Response to Referee 3

We would like to thank the reviewer for his/her fruitful comments that helped to improve the manuscript.

The study presents an evaluation of CAMx model against LIRIC output profiles retrieved above the city of Thessaloniki. It is an interesting study with valuable results for the scientific community. However, the authors need to address some issues before publication. As it is currently presented, the idea of the validation is sometimes lost along the manuscript and the paper becomes a little too descriptive. The manuscript would benefit from a more in-depth discussion regarding the validation and more discussion including uncertainties is definitely needed. A review of the writing, which is sometimes confusing, and a possible shortening in length would also be useful to improve the manuscript. Find some more detailed comments below:

Page 1
Line 7: A fractional bias of 24.8% does not seem “close”. I suggest you use the absolute value here instead of percentage.

The text has been modified to: “mean bias of 0.57 km.”

Page 2
Line 3: Rephrase this sentence. As it is written, it looks like EMEP is a model instead of a programme.

The following text was removed: “European Monitoring and Evaluation Programme EMEP”

Lines 27-35: The identification of PM2.5 and PM10 particles with the fine mode and the coarse mode from LIRIC is not completely accurate. Please, rewrite.

See the relative comment response in Reviewer 1.

Page 2, Line 34: The text was modified to: “Instead of evaluating the performance of CAMx only for the PM10 particles, we separate the fine from the coarse particles by applying the LIRIC technique, then we convert the fine and coarse concentration profiles of LIRIC to PM2.5 and PM2.5-10 profiles and perform the validation for the PM2.5 and PM2.5-10 individually.”

Page 3

Lines 1-9: This information seems more appropriate for the methodology section than for the introduction.

In the introduction we briefly mention the tools used in our study which are described in more detail in the methodology section.

Line 25: “pre-processing”

The text has been modified according to the reviewer’s suggestion.

Line 27: Parenthesis are missing for the reference Schneider et al. (2000). Please, also add the more recent reference Pappalardo et al. (2014)

The text was changed to: “(EARLINET) (Schneider et al., 2000; Papalardo et al., 2014)”

Line 31: Was the sun photometer deployed at Thessaloniki just for this study?

The lines 31-32 have been rephrased to: “We used measurements from a CIMEL multiband sun-sky photometer which was installed in Thessaloniki in 2003 as part of the AERONET Global Network.”

Page 6

Line 21: Please, rewrite. It is not clear what you mean by “user defined uncertainties”. Does the study by Filioglou et al. (2016) take into the account the uncertainties in the input lidar and radiometer data or just the user defined input parameters? In that case, what is the estimated uncertainty of the output profiles? Include also here that LIRIC has been validated against in-situ aircraft measurements to emphasize that it can be used as an independent reliable tool for the validation of CAMx (see e.g. Granados-Munoz et al., 2016 and Kokkalis et al., 2017)

See the relative comment response to Reviewer 1.

The text was modified to: “The effects of multiple user defined uncertainties, such as the upper and lower limit heights of the profile and the algorithm’s regularization parameters, on the final result has been studied by Granados-Muñoz et al. (2014) and Filioglou et al. (2017) for selective case studies in Granada and Thessaloniki respectively. They agree that the parameter that produces the biggest uncertainties is

the lower limit height of the profile. Furthermore, the LIRIC retrievals have already been evaluated for volcanic and desert dust particles by Wagner et al. (2013) showing that the inversion can be accurate for two quite different types of aerosol. The aerosol extinction products of LIRIC has also been compared against the respective products from the Generalized Aerosol Retrieval from Radiometer and Lidar Combined data (GARRLiC) algorithm and against the retrievals from raman lidar measurements (Bovchaliuk et al., 2016). Finally, LIRIC has also been validated against in-situ aircraft measurements (e.g., Granados-Muñoz et al., 2016a; Kokkalis et al., 2017). Granados-Muñoz et al. (2016a) compared the LIRIC retrievals with airborn in-situ measurements and found a promising agreement with the differences between the two staying within the expected uncertainties. Kokkalis et al. (2017) analyzed a biomass burning case. Their comparison between the LIRIC retrievals and the aircraft measurements resulted in a good performance of the algorithm for the fine particles. As a result it can be used as an independent reliable tool for the validation of CAMx."

line 29: What do you mean by characterization procedure of the lidar profiles?

