Response to the Reviewer #1

First of all, we would like to thank the reviewer for their efforts. The comments and questions have been very helpful for improving the manuscript. Please find below the reviewer comments in black, followed by the author's response in blue.

General comment:
The article presented by Sakai et al. is within the topics of AMT. It is clear in its approach and presents a very detailed validation of a new H2O Raman lidar. In addition, the opportunities offered by the water vapor lidars are relevant considering the increase of the extreme raining events. It deserves to be published. I have some minor corrections and remarks listed in the following.

Thank you for reading our manuscript and giving insightful comments and suggestions. We have read your comments carefully and revised the manuscript in accordance with the suggestions. Please find our point-to-point answers below.

1) Abstract L15: Changed their by the Introduction The requirement for data assimilation is more on the absolute value of the root-mean-square-error, less than 0.4 g/kg in the planetary boundary layer (Weckwerth et al). Biases are more problematic for data assimilation process and may induce large discrepancies.

In accordance with the reviewer's suggestion, we have provided the absolute values of the RMSD (0.98 g/kg) as well as the relative error in the abstract, which is larger than the required value of 0.4 g/kg reported by Weckwerth et al. (1999). However, in a recent review by Wulfmeyer et al. (2015) reported that the noise error lower than 10% and bias error smaller 5% are required for the data assimilation. These values can be translated into <1 g/kg and < 0.5 g/kg for 10 g/kg that is typical value in the lower troposphere. Thus, we have provided these percentage values in the Introduction.

2) Section 2.4: The investigation about the variation of K is a very interesting study. From our experience, it may be due to temperature instabilities in the trailer, although in our case we could not pinpoint whether it was due to PMT gain or filter CWL variations. Maybe the temperature of the air conditioning was set differently during the summer and the fall? The high voltages of the photomultipliers seem to be fixed; some authors vary the PMT gains to adjust for the signal/sky background noise ratio during daytime and nighttime. This means the PMT gain has been optimized for daytime limitations, and that the lidar could be more effective at night. Could you comment? Why was it necessary to adjust the focus? Because of the displacement of the trailer? Is your collimating lens an achromat? If not, it could explain the change of K with the change of focus.

We set the temperature of the air conditioner 23°C during the summer and autumn. The variation of the temperature in the trailer was at most ± 5 K (21-28°C). According to the manufacturers, the temperature variation of the sensitivity of PMT is <0.4%/K (Hamamatsu Photonics, Japan) and that of the filter CWL is < 0.0035 nm/K (FUJITOK, Japan), which corresponds to <6% variation of the effective Raman backscattering cross section ratio (N_2 / H_2O) for ± 5 K variation (Fig. A1). Accordingly, we have added comments on these variations in the revised manuscript.

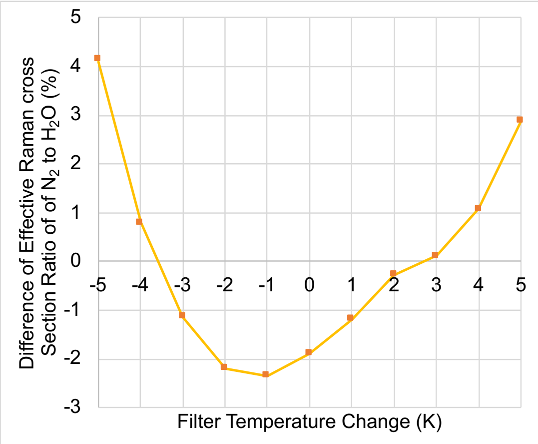

Figure A1. Calculated differential effective Raman backscattering cross section ratio of N_2 to H_2O channels as a function of interference filters used in this study.

As the reviewer has suggested, changing the high voltages of the PMTs during daytime and nighttime is effective to optimize the PMT gain. During the experiment in 2016, we did not change the high voltages (i.e. –1300 V) so that the measurement performance was limited. In particular, the sensitivity of PMT decreased around midday in Summer when the solar zenith angle is high. To improve the performance, we upgraded the lidar control program to automatically change the high voltages during day and night in 2018.

