We would like to thank the two reviewers for devoting time to read our manuscript and provide valuable comments for improving it and increasing its scientific value. We have modified our manuscript following the guidelines given by the two reviewers. Below we answer to each reviewer's comment (RC) separately. The **RC**s are given in **bold**, our replies in plain font and the corresponding *changes in the manuscript* are given in *italic*.

Kind regards,

Konstantina Nakoudi, on behalf of all the co-authors

Comments by Anonymous Referee #2 and our replies and changes in manuscript:

We would like to thank the Anonymous Referee 2 for the constructive comments and recommendations. We made strong efforts to revise the manuscript. More emphasis has been given on the reasons behind the arisen discrepancies between the different methodologies. Comparisons were directed between ground-based and space-borne lidar measurements as well as numerical estimations from WRF. The comparison between lidar and radiosondes is not included in the new version of the manuscript, since the nature of the two methods is different. Furthermore, we decided to include one atmospheric model in the discussion and we selected the WRF, which has higher horizontal resolution than ECMWF. The comparison between WRF and radiosonde profiles and the application of the temperature, potential temperature and relative humidity criteria is excluded as well, since the comparison with lidar is indirect. As suggested by the two reviewers, Section 5 (Comparison to another location) was left out. A new Section (4.2) has been added, in which the sensitivity analysis regarding the WCT threshold is discussed.

1) **Lines 45-47, I suggest adding more details about the detection of PBL height using radiosonde measurements and other instruments (e.g. microwave radiometers) with the appropriate references.**

   More details have been added regarding PBL height detection from radiosonde measurements and other instruments such as microwave radiometer and Doppler wind lidar together with the respective references.

   The corresponding part in Section 1 (Introduction) has been modified as follows:

   Several methods have been proposed to estimate PBLH, utilizing vertically resolved thermodynamic variables, turbulence-related parameters and concentrations of tracers (Seibert et al., 2000; Emeis et al., 2004). More specifically, different methods for the determination of the PBLH from radiosonde measurements have been compared and the associated uncertainties have been estimated (Seidel et al., 2010; Wang and Wang, 2014). Radiosondes have been routinely used for decades and therefore are a valuable method for long-term climatology analyses (Seidel et al, 2010; Wang and Wang, 2014). Restrictions of radiosondes refer to the coarse vertical resolution of standard meteorological data with respect to boundary layer studies as well as the smoothing due to the sensor lag constant bounded by the high ascent rate of the radiosonde (Seibert et al., 2000). Remote sensing systems such as aerosol lidar, microwave radiometer (Cimini et al., 2013), wind-profiling radar (Cohn and Angevine, 2000) and Doppler wind lidar (de Arruda Moreira et al., 2018) are suitable for long-term measurements of various atmospheric quantities with high temporal resolution and they can be used either independently or synergistically so as to retrieve the PBLH. Space-borne lidar systems provide the advantage of spatial coverage, *although for studies focusing on a specific area of interest, measurements are constrained by the overpass frequency. Ceilometers entail less cost, but on the other hand, they include fewer channels and, thus, cannot be used for detailed aerosol studies. In elastic and Raman lidar systems atmospheric aerosols are used as tracers and the PBL top is indicated by a gradient in the range-corrected lidar signal (Menut et al., 1999; Brooks 2003; Amiridis et al., 2007; Morille et al., 2007; Baars et al., 2008; Engelmann et al., 2008; Groß et al., 2011; Tsaknakis et al., 2011; Haeffelin et al., 2012; Scarino et al., 2013; Summa et al., 2013; Korhonen et al., 2014; Lange et al., 2014; Bravo-Aranda et al., 2016).*

