

## ***Interactive comment on “Remote sensing of PM<sub>2.5</sub> during cloudy and nighttime periods using ceilometer backscatter” by Siwei Li et al.***

**Siwei Li et al.**

siweiligm@gmail.com

Received and published: 19 April 2017

We thank Dr. Koelmeijer for providing his comments and sharing his own thoughts which helped to improve the manuscript. This document includes all the comments as well as our responses to every one of them.

This paper describes a method to extract information on PM<sub>2.5</sub> from lidar ceilometer observations. This may improve knowledge of PM<sub>2.5</sub> concentrations at sites where no other PM<sub>2.5</sub> measurement equipment is available. The advantage of using lidar data is that it can be used to determine PM<sub>2.5</sub> also during cloudy or nighttime conditions, in which case methods based on satellite measurements of AOD cannot be used. The paper is interesting as it shows the dependence of the relation between PM<sub>2.5</sub> and lidar backscatter on meteorological variables. In general the paper is well written and

C1

well structured.

Page 1, line 18-19. I have some doubts about the validity of the conclusion reached in the Abstract, line 18-19. The numbers given there seem to be flawed because these correlations pertain to the so-called 'best fit' rather than averaging over the 100 fits that were performed. I would like to replace these correlations by the average correlations of the 100 fits instead of using the results for the best fit. The point is that if one wants to apply the method for a site with no PM<sub>2.5</sub> measurement instrument and only a ceilometer, one does not know what fit-parameters would give the best fit, and these parameters will certainly depend on site and meteorological conditions and aerosol characteristics. Hence, showing best-fit results gives a too optimistic view on the accuracy of ceilometers to estimate PM<sub>2.5</sub> concentrations at sites where no other instruments are available.

Responds: Thanks for the comments and suggestions. We used the average CVR-square out of the 100 fits instead of R-square for the best fitting result in the revised version. They are very close. For daily PM<sub>2.5</sub>, the R-square for the best fitting results are 0.67 and 0.83 at ARM SGP and HUBC sites respectively while the average CV R-square are 0.66 and 0.82 respectively.

Page 5, line 16. Please explain what is meant by "overfitting". Also introduce "best fitting", as results are given in the paper both for the average correlations of the 100 fits as well as for the fit that gives the best fit.

Responds: Thanks for the comments and suggestions. We added the introduction in the revised version. "A complicated model has a potential to be overfitted. The overfitted model describes random error or noise instead of the underlying relationship. To test our model and avoid overfitting, cross-validation are used in this study." The best fitting is just based on the random test data which is only 10% of the total data. The model with the parameters of the "best fitting" out of the 100 cross-validation may be not the "best" model for the whole data. To be consistent with some other studies, we

C2

chose the parameters of the best fitting of the cross-validation to show the correlation between the measured PM2.5 and modeled PM2.5 for all the data. Meanwhile, we presented the average CVR-square out of the 100 cross-validation for each model. In the manuscript, we indicated that in page 9 line 29 "With the parameters (table 1, 2) from the best fitting out of the 100 independent cross-validations (10% of the total data)..."

Page 9, line 1. Correlation coefficient, is R or R\_square meant here?

Response: The correlation coefficient is R at here.

Page 9, figure 4. Is R or R\_square equal to 0.40, 0.46 and 0.29, respectively? On the x-axis, also give the unit of "surface backscatter". Is this the same as the backscattering coefficient as defined in equation 6? If so, please use the same name for the same quantity.

Response: Thanks for the comments. In figure 4, it is R. This figure showed the relationship between the meteorological conditions to PM2.5-to-backscattering ratio. We added the unit of "surface backscatter" in the revised version. The surface backscatter is the integration of backscattering coefficient from surface to the height of 150 m.

Page 12, figure 9/Page 13, figure 10. Is R or R\_square presented, and why is the relation shown using the best fit only? To me this gives a flawed impression to the reader, see comment on the Abstract of this paper.

Response: Thanks for the comments. In the figure 9 and figure 10, it was R-square. In the revised version, we used R instead of R-square to be consistent with previous figures. The best fitting is just based on the random test data which is only 10% of the total data. The model with the parameters of the "best fitting" out of the 100 cross-validation may be not the "best" model for the whole data. To be consistent with some other studies, we chose the parameters of the best fitting of the cross-validation to show the correlation between the measured PM2.5 and modeled PM2.5 for all the data.

C3

Page 13, line 13. "ARM SGP site were" → "ARM SGP site was"

Response: Thanks for the correction. We corrected it in the revised version.

Page 13, line 19. I think it is better to remove this sentence. I do not know by what criterion one can arrive at the conclusion that the regression model works well. The correlations presented in line 17 are based on the best fit, and to my opinion give flawed results (see comment on the Abstract)

Response: Thanks for the comments. We removed this sentence and used average CVR-Square instead of the R-Square of the best fit at here and also in the abstract in the revised version.

Page 14, line 11. "laser power are" → "laser power is"

Response: Thanks for the correction. We corrected it in the revised version.

Page 15, line 24. Also here the best fit results are presented, which I find problematic.

Response: Thanks for the comments. We used average CVR-Square instead of the R-Square of the best fit in the revised version. Actually, they are very close.

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-305, 2016.

C4