The text has been changed to: "aerosol type identification"

Page 7
Line 20: How did you calculate the full overlap height? Add references here and/or provide more details.

From the method of Wandinger et al. 2002 both the overlap function and the full overlap height are calculated. In this study we have applied the overlap correction per case using a typical overlap function. The overlap correction, however, cannot be extended to the ground. Typically, we limit the profile at the height where the function is higher than 0.7. For the CAMx validation however we preferred to use 0.9 (600m) to be on the safe side. This not clear in the text and it will be added.

Furthermore, in the original analysis we kept the lidar signals constant below 600m during the LIRIC inversion but the concentration product of LIRIC can be slightly variable even below the lower limit. To be entirely consistent with the idea of constant products below 600m we decided to keep the concentration profiles constant below this lower limit. This slightly affects the figures 3,4,5,6 and the tables 3,4,5,6. The text was also modified in order to clarify this adjustment.

Page 7, Line 20: The text has been modified to: "A lower height boundary has to be determined due to the overlap function of the lidar system. Operationally, we apply the

method of Wandinger et al. 2002 for the calculation of the overlap function and the full overlap height. In the current dataset the full overlap height was calculated at 900m. The correction however cannot be trusted down to the ground (Wandinger et al. 2002). In this study, we apply the correction down to 600m where the overlap function is still above 90% and use this height as the lower boundary of the LIRIC inversion. Below this height the lidar signals are considered constant during the LIRIC inversion. The concentration retrievals are also kept constant below 600m."

Line 26: Be more specific for the maximum height, what it is consider a significant quantity?

The text was rephrased to: "The upper boundary depends on the maximum height where aerosol exist in a significant quantity, that is, a region where the lidar signal from the aerosol backscattering can no longer be separated from the noise. This height can vary depending on the atmospheric conditions."

Line 27: Replace summing by adding

The text was modified according to the reviewer's suggestion.

Page 8
Line 5: Why are you using 1.5 and 2.6 g*cm-3? Why don't you use the known aerosol densities provided by CAMx for each case? That would lead to a more accurate comparison between LIRIC and CAMx.

The use of a CAMx derived density on the LIRIC profiles presupposes that the mixing ratio of each species is well predicted by CAMx. Otherwise the reference data of LIRIC would be affected by uncertainties originating from CAMx. Thus we preferred to use constant conversion values that are commonly used in the literature. This is also the way of Binietoglou et al. 2015. A more direct comparison would be to convert the CAMx profiles to ppbv. This conversion, however, was rejected after testing it because we wanted to avoid confusion by using a unit that is not adopted by the modeler's community like ppbv and the results were also pretty similar.

Lines 8-14: Since CAMx lacks of biomass burning aerosol emissions and does not consider desert dust emissions directly, I understand that the fires and dust categories are only used to evaluate the impact that this cases have on the model performance. However, for the evaluation purpose it would make more sense to me to include a

category excluding biomass burning and dust cases. That way you would be comparing apples to apples.

The main reason that we didn't isolate the cases that aren't biomass burning and dust is that our dataset is limited. By removing the 6 dust cases from the "non fires" group, the dataset is reduced to 11 measurements. Furthermore, by checking the dust cases individually we observed that unlike the coarse mode, the fine mode is generally in good agreement between LIRIC and CAMx. In order to provide more information on the dust cases the old Figure 4 is modified and the dust cases are displayed with orange color.

Line 21: Please, specify the criteria you use to detect dust cases. Some trajectories do not seem to originate in dust source regions in Figure 1. Idem for continental.

To characterize the cases we check the trajectories separately in the PBL and the FT. Then, one trajectory, either in the PBL or in the FT, is required in order to identify the measurement as dust or biomass burning. Additionally, an empirical criterion of a maximum dust concentration above 10ugr/m^3 in the DREAM profile is also applied to ensure that the trajectory carries dust.

Line 33: This sentence is confusing. Rewrite. What is the diameter for separation between fine and mode in CAMx? Is it the same as in LIRIC?

See the relative comment response to Reviewer 1.