2) **Lines 47-54, not all the lidar systems can provide continuous, systematic and quantitative measurements of atmospheric aerosol profiles. Please clarify this aspect and expand this section, adding strengths and weaknesses between space-borne and ground-based systems and between lidar ceilometers and research lidars.**
We thank the reviewer for this remark. We included information about the advantages and disadvantages of satellite lidar systems and ceilometers:
Space-borne lidar systems provide the advantage of spatial coverage, although for studies focusing on a specific area of interest, measurements are constrained by the overpass frequency. Ceilometers entail less cost.
Furthermore, we replaced the sentence Lidar (Light Detection And Ranging) systems can provide continuous measurements … with the sentence:
Remote sensing systems such as aerosol lidar, microwave radiometer (Cimini et al., 2013), wind-profiling radar (Cohn and Angevine, 2000) and Doppler wind lidar (de Arruda Moreira et al., 2018) are suitable for long-term measurements of various atmospheric quantities with high temporal resolution and they can be used either independently or synergistically so as to retrieve the PBLH.

3) **Lines 93-104, to understand if the anomalies in temperature and precipitation are significant, you should also report in the discussion and in figure 1, the standard deviation of the measured and climatological values.**
The standard deviation of the climatological values is not available. The World Meteorological Organization provides only the climatological value for the mean daily maximum temperature and mean total rainfall, which can be accessed at http://worldweather.wmo.int/en/city.html?cityId=224. Standard deviations of the average maximum temperature and cumulative precipitation are discussed in the new Section 4.4.2 and their relation to the seasonal PBLH cycle is investigated.

4) **Lines 121-122, please specify the differences between nighttime and daytime configurations and the corresponding vertical sounding ranges.**
The configuration of the lidar system FMI-Polly^{XT} was the same during daytime and night-time. The lower limit of the vertical sounding range was depended on the height, where full overlap between the emitted laser beam and the receiver field of view was achieved. During the measurement campaign, the altitude of full overlap varied from 550 to 850 m. Vertical range covers the whole troposphere under cloudless conditions. This is sufficient for PBL studies considering the heights needed in the study. Engelmann et al. (2016) reports a maximum vertical range of 40 km, which depends on the capabilities (height bins) of the data acquisition.
These aspects are discussed in Section 3.1.1 of the manuscript:

The system vertical resolution was 30 m and the vertical range covered the whole troposphere under cloudless conditions. This is sufficient for PBL studies considering the heights needed in the study. Engelmann et al. (2016) reports a maximum vertical range of 40 km, which depends on the capabilities (height bins) of the data acquisition. The FMI-Polly^{XT} lidar system is described in more detail in Althausen et al. (2009) and Engelmann et al. (2016).

The incomplete overlap between the laser beam and the receiver field of view L-R (Laser-Receiver), restricted the observational detection range to heights above 200-300 m. This was partly counterbalanced by the overlap correction function. In this study, overlap corrections were performed at 532 nm following the methodology proposed by Wandinger and Ansmann (2002). During the measurement campaign, the L-R was completed at 550-850 m.

5) **Lines 125-127, please add more details about the overlap factor of the system. Is a correction applied? Which is the height of full overlap? Since incomplete overlapping could hamper the PBL height detection, this characteristic should be well specified.**

   In this study, a correction was used for incomplete overlap following the methodology proposed by Wandinger and Ansmann (2002). Full overlap between the emitted laser beam and the receiver field of view was achieved between 550 and 850 m.

   These aspects are discussed in Section 3.1.1 of the manuscript:*The incomplete overlap between the laser beam and the receiver field of view L-R (Laser-Receiver), restricted the observational detection range to heights above 200-300 m. This was partly counterbalanced by the overlap correction function. In this study, overlap corrections were performed at 532 nm following the methodology proposed by Wandinger and Ansmann (2002). During the measurement campaign, the L-R overlap was completed at 550-850 m.*

6) **Lines 143-145, and 151-154, see the comment for lines 125-127. The overlap characterization should be discussed in details in section 3.1.1.**

   As the reviewer suggests, we have included more information concerning the overlap characterization in Section 3.1.1. Please see our replies in Comments 4 and 5.

