Reply to Referee #3

The authors would like to thank the editor and reviewers for their constructive and valuable comments. Accordingly, point-by-point answer is given as follows:

(1) ***Since the existing of gravity wave is an important factor for the method, should you declare the application limitation of the method you developed?***

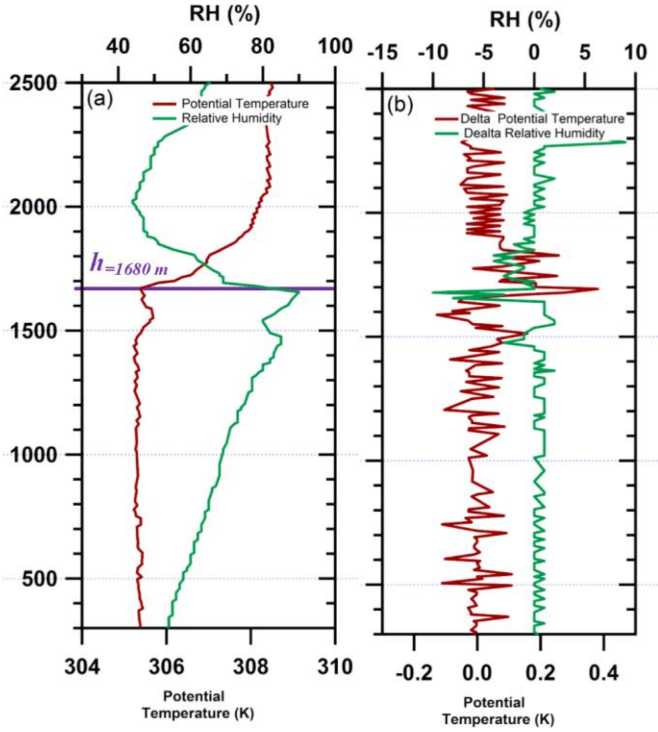
The favorable conditions for generation and maintaining gravity waves over Beijing (Gong et al., 2013; Gibert et al., 2011) provide us with great opportunity to develop the new BLH retrieval algorithm based on long-term Lidar observations. Such new method is thus supposed to be effective for the BLH retrieval from Lidar under similar condition of gravity wave impact over other areas. However, this method is out of application under very shallow nocturnal boundary layer below the useful Lidar signal (before the overlap reaches 1), especially in winter night. In addition, more applications of the new algorithm under gravity wave impact conditions over various areas worldwide are necessary to properly evaluate its fluctuations and possibly further limitations in the future.

Gibert, F., Arnault, N., Cuesta, J., Plougonven, R., and Flamant, P. H.: Internal gravity waves convectively forced in the atmospheric residual layer during the morning transition, Quart. J. Roy. Meteor. Soc., 137, 1610-1624, 10.1002/qj.836, 2011.

Gong, S., Yang, G., Xu, J., Wang, J., Guan, S., Gong, W., and Fu, J.: Statistical characteristics of atmospheric gravity wave in the mesopause region observed with a sodium lidar at Beijing, China, J. Atmos. Sol.-Terr. Phy., 97, 143-151, 10.1016/j.jastp.2013.03.005, 2013.

(2) In general, the data from radio sounding is usually too rough to define boundary height. You might also need to declare what kind of the method you define from radiosondes (same as Stull (1988)?), and in case have multi-levels, how do you define boundary layer height?

The BLH has been defined from the radio sounding based on an elevated inversion in potential temperature (or the height of a significant reduction in air moisture), which is the classic and easy way, widely employed by Stull (1988), Seibert et al (2000) and He et al (2006). Accordingly, intensive radio sounding has been achieved 4 times per day with vertical resolution of ~ 7-10 m. An illustration of BLH determination based on such method is displayed on ReFig.3-1 (14:00 on 9 Aug).



ReFig.3-1 (a) potential temperature and relative humidity at 14:00 Aug.9 (b) the change of potential temperature and relative humidity with altitude

For the cases with multi-layers, the potential temperature presents multi-inversion associated with complex relative humidity profile. As such, wind and wind shear profile method could be applied to define the BLH where the wind shear first becomes less than a detection criterion (Hyun, et al.2005) as described below.

$$\sqrt{(\frac{\partial \overline{U}}{\partial z})^2 + (\frac{\partial \overline{V}}{\partial z})^2} < S_c \quad (2)$$

$\overline{U}$ and $\overline{V}$ are the mean wind components in the east-west and north-south directions, respectively; *Sc* is the detection criterion, Usually, *Sc* is set to be 0.04 s^{-1}.