The text has been modified according to the reviewer's suggestions: "Another hindrance in the analysis is that the fine and coarse mode of LIRIC are not directly comparable with the PM2.5 and PM2.5-10 modes of CAMx. The PM2.5 particles should include all the fine particles and a small part of the coarse particles that changes depending on the case. Additionally, the size distribution of the sunphotometer usually surpasses the PM10 diameter limit. Fortunately, it is possible to convert the fine and coarse modes of LIRIC to PM2.5 and PM2.5-10 particles. In the LIRIC inversion, the normalized volume size distribution of each mode is derived by separating the columnar size distribution of the sunphotometer in the two modes. The normalized distribution of each mode remains constant with height. Taking that into account, the fractions of the sunphotometer's coarse mode that belong in the PM2.5 region and the region outside the PM10 particles can be calculated from the sunphotometer's volume

size distribution. Then, the fine and coarse concentration profiles of each LIRIC case can be converted to PM2.5 and PM2.5-10 profiles using the equations 2 and 3. ”

Page 9

Line 1: Specify here the number of cases for the comparison. Why does this number emphasize the need of statistics?

The text was modified according to the reviewer’s suggestions: “A total of 22 cases take part in the comparison. We preferred a statistical approach in the analysis rather than comparing each case individually since the size of the dataset permits it.”

Lines 11-12: Provide more updated references.

The text was modified to: “(e.g., Flamant et al., 1997; Menut et al., 1999; Brooks, 2003; Tomasi and Perrone, 2006; Bravo-Aranda et al., 2016)

Line 15: Is this identification criteria based on a sensitivity analysis, previous studies, etc? Please, explain.

The reason why we use the upper limit criteria is that by applying the WTC method it is possible that a strong elevated layer could be identified as the PBL. This is also specified in Baars et al., 2008. In one of the cases they analyzed, an elevated dust layer complicated the derivation of the PBL top. Garrett, 1992 mention that the ABL typically extends from the ground to 2–3 km. Additionally, Georgoulias et al. 2009 in their study show that for noon measurements the mixing layer top is most of the time below 2600m for Thessaloniki. The selection of the upper limit value at 2500m is based on these studies.

The lower boundary corresponds to the height where the overlap function of the lidar system is above 0.9, that is 600m. The value of 400m is a typo and will be corrected. This is also mentioned in the response to Reviewer 1.

The text is modified according to the reviewer’s suggestion “Identification criteria are necessary for the selection of the PBL height. The top of the layer between 600m and 2.5km with the minimum value in the transformed signal is chosen as the boundary layer height. The upper limit is necessary in order to avoid identifying the top of sharp elevated layers as the PBL. According to Georgoulias et al. (2009) the upper limit of 2.5km is realistic for Thessaloniki. Baars et al. (2008) presented a case where an

elevated dust layer complicated the PBL height retrieval with the WCT method. The wavelet transform is applied to the LIRIC concentration profiles before the upscaling of the resolution."

Line 17: What it is the advantage of applying the WCT to LIRIC output profiles instead of the range-corrected signal as in previous studies?

The range-corrected signal is an optical product that is typically used for the boundary layer height calculation because it is representative of the aerosol quantity and it is also much more straightforward to calculate than i.e. the aerosol backscatter or the aerosol extinction coefficient. Here, the aerosol concentration is already available and it provides direct information of the quantity of the aerosols. Thus, we preferred to use the LIRIC products instead.

Do you obtain similar results using the volume concentration profiles and the RCS?

The application of the WCT either in the range-corrected signal or in the LIRIC concentration provides similar results.

Line 23: No aerosol is expected above the upper limit in LIRIC, why don't you set these values to zero instead of a constant value?

The vertical profiles of CAMx extend up to 9.5km. As it can be seen from figures 3a and 3b the model typically provides non zero values in the whole profile so it will not be realistic to assume that the concentration is zero above the upper limit. Consequently, we use the information of the last point of the profile as the best guess of the aerosol load above that height. In figures 3a and 3b it can be seen that this choice produced a mean concentration of 1-2ugr/m^-3 at 10km for both the fine and the coarse particles which is not abnormal for this atmospheric region.
However, it is true that in rare occasions when the SNR in the lidar signals is quite low, especially near the upper limit, the LIRIC inversion can be affected. Higher than expected concentration values can be produced near the upper limit resulting in unrealistic LIRIC overestimations for the integrated mass values in the FT. We detected that this is the case for two measurements in the current dataset, one that belongs to the "continental" category and one that belongs in the "fires" category. In order to ensure the quality of our reference data we decided to remove those two cases from

the analysis. This reduces the total number of cases to 22, the number of the "fires" cases to 5 and the number of the "non fires" cases to 16.