7) **Lines 148-150, the WCT method was also applied for the detection of cirrus clouds height base (Dionisi et al., 2013, ACP) where a sensitivity study was made to fix a proper threshold. Please add this reference.**

   We thank the reviewer for letting us know about the study of Dionisi et al. (2013). The reference to this study has now been added to the manuscript as follows:

   The WCT method has also been applied for the detection of cirrus cloud base height (Dionisi et al., 2013, Voudouri et al., 2018) over different geographical regions.

8) **Line 220, please specify the CALIOP version dataset used in this study.**

   In this study, CALIPSO Level 2 aerosol layer, Version V4-10, product was used.

   This information has been added in in the corresponding Section of the manuscript (now 3.2, previously 3.3):

   In this study, CALIOP Version V4-10 dataset was used. Currently, no operational CALIOP PBL product is available.

9) **Lines 225-227, please specify if the used CALIOP overpasses are nighttime or daytime measurements.**

   The cited study of Leventidou et al. (2013) used daytime lidar measurements. This is now specified in the corresponding Section of the manuscript (now 3.2, previously 3.3):

   Leventidou et al. (2013) evaluated the daytime PBLH derived by Level 2 Aerosol Layer products over Thessaloniki, Greece, for a 5-year period, making the assumption that the lowest aerosol layer top can be considered as the PBLH.

10) **Please specify the impact of the WCT threshold on the results. A sensitivity study could be of help to interpret the results. Which is the associated error to the estimated PBL height? Is there any correlation between the magnitude of this error and the agreement between the different datasets? The effect of the different model horizontal and vertical resolution should be also added in the discussion.**

    We would like to thank the reviewer for this recommendation. As suggested, we performed a sensitivity analysis regarding the WCT threshold impact on the PBLH values. For this reason, we used the case of 2 March 2009, which was already discussed in the applicability of the WCT Section (4.1). The threshold was modified from 0.03 to 0.08, which corresponds to 6-16% signal gradients. It was found that the overall performance of the WCT algorithm is stable, with the exception of elevated layers and strong gradients appearing inside the PBL. Subsequently, it has been observed that the agreement with the simulated PBLH from WRF was affected in the presence of elevated layers or internal aerosol content gradients. However, this deviation was dependent on the altitude of the aforementioned atmospheric features. Furthermore, during early morning, where the convective activity has not been initiated yet, a small fluctuation (30 m) was identified**.**

    A new Section (Section 4.2) has been added in manuscript, discussing the sensitivity analysis as follows:

    In cases of elevated layers or aerosol gradients within the PBL, it has been revealed that the signal decrease threshold needs to be properly adjusted (Section 4.1). In this study, we adapted the threshold (t) so that the WCT algorithm was allowed to identify signal gradients in the order of 6-16% (t=0.03-0.08, correspondingly). In this Section, we investigate the effect of the WCT threshold on the estimated PBLH. For this reason, we performed a sensitivity analysis modifying the signal decrease threshold for the case of 2 March 2009, where elevated layers were injected into the PBL.

    The overall performance of the WCT technique was stable (Figure 6), with the threshold affecting the results in only a few cases. When the lowest and more sensitive to detect weak layers threshold (0.03) was applied, a thin aerosol layer (around 1300 m) was identified (see Figure 4). At this time, increased thresholds (0.04-0.08) detected a stronger elevated layer (approximately at 2 km). The lowest threshold was also more efficient when gradients appeared inside the PBL (around 17:00 UTC), with the higher thresholds yielding increased PBLH by approximately 300 m. When the elevated layers were characterized by higher aerosol load (18:00-19:00 UTC), lower thresholds (0.03-0.05) performed better as well, with the higher ones identifying stronger layers (around 1 km). Thus, the PBLH deviation, introduced by the modification of the WCT threshold, appeared to depend on the altitude of internal gradients or elevated layers. However, in the early *morning (00:00-03:00 UTC), where the convective activity was not initiated yet, a minor fluctuation (30 m) was observed, related to the algorithm's sensitivity towards aerosol content gradients.*

An adequate threshold adaptation also affected the agreement with the modelled PBLH. More specifically, it is shown (Figure 6) that during cases where the applied threshold induced a deviation from the smooth PBLH evolution, the disagreement with modelled PBLH increased as well. Besides, the agreement with the simulated PBLH appears to depend on the altitude of the atmospheric features (internal or elevated aerosol gradients) that affect the performance of the WCT algorithm.