3) Section 2.5: To improve the calibration process, especially for the overlap correction function in the lower layers, tethered balloon or kite can be used as in Totems and Chazette (2016). We are then certain of the location of the reference measurements, and we can renew it at will. The accuracy on w is then better. Totems, J. and Chazette, P.: Calibration of a water vapour Raman lidar with a kite-based humidity sensor, Atmos. Meas. Tech., 9, 1083-1094, doi:10.5194/amt-9-1083-2016, 2016. Could you comment on whether this correction of the overlap factor needs to be re-evaluated at the same time as K when the telescope is re-aligned/re-focused? Rather than PMT inhomogeneity, the incidence on the interference filters may have been modified by the change of focus, which is known to have a large impact.

The use of kite for the lidar calibration and determination of the overlap function is a promising method because it can measure the air close to the lidar. One important issue for use of it is that we need to get permission of the Minister of Land, Infrastructure and Transport if we fly it over a height of 150 m in a densely populated area or near airport in Japan. In accordance with the suggestion, we have added the comment on that and referenced Totems and Chazette's paper in the text.

We agree that the correction of the overlap needs to be re-evaluated when the telescope is re-aligned/refocused. Thank you for the information that change of the telescope's focus has large impact on *K* by modifying the incidence on the interference filters that.

4) In Figure 5, there is a great variability of the observed overlap correction, what can explain this? Lidar noise? Radiosounding error? It may be necessary to distinguish different cases because it is an important point for the robustness of the measurement in the lower tropospheric layers. Can you evaluate or at least comment on the resulting uncertainty on w below 1 km altitude?

We have checked the individual profile of the comparison of *w* between the lidar and radiosounding (Fig. S1) . It is difficult for us to distinguish the reasons for the variability. The possible reasons are 1) the difference of the measurement period (i.e. 20 minutes average for the lidar and approximately 1 second for the radiosonding), 2) the difference of the vertical resolution (i.e. 75 m for the lidar and 20– 300 m (it depends on the significant pressure level interval) for the radiosounding data), and 3) lidar noise. We have added these comments on the sources of the variability in the text. In accordance with the reviewer's suggestion, we have added comments on the uncertainty of the correction based on the standard deviation of the difference between the lidar and radiosounding.

5) Section 3.1.1: Radiosounding errors should also be shown in Figure 6.

We have added the error bars of the radiosonding in Fig. 6.

6) L24-27: The decrease of the water vapor concentration could be seen on the in-situ measurements of weather stations. Perhaps the temporal evolution of one of these measurements should be added. Why would the laser energy have decreased? Is it because of cold, flash lamps and/or damage on optics? The differences with the modeling can be related to local effects and thus to the representativity of the measurement site at the mesoscale. They can also be due to a problem in the assimilation process if it does not integrate well the error matrices.

Because the decrease of the water vapor concentration can be seen in Fig. 9, we would like to retain the figure without showing the temporal evolution of *w* at the surface. Instead, we have added the comment in the text that the monthly mean *w* values decreased from 17 to 4 g/kg at 1000 hPa and from 8 to 1 g/kg at 700 hPa between August and December in 2016.

The primary reason for the decrease in the laser power was the aging of the flash lamp because the emission power decreases as increasing shot number (the lifetime is 20 million shots, or about 3 weeks for the continuous operation). In fact, the laser power increased from 110 mJ/pulse to 220 mJ/pulse after replacing the flash lamp and adjusting angles of second and third harmonic crystals after the experiment on 8 December 2016. We have added the comment in the manuscript.

We agree with you that the differences with the modeling can be related to local effects and thus to the representativeness of the measurement site at the mesoscale. We have added the comment that in the revised manuscript.

7) Section 3.1.2, L11: Typing error on "difference-wsonde"?

We have deleted the word because it is unnecessary.

8) Section 3.1.3: This section should be merged with section 3.1.1.

We have merged Section 3.1.3 with Section 3.1.1 as the reviewer's suggestion. To be consistent with this change, we have changed the order of subsections in Section 3.3.

9) Section 3.2, L11: A ground level in-situ measurement could have helped.

Following the reviewer's suggestion, we used the ground level in-situ measurement of *w* to compute PWV from the lidar data instead of interpolating the lidar-derived w at 0.14 km to the ground level. By this change, the lidar-derived PWV values has slightly changed and thus result of the comparison between RL- and GNSS-derived values also changed (Table 3).