He, Q. S., Mao, J. T., Chen, J. Y., and Hu, Y. Y.: Observational and modeling studies of urban atmospheric boundary-layer height and its evolution mechanisms, Atmos. Environ., 40, 1064-1077, http://dx.doi.org/10.1016/j.atmosenv.2005.11.016, 2006.

Seibert, P., Beyrich, F., Gryning, S.-E., Joffre, S., Rasmussen, A., and Tercier, P.: Review and intercomparison of operational methods for the determination of the mixing height, Atmos. Environ., 34, 1001-1027, http://dx.doi.org/10.1016/S1352-2310(99)00349-0, 2000.

Stull, R. B.: An introduction to boundary layer meteorology, Springer, 1988. P13

Hyun, Y.-K., Kim, K.-E., and Ha, K.-J.: A comparison of methods to estimate the height of stable boundary layer over a temperate grassland, Agricultural and Forest Meteorology, 132, 132-142, 10.1016/j.agrformet.2005.03.010, 2005.

(3) Page 8, line 33 and Table 2, there are 298 radiosondes in total for the comparison of RMSE and Lidar retrieval method, how many cases in each pollution types? I suggest you need to show in the table.

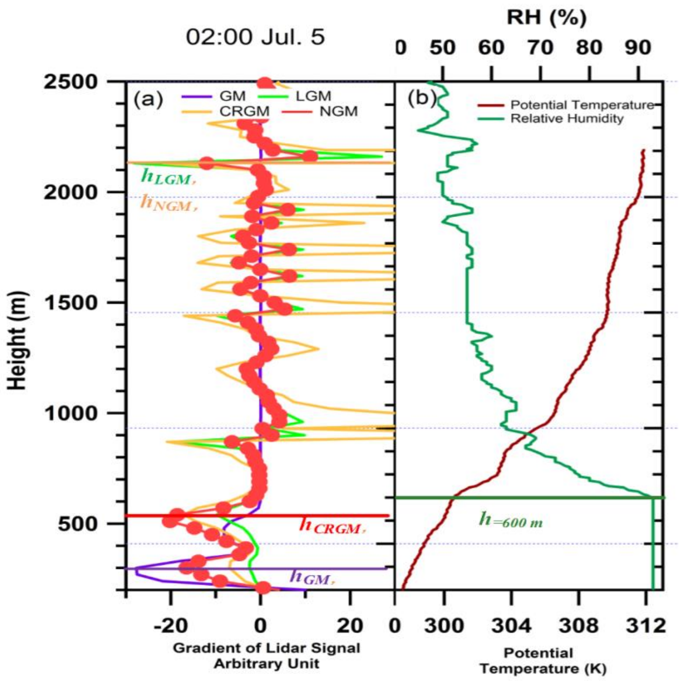
We have updated Table1 as following:

Table 1: Root-Mean-Square Error (RMSE) for each Lidar retrieval method compared with radiosonde measurements and Sample size in each comparison level

$PM_{2.5}$ (μg/m^3)	**CRGM (m)**	**GM (m)**	**LGM (m)**	**NGM (m)**	**Samples / Cases**
0-35	124	124	137	129	114
35-75	123	133	238	227	88
75-115	135	213	320	418	53
115-150	154	310	346	434	19
150-250	137	629	636	643	24

(4) Figure 6, from the case under clean atmosphere (9 Aug.), they are with good performance for all the method of retrieval algorithms, even in the vertical distribution. However, in Figure 7, at low concentration level (green dots, less than 35 ug/m3), the scatter distribution of some cases are diverse (i.e. far away 1:1 line) in different method. Why?

Basically, under low aerosol loading condition (clean condition), the RMSE of the retrieval algorithms is found to be relatively weak (good performance). However, persistent high bias might be noticeable under some exceptional conditions as pointed out by the reviewer. Actually, the algorithms result is determined by moment profile, and in clean condition, some occasional ground floating dust might disturb the moment signal of ideal vertical distribution, resulting in a bias of the algorithm retrieval result. ReFig.3-2 illustrates such a case with $PM_{2.5}$ concentration of 9.5μg/m^3 (clean condition), with large bias between BLH determined by radiosonde and Lidar algorithms. In clear, we believe that observed bias might be due to such occasional floating dust.



ReFig.3-2 (a) the Profiles of CRGM, LGM, GM and NGM, and the corresponding retrieval BLH at 02:00 Jul.5 (b) potential temperature and relative humidity at 02:00 Jul.5