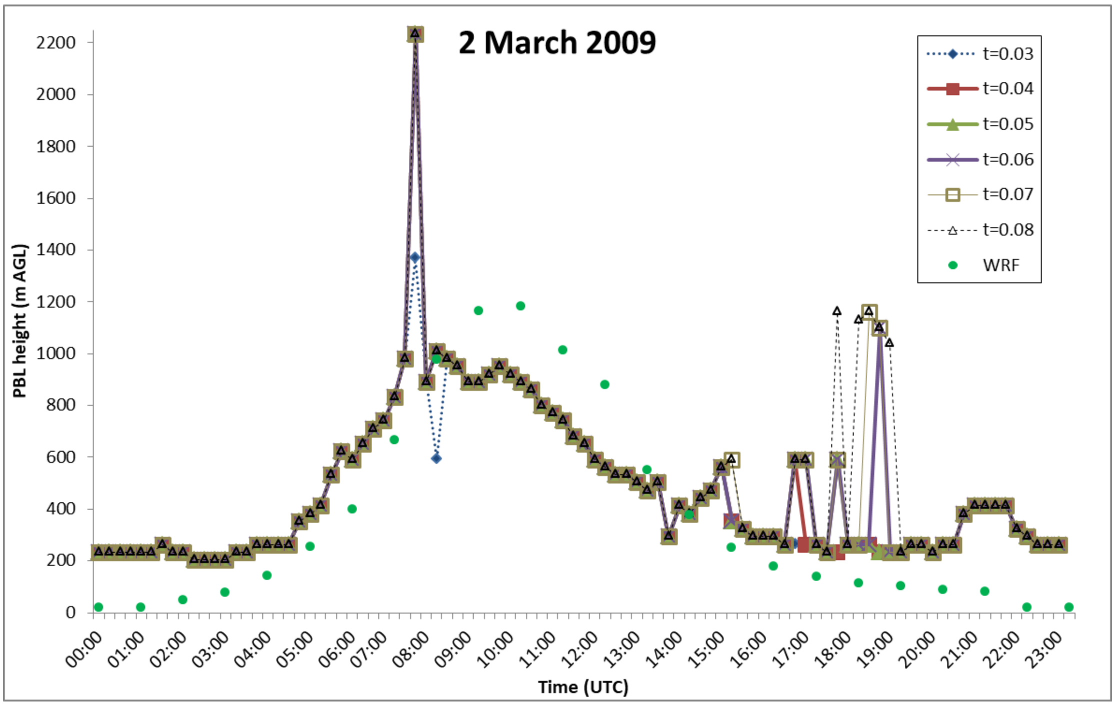

Figure 6: Sensitivity analysis of the WCT method for the case of 2 March 2009. PBLH was estimated by FMI-Polly^{XT}, after modification of the WCT threshold, and by WRF model.

11) **Lines 313-317. Please explain the difference on PBL heights retrieved by radiosonde and WRF model.**
The differences between the PBL heights derived by radiosonde and WRF model can be possibly attributed to the different vertical resolution of the radiosonde and WRF profiles. In the revised manuscript, we decided to exclude the WRF-radiosonde profile comparison because there was no direct comparison with lidar.