10) Section 3.3.1: It is not so clear whether assimilation is only about radiosounding. Are there no other types of data assimilated, such as spaceborne data? It would be better to show the scatter plot of the radiosounding/LA also.

The LA assimilates the multiple sources, including surface measurements, radiosounding, satellites, and GNSS-derived PWV data, as has been described in Sect. 3 (P5, L15). Because the main purpose of this paper is the validation of the RL system, we would like to show the scatter plot of the radiosounding and LA only in the response (Fig. A2) but not in the original manuscript. We can see in Fig. 2A that there was a negative bias in the LA data.

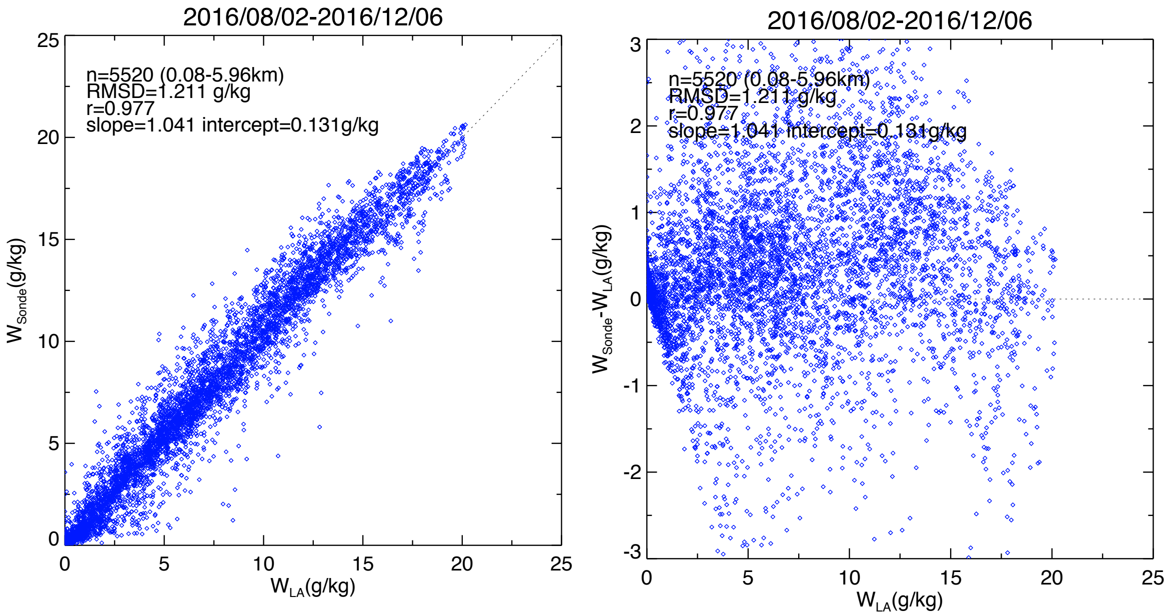

Fig. A2. (Left panel) Scatter plot of w obtained with the LA (w_{LA}) versus w obtained with radiosondes (w_{Sonde}) from 2 August to 6 December 2016. (Right panel) Scatter plot of the difference ($w_{\mathrm{LA}} - w_{\mathrm{Sonde}}$) as a function of w_{LA}.

11) Section 4: Change numbering.
Collected.

We wish to thank the reviewer again for his or her valuable comments.

References:
Hamamatsu Photonics K. K. (2007), Photomultiplier tubes, basics and applications, 3rd Ed., Figures 8-11 and 13-1 (available from https://www.hamamatsu.com/resources/pdf/etd/PMT_handbook_v3aE.pdf).

Weckwerth, T. M., V. Wulfmeyer, R. M. Wakimoto, R. M. Hardesty, J. W. Wilson, and R. M. Banta (1999), NCAR-NOAA lower tropospheric water vapor workshop, *Bull. Am. Meteorol. Soc.*, 80, 2339–2357.

