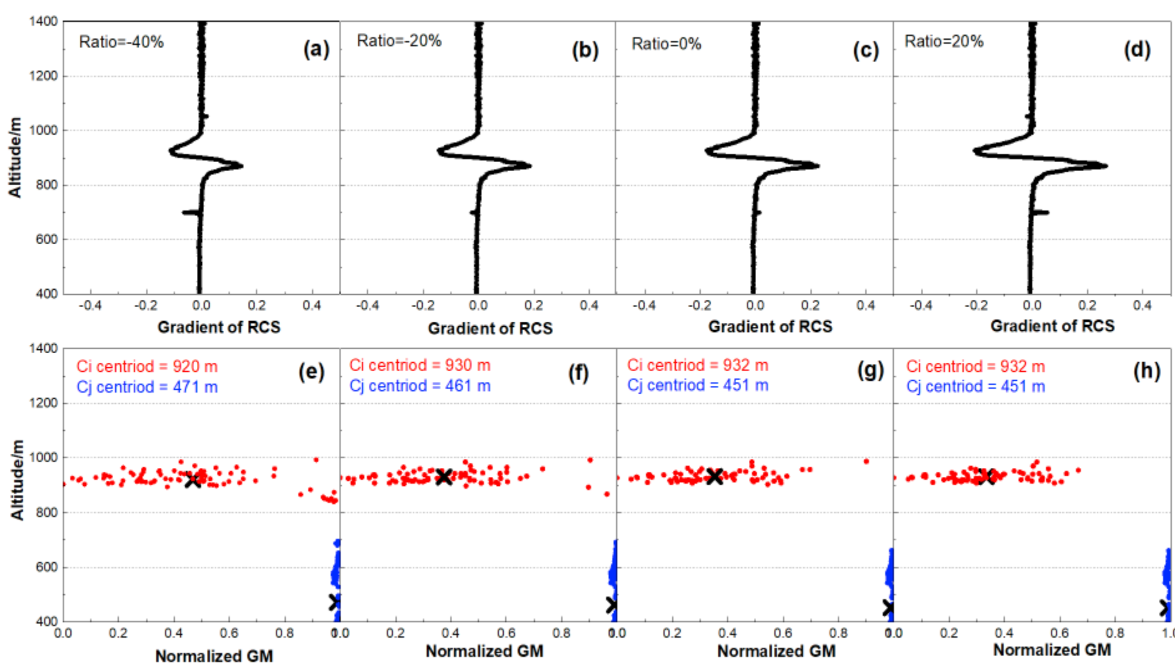

Figure S2-3. The real lidar RCS for the cloud case (5 Jan 2016 00:00-1:00 LST). (a-d) the different ratio of strength of the cloud layer intensity. (e-h) the first weighted k-means clustering with the distribution of clusters in blue and red, respectively. The corresponding centroid is marked at top of figure.

Add the signal of the cloud layer on the raw data, the ratio of the intensity for cloud layer changes from -40% to 40%. As the figure shown the first k-means clustering in figure S2-3(e-h),the intensity of the cloud layer will not influence the CA-GM.

In summary, these results indicate that the degree of estimation of the NBL top by applying CA is weaken affected by the signal noise. In fact, the discrete point in RCS gradient distribution increase the fluctuation of the NBLH. However, CA determines the NBL by taking into account the overall set of observations of a given point, thus decreases the dependence of the method on the RCS values in single moment. The intensity of the CLs changes ±40% and will not affect the cluster of the CA-GM, it can be significant stratified due to the relative significantly signal difference on the backscatter signal. As for the EALs, the strict threshold will defined the EALs accurately. Therefore, the CA-GM approach is able to accurately obtain the NBLH with the effect of noise, cloud layers and elevated aerosol layers.