Page 10

Line 17: More discussion, including numerical values, is missing here. For the case on January 13, 2014, it looks like most of the aerosol concentration is below the full overlap height. How does this affect the output profiles? How reliable are LIRIC output profiles in this case? Please, add some discussion in this respect.

See the relative comment response in Reviewer 1

Lines 19-26: As it is presented, it is not very clear what the contribution of the analysis of the optical properties to the evaluation is. Considering that the goal of the paper is the evaluation of CAMx, I think this section should be shorter or rewritten to clarify its purpose. Additionally, previous studies have shown that backscatter provided by LIRIC is affected by large uncertainties, especially for non-spherical particles (see Wagner et al., 2013 or Granados-Munoz et al., 2014). How do these backscatter profiles compared to those retrieved with a different method (e.g. Klett-Fernald)?

The agreement between LIRIC and Klett derived optical properties is very good. We present in the paper two typical cases as a demonstration of the methodology used to examine the aerosol profiles for each individual case. The inclusion of lidar ratio and angstrom exponent profiles provides further evidence for identifying different aerosol types and layers, but such profiles can be misleading if someone does not examine in parallel the extinction and backscatter profiles. A relevant comment has been added in the text

Add text

Line 20: specify if it's extinction or backscatter related Angstrom exponent.

It's the extinction Angstrom exponent. The text was modified according to the reviewer's suggestions.

Line 20: It should be figure 2e instead of figure 2d. Include also the CAMx profile in Figure 2f

The figures have been renamed according to the reviewer's suggestion.

The CAMx profile in figure 2f is not provided since it is biased in a similar way with figure 2d. This is specified in the text. The space inside this small figure is also limited. Furthermore, the main point in this section is to shortly demonstrate the capabilities and all the possible products of LIRIC for two different aerosol cases.

Page 11

Lines 10-15: Add numerical values in the discussion. In general in this section 4.1, add more discussion taking into account the uncertainties and shortcomings in LIRIC (and the model if provided by the modellers).

Section 4.1 was modified according to the reviewer's suggestions

Page 12

Line 24: Can you provide some information about the boundary layer height values obtained in the study?

The boundary layer height retrievals of the cases vary between 600m and 2500m without showing any strong pattern. A slight preference for PBL values in the range 1000-1500m can be observed. However, one has to take into account that the cases are not uniformly distributed either in the annual and daily cycle. Both of these variables highly affect the PBL height.

Besides, because of the incomplete overlap, LIRIC uncertainty in the PBL should be higher than in the troposphere. Take it into account when discussing the results.

As it was mentioned in the previous comment responses we will include in the text that the lidar signals are overlap corrected down to 600m since it was not clearly specified. Consequently, the signals can be trusted down to 600m. Indeed the missing part of the signal (0-600m) that is assumed to be constant can produce uncertainties in the retrieval. Munoz et al. 2014 have studied the uncertainty of the LIRIC retrieval using different parts of the signal that were not overlap corrected, and thus always underestimated, within acceptable overlap values (above 0.8). They found that the produced uncertainty is higher in the near range in terms of absolute values. This approach, however, includes both the uncertainty of the part of signals that is not overlap corrected and the uncertainty of the assumption of constant signals below the lower limit. For that reason, it is uncertain if the height variability that they observe applies to our case.

Page 14
Line 7: “are presented”

The text was modified according to the reviewer’s suggestion.

Page 15
Line 2: Provide more details on the results obtained removing the dust cases

The following text was added: “The comparison for the PM2.5 particles is actually affected in a negative way due to the limited number of measurements in the dataset.”

The dust cases in all the scatterplots will also be marked with an orange color.

Line 7: Do you have information about the relative humidity above Thessaloniki during the study period? This could give an idea about how important the hygroscopic growth is and how much it could affect the comparison. Consider rewriting the conclusions section after all previous comments.

Unfortunately, the only information available for this period is the water content which is added in the PM2.5 calculation. This could be analyzed in a future study.

Table 2: Should be a instead of z (or vice-versa)?

See the relative comment response to Reviewer 1

Figure 3: Add also the number of cases for the no fires category in the figure

Figure 3 has been updated according to the reviewer’s suggestions.