Wulfmeyer, V., R. M. Hardesty, D. D. Turner, A. Behrendt, M. P. Cadeddu, P. Di Girolamo, P. Schlüssel, J. Van Baelen, and F. Zus (2015), A review of the remote sensing of lower tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles, *Rev. Geophys.*, 53, doi:10.1002/2014RG000476.

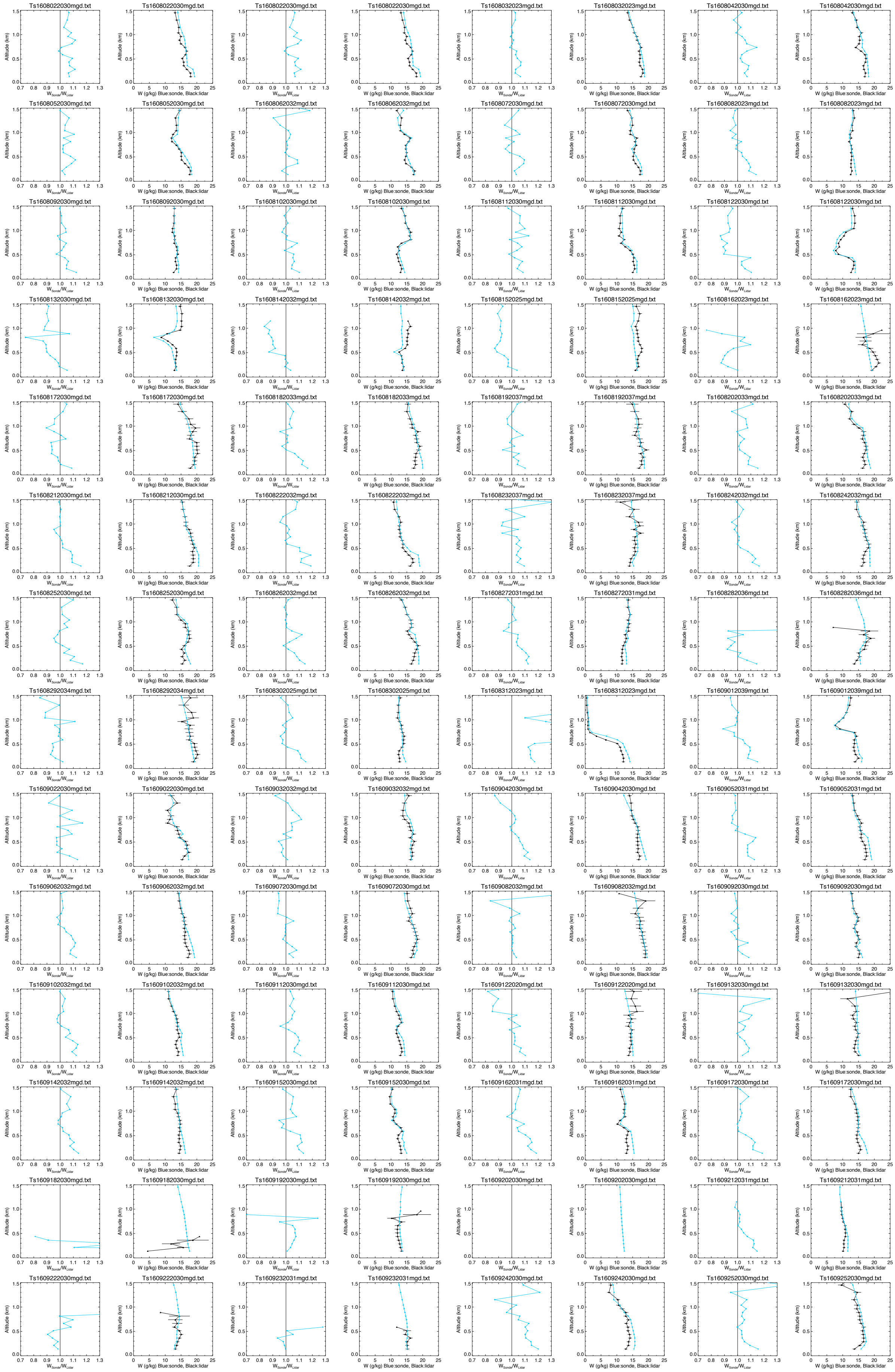

Figure S1. Vertical distribution of the ratio of *w* obtained by radiosonde (light blue) to *w* obtained with the RL system without beam overlap correction (black) from 2 August to 6 December 2016.

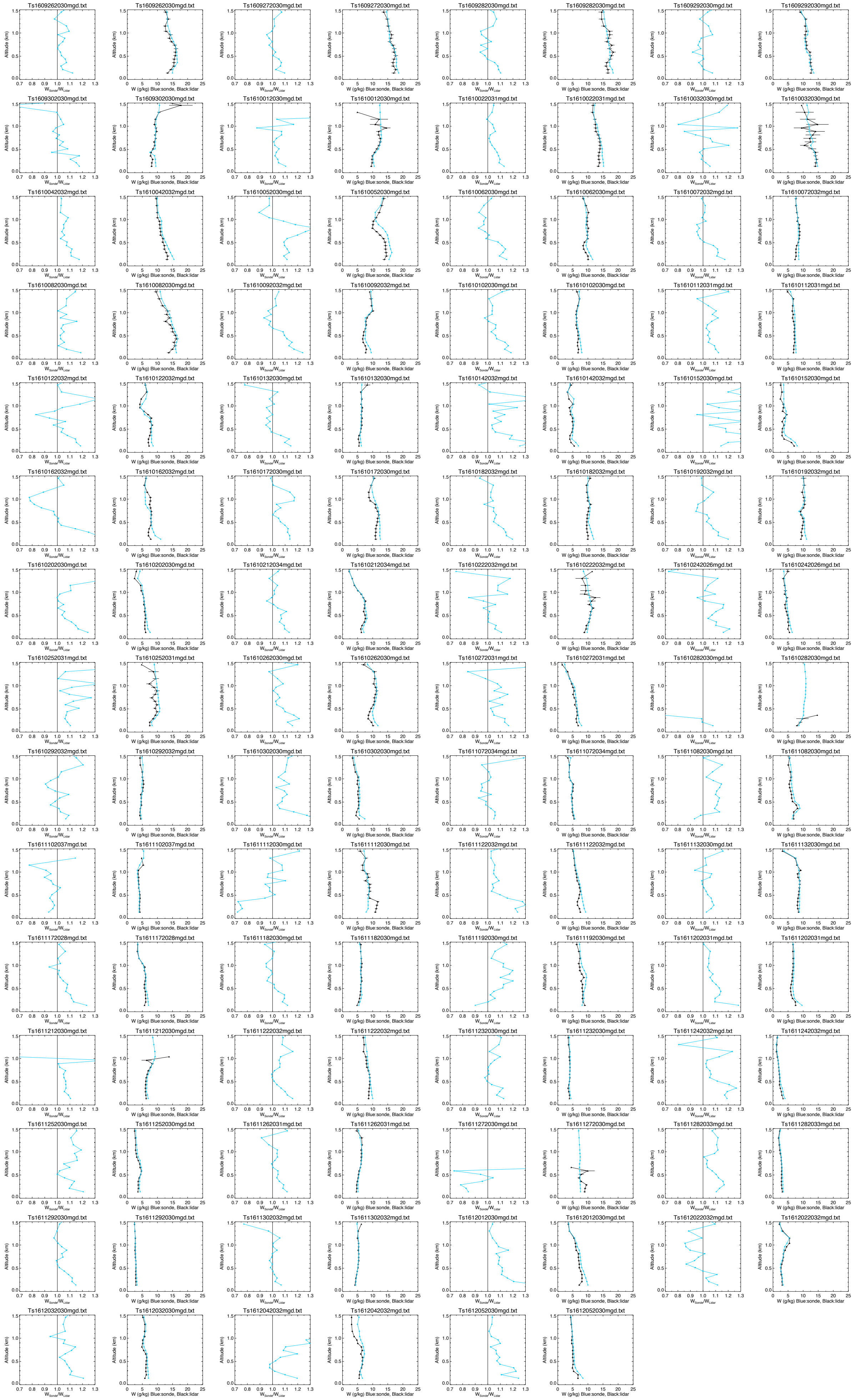

Figure S1. (Contd.)