Review of “Ceilometers as planetary boundary layer height detectors and a corrective tool for ECMWF and COSMO NWP models” by Uzan et al.

Uzan et al. combine observations by ceilometers at various sites and radiosoundings at one location to evaluate the planetary boundary layer height in two numerical models, namely the global IFS and the regional COSMO, in a geographically varying region (Israel). Only daytime planetary boundary layer (PBL) height in summer is considered. A good agreement between the PBL height retrieved from ceilometer measurements and soundings is shown, in line with previous studies, and indicates that the ceilometer is a suitable tool for evaluating model performance. The comparison of the PBL height between the models and observations show that the COSMO model generally performs better than the IFS. Two methods to estimate the PBL height from the models are used, neither one showing superior performance for both IFS and COSMO. The study implies that the COSMO PBL height has a bias depending on the distance from the shoreline and the topography, and a correction using these two parameters is presented. As the main finding, the manuscript claims to demonstrate that the PBL height retrieved by ceilometers can be used to improve the PBL height estimate from the COSMO model.

The region of study seems meteorologically interesting, and previous work using various observational techniques have demonstrated the influence of synoptic conditions, topography, and sea breeze on the PBL height in this region (Dayan et al, 1988, 2002; Uzan et al. 2016). Evaluating model performance regarding PBL height in this environment is a worthwhile effort. The methodology is given consideration: two techniques for calculating PBL height from the models and soundings are used, and most of the limitations of the measurements are fairly presented. The methods applied to estimate the PBL height (the bulk Richardson and the parcel method for model and radiosonde data, and the wavelet covariance transformation method for ceilometer) are standard and have been used in many previous studies. However, the manuscript falls seriously short in its analysis of the data, and the presentation and discussion of the results. The main conclusion, that the ceilometers can be used to improve the COSMO PBL height, is not sufficiently backed up by the results presented. A considerable drawback of this work is the small amount of data, which is surprising given that the main strengths of the ceilometer compared to other observational techniques available are the robust performance and low cost that allow continuous observations at multiple sites simultaneously. I strongly encourage the authors to obtain more measurements, if possible. Considering the lack of novelty in the methods and the small sample size, the authors should make considerably more effort in the careful evaluation and interpretation of their results.

Major comments

1. Motivation

The motivation, strengths and the central question of the paper could be made clearer in the introduction of the paper. As explained in Sect. 2, the region studied is interesting and different aspects are impacting the PBL height. The interesting aspects of the spatial variability of the studied region could be included in the introduction. In the light of the spatial variability, evaluating model performance on a single site would have limited value. One of the strengths of the study is the use of a network of ceilometers that can estimate the temporal development of the PBL at various locations simultaneously. This aspect deserves to be mentioned in the introduction.

Secondly, the introduction does not provide enough information to motivate the development of a post-processing tool for the modeled PBL height. The the goal to use ceilometer detected PBL height to correct for modeled PBL height could be simplified to “use A to correct B”. Currently, it is demonstrated that to “use A” is possible, e.g. PBL height can, with some limitations, be retrieved from the ceilometer measurements. However, “to correct B” is neglected in the introduction. The introduction only states the need for accurate PBL estimate, but no literature on identified shortcomings, methods found for improvement or anything else that would have been done previously to evaluate or improve PBL height estimates in NWP models is presented. Do previous studies suggest that it is more feasible to correct the end product (e.g. the PBL height) than to improve model parametrizations in order to obtain a better result from the model? Do the authors envision a use for the corrected PBL height? The authors could also consider whether their main aim should be on developing a correction, or rather a rigorous evaluation of model performance in the complex region. The latter could be helpful for understanding model shortcomings and would be a more general result than a location and time specific correction.

2. Amount and selection of data
One of the confusing aspects of this paper is the small number of days analyzed. The strength of the ceilometer is that data acquisition is cheap (see Sect. 1), however the small dataset is undermining this specific strength. The conclusions drawn are seriously undermined by the small sample size. For example, Sect. 6.2 seems to describe statistical results obtained from 13 data points. If possible, the authors should obtain more data. Alternatively, the study could be shifted to focus on case studies evaluating the shortcomings of the models in more detail.

Although the reasons for focusing on daytime PBL only in summer are given, further selection seems to have taken place. Why are only 13 days included from August 2015, and 20 days from August 2016 (L. 292-293) in Sect. 6.1? Why does Sect. 6.2 only include 5 ceilometer sites, when Sect. 6.3 includes 8 ceilometer sites (L. 319-321 and 345-346)? Why do Sections 6.2 and 6.3 only include data from August 2015, and not from August 2016? Are the 13 days used in Sect. 6.2 a subset of the 33 days in Sect 6.1? The authors should provide an explanation for the small number of days analyzed and why certain days and sites were selected at different stages of the study.

3. Significance
Related to the comment above about the amount of data, the authors should consider the statistical significance of the presented results. Specifically, wherever R-values are given (L. 298, Table 3, and elsewhere), the corresponding p-value should also be presented. Other techniques to analyze the statistical significance of the results are also welcomed, and the results should be discussed from the point of view of statistical significance.

4. Spatial variability (Sect. 6.2)
Section 6.2 could provide possibly the most interesting results for considering model performance in terms of PBL height in complex environments. If model under- or overestimation could be connected to certain processes (e.g. the sea breeze), the results would be more generally interesting. Mountainous coastlines are not unique to Israel, and many people inhabit such areas. This section deserves a proper evaluation, and the analysis and discussion should be extended.

Specifically, this section is hard to understand for someone not familiar with the geography of Israel. I would advice the authors to consider the presentation of their results. For example, the mean error at each site for each model and method could be presented with a symbol on a map having the color indicating the value. This would make any spatial structures in the mean, mean error (ME) or root mean square error (RMSE) more apparent. The authors could also plot the ME and/or RMSE as a function of the distance of the site to the shoreline and altitude above sea level (these are the two variables used for the correction in the next section).

From the authors description of the situation, it seems that the sea breeze has a clear influence on the PBL height. Is it to be understood, that the model does not correctly produce the sea breeze circulation, or is the model lacking in terms of the effect of the sea breeze on PBL height? It would be interesting if the authors could evaluate the discrepancy between ceilometer and model PBL height in terms of the strength, and spatial and temporal development of the sea breeze circulation during the day. Furthermore, in Sect. 6.3 data for 9-14 UTC are used, and I suggest the authors consider including the temporal development of the PBL height in their analysis in Sect. 6.2 as well.

5. The rationale of the correction presented in Sect. 6.3
Before a correction is developed and presented, it should be made clear that a corrections is needed and that there is a systematic bias that can be corrected for. Table 3 (and Section 6.1) show that the mean error of $COSMO_R$ compared to radiosondes is -3 m, which does not leave much room for improvement. Also Table 5 shows that at different sites the mean error of $COSMO_R$ is within a few tens of meters at most. (However, I would be cautious to draw conclusions from statistics comprising of 13 data points, and the authors should obtain a larger sample size if possible. See comments 2 and 3). For a 1 km deep PBL, an error of 30 m is 3%. For which application is this not good enough, and how good should the model performance be? Furthermore, considering that the definition of the planetary boundary layer is slightly ambiguous, can a perfect agreement between different methods be expected? The authors should explain why they think the model performance is not good enough and requires improvement. Furthermore, the authors could consider if the correction they presented would actually be more useful for the IFS model that shows clearly worse performance than the COSMO in terms of PBL height prediction.

Sect. 6.2 should demonstrate the basis of the correction presented in Sect. 6.3. The fact that the mean error in Tel Aviv, Beit Dagan and Weizmann are so similar suggests a spatial consistency that is more clear for $COSMO_R$ than $COSMO_P$. (Table 5). Is this the reason $COSMO_R$ was used for the correction in Sect. 6.3 instead of $COSMO_P$? The fact that there seems to be some spatial structure in the mean error is promising for developing a correction. The RMSE does not seem so spatially consistent.

To justify the correction method presented in Sect. 6.3, it should be established that a bias exist in the models' PBL height estimation that depends on altitude and distance from shoreline, that could consequently be corrected for. The authors should evaluate how the discrepancy between ceilometer and model PBL height depends on the topography and distance from shoreline. Furthermore, this could be done for different hours of the day, as the correction procedure is also applied for each hour separately.

6. Conclusion not supported by data
Perhaps the most serious shortcoming of the manuscript is that it is not demonstrated that the model result is better after correction. The authors should include a quantitative evaluation of the improvement of the model PBL after the correction. For example, the radiosondes at Beit Dagan could be used as an independent reference for the model PBL height. Another approach would be to estimate the correction parameters using only some of the available ceilometer stations, and using the remaining stations as a references to estimate the improvement in PBL height achieved by the correction. Varying the number of stations and the locations of the stations included for fitting the the correction parameters also gives an indicator for how many ceilometers needs to be included, or how they need to be located, for achieving a significant improvement for the $COSMO_R$ PBL height. If the authors aim is to show that the ceilometer is a useful tool to improve the modeled PBL height, the strength of their paper relies on the extent and rigor that this kind of analysis is carried out.

7. Presenting the research area
More attention should be paid to make the reasoning understandable for readers that are not so familiar with the specific geography and climatology of the region. Firstly, the studied region and its interesting aspects could be mentioned in the introduction. The first time the the location is given is the very end of the introduction, on line 97. This should be included already in the previous

paragraph that outlines the purpose of the study, as well as in the abstract. Secondly, a topography map should be included. Global topography data is available (for example from NOAA https://doi.org/10.7289/V5C8276M) and a map can be drawn using openly available tools (such as python). Depending on the weight the authors want to give to the humidity (mentioned on lines 103-104) and the prevailing synoptic conditions (line 125), they could also include a map of mean precipitation and pressure in August to help the reader to follow their argumentation.

Minor comments

8. Lines 1-2.
The authors should reconsider the title of the manuscript. The current title is somewhat misleading because it implies that the correction for PBL height was considered for both models, when in the manuscript only the COSMO PBL height was corrected. Furthermore, the journal guidelines recommend avoiding the use of abbreviations in the title, so the authors might want to avoid the use of "NWP" in the title.

9. L. 23-25.
Here results are given for flat and elevated terrain. Consulting Tables 4 and 5 it seems that flat terrain refers to Tel Aviv, and elevated terrain to Jerusalem. The authors should consider mentioning the sites for which the numbers refer to to avoid ambiguity, or at least mention that the values presented are from single stations.

10. Abstract.
The abstract does not mention Israel or give any other indication over the geographic locations apart from "heterogeneous area" and mention of the Beit Dagan radiosonde launch site. Location should be given.

11. L. 33-40.
Considering that this paragraph states the broad motivation and importance of this study, some references would be appropriate.

12. L. 56-57.
"ceilometers obtain a wide spatial resolution per lidar" - I'm afraid I do not understand the meaning of this phrase. Perhaps the authors mean that the a wider spatial resolution is achieved by ceilometers than lidars?

13. L. 53-65.
This paragraph seems to suggest that ceilometers are better than lidars in every aspect. It would be fair to mention a shortcoming of the ceilometer compared to a lidar.

14. L. 89-91.
It is not obvious here why the summer season is more appropriate for a approach that is limited by precipitation. It is later explained that this season has low precipitation. This should also be mentioned here to help the readers not familiar with local climatology.

15. L. 92-97.
It would be possible to help the reader further by outlining the structure of Sect. 6, either here or at the beginning of Sect. 6.

16. L. 85-86.
The introduction demonstrates the strengths of ceilometers compared to other available

observational techniques to estimate PBL height, but only states that ceilometers have not been used often for evaluating model performance. However, other observational techniques have, and this should be mentioned. Specifically, have other observational tools been used for evaluating PBL height in NWP models in Israel, or other mountainous coastlines?

17. Introduction.
I find the extent of presenting the literature for the use of ceilometer to detect PBL height satisfactory. However, no mention of previous work using ceilometer to derive PBL hight in Isreal is presented. The authors should site at least Uzan et al. (2016) and any other studies employing the measurement technique in their region of study.

18. L. 106.
"IMS weather reports" - The authors should provide a more specific reference, if possible.

19. L. 100-103.
Here could cite Fig. 1.

20. L. 111.
PBL height detection becomes increasingly difficult with increasing range (because of the decrease in the signal-to-noise ratio), and because of the low power of the ceilometer deep boundary layers are hard to detect. The moderate PBL height means that it is less of an issue in this study, and the authors could mention this to support their choice of instrumentation.

21. L. 112-115.
"Summer dust outbreaks in the eastern Mediterranean are quite rare (Alpert and Ziv 1989, Alpert et al., 2000) therefore, they were not addressed here, especially in the height levels below 1 km (Alpert et al., 2002)." - The sentence structure is unclear. Do the authors mean that especially dust outbreaks below 1 km were not addressed, or perhaps that the dust outbreaks below 1 km were especially rare and therefore not addressed? Should be clarified.

22. L. 119.
The abbreviation LST is not defined.

23. L. 116-138.
This is a paragraph about PBL structure and development in the studied region based on literature. It is useful and informative, even though it is concise and provides a lot of information for someone not familiar with the region. This paragraph is crucial for understanding the results, and the authors should not be afraid to extend if necessary to better understand the results. They should also refer back to this section at later parts of the manuscript when the concepts described are discussed. Furthermore, Fig. 3b could also be referred to as an example to aid the description of the diurnal cycle.

24. L. 116-138.
The use of abbreviations seems excessive: SBF and RL are only used once after being introduced, and could therefore omitted. Also CBL and SBL are only used 1-2 times after this paragraph and the need for the abbreviations is questionable and does not aid readability of the manuscript.

25. L. 136-138.
Please provide reference(s) for nocturnal PBL in Israel, if available.

26. Sect. 4.1
The placement of ceilometers in the heterogeneous research area should be described. Do the

ceilometer sites adequately represent the variability of the region? Are the different regions mentioned in the text (humid, arid, coastal, complex terrain) covered by the measurements?

27. Sect 5.3
The ceilometer backscatter profile is related to the aerosol loading, and therefore the layer that is detected is actually a aerosol layer. Implicit in the method described is the assumption that the PBL height corresponds to the height of the aerosol layer directly above ground. This assumption should be stated, and potential consequences to the results discussed. It is especially a limitation for detecting internal boundary layers which might develop due to the sea breeze circulation or katabatic winds.

28. L. 143 & Tables 1 and 2.
Table 2 is mentioned before Table 1 in text, the order of the tables should be swapped.

29. L. 156.
The authors could consider using the word “increased” rather than “improved” because it is more neutral. Although the model performance might have improved in important aspects due to increase in resolution, the