12) **Lines 384-386. The results of the PBL height comparison between lidar and rds during daytime seems due to the sum of two opposing effects: the overestimation and underestimation of PBL height by rds, respectively. In fact, few points are along the 1:1 black line of figure 6, with two clouds of points on the right and on the left of the 1:1 line. This is confirmed by a significant but not very satisfying correlation (R2 = 0.46). Please explain this effect and rephrase this section. Please add the statistical significance of the comparison.**

As the reviewer points out, in the scatter plot for comparing lidar to radiosonde PBLH there are data points on the left and right side of the 1:1 line. This effect can be explained by the different vertical resolution of lidar and radiosonde measurements. Furthermore, the different methodologies for the determination of the PBL height can account for discrepancies as well as the distance between the lidar station and the radiosonde launch site. The correlation coefficient for daytime measurements was found 0.68.
As the reviewer suggests, we performed a statistical significance test. More specifically, if p-value is lower than the significance level (0.05 in this case), then the corresponding correlation between the two datasets is considered significant. In other words, there is 5% probability that there is no relationship between the two datasets (null hypothesis).
The p-value for the correlation between lidar and daytime radiosondes was found equal to 0, while for lidar and night-time radiosondes was equal to 0.03. Since, the p-values are lower than the statistical significant level of 0.05, the correlation between the two datasets is considered significant.
In the revised version, we do not discuss the lidar-radiosonde comparison due to the different nature of the two methodologies. Therefore, previous Sections 3.2 and 4.2.1 as well as previous Figure 6 have been removed.

13) **Lines 405-409. Please add the statistical significance of the comparison.**
As the reviewer suggests, we performed a statistical significance test as in Section 4.2.1. We found that for ground-based and satellite lidar the p-value was equal to 0.005, which is lower than the statistical significant level of 0.05. Hence, the correlation between the two datasets is considered significant at the 95% confidence level. In other words, there is 5% probability that the null hypothesis (no correlation between the two datasets) is true.
In the revised manuscript we also comment that longer measurement periods or more extended comparison to ground stations are needed in order to perform robust comparison between ground-based and satellite lidar measurements.

14) **Lines 409-413. The number of considered cases is probably too small to generalize these results. Is there any noticeable difference between the different aerosol type of detected layers or between nighttime and daytime comparisons?**
Based on the analyzed CALIPSO observations, in the majority of cases the detected Aerosol Subtype was dust, with a few cases comprising dust-polluted dust and dust-polluted smoke mixtures. Furthermore, the layer top altitude did not appear to change systematically between daytime and night-time.
The relevant information has now been included in the manuscript:

Based on the analyzed cases, it was found that the overpass distance (here 20 and 101 km) from the lidar station and time difference between the measurements did not affect the agreement of the PBLH. Furthermore, the layer top altitude did not appear to change systematically between daytime and night-time. However, the small number of measurements does not allow us to generalize these findings. Hence, longer measurement periods or more extended comparison to ground stations are needed in order to draw more robust conclusions.

15) **Is it possible to add in this analysis the mean diurnal PBL evolution estimated through WRF simulations?**
The simulation of the diurnal PBLH evolution by WRF was dedicated to a specific number of cases, which are presented in Section 4.1 in order to justify the PBLH derived by the WCT method under different aerosol load and meteorological conditions.

16) **Is there an explanation for the ECMWF overestimation (Polly underestimation) of PBL top height during convective hours for Winter and Pre-monsoon seasons and the ECMWF underestimation (Polly overestimation) for Monsoon season? The good agreement found at 12 UTC should also be highlighted.**

As the reviewer points out, during convective hours in the winter and pre-monsoon, ECMWF overestimated PBLH, while in the monsoon season an underestimation was observed. During the monsoon period high amounts of precipitation are expected, whereas in winter and pre-monsoon much lower amounts are expected. This opposite behavior can possibly pertain to the modelled amount and initiation time of precipitation and subsequent evaporation. In addition, the soil and vegetation parameterization schemes significantly affect the energy and moisture fluxes inside PBL, which depend on the thermal properties of the underlying surface such as heat capacity and heat conductivity. In particular, the phase of soil water plays a key role in latent heat fluxes. It has been suggested that a non-proper representation of water soil phase can lead to a delay in soil cooling in the beginning of the cold period and a corresponding delay in soil warming in spring, an effect which is more intense if the solar forcing is significant as in the subtropical region of Gual Pahari. Both effects make soil temperature less responsive to the atmospheric forcing (ECMWF, 2010b, p.119), and, thus, can possibly account for the seasonal patterns appearing in the PBLH diurnal cycle.

Therefore, the partition between latent and sensible heat fluxes by the surface parameterization scheme of the ECMWF model could explain the reversed behavior during rainy and relatively drier periods.

During times of maximal insolation (6:00 and 9:00 UTC) the overestimation of PBLH by ECMWF was higher, especially in the winter and pre-monsoon seasons, where cloud cover is in general lower. However, in the presence of lower solar irradiance (12:00 UTC) and, thus, weaker thermal turbulence, FMI-Polly^{XT} and ECMWF exhibited the highest agreement, particularly in the winter and pre-monsoon periods, where the solar radiation is expected to be the main driver in the formulation of PBLH. Hence, the good agreement at 12 UTC is most likely related to the intensity of solar irradiance. On the other hand, during the monsoon period, the performance of ECMWF comparison is fairly the same during all convective times (6:00, 9:00 and 12:00 UTC). This can be attributed to the fact that more and more complex factors, such as cloud cover and precipitation, arise and contribute to PBLH development during the rainy monsoon season.

In the revised manuscript, we decided to exclude results from ECMWF due to its low horizontal resolution.

17) **The considered cases for this analysis are only 44 whereas for the previous section the number is higher (72). Please explain this difference. Statistical significance should also be specified. The measured differences in the growth rates between premonsoon**

and monsoon season can be attributed to a real signal or the poor significance of the sample does not allow any physical explanation? Please clarify these aspects.

Following the guidelines given by Baars et al. (2008), PBL growth period began when the PBL height started to increase (typically 2-4 h after sunrise) and was complete when 90% of the daily maximum height was reached (typically between 08:00 and 10:30 UTC). Regarding the daily evolution rate, this was determined through the slope of a linear fit to the hourly height values (between the start and the completion of the growth period). Furthermore, the calculation of the evolution rate was restricted to cases where at least 4 consecutive or 3 non- consecutive hourly values were available. Due to these criteria, the number of the mean growth rates data used in the analysis is lower than the number of the mean and maximum PBLH data.

The two-sided Wilcoxon rank sum test has been applied in order to examine whether the samples of pre-monsoon and monsoon growth rates are statistically different. The test has yielded that the two samples are statistically different at the 95% significance level (p-value=0.03). The differences in the growth rates between pre-monsoon and monsoon could be explained physically. More specifically, the slightly lower growth rates that were observed during monsoon season can be related to the weaker diurnal PBLH cycle that was found during this season. The above mentioned behaviour can be possibly explained by the differences in precipitation between the two seasons. The pre-monsoon season was characterized by less precipitation compared to the rainy period of monsoon. The relevant information has been added in the new manuscript as:

The distributions of daily growth rate during pre-monsoon and monsoon show similarities. In order to examine whether the distributions are statistically different we applied the two-sided Wilcoxon rank sum test (Wilcoxon, 1945; Wilcoxon and Wilcox 1964). The test yielded that the two distributions are statistically different at the 95% significance level. Hence, the differences in the growth rates between pre-monsoon and monsoon could be explained physically. More specifically, the slightly lower growth rates observed during monsoon are possibly related to the weaker diurnal PBLH cycle that was found during this season (Figure 8c). The above mentioned behavior can been explained by the different precipitation patterns during the two seasons, since pre-monsoon was characterized by less precipitation compared to the monsoon.

18) **I'm not sure that this section is bringing any relevant information. Please motivate this comparison with further details and results or remove the section.**

As suggested by the Reviewer, this section has been removed from the manuscript. The comparison with the PBLH characteristics over Elandsfontein site is performed in parallel with the corresponding results (PBLH diurnal and seasonal cycle) from Gual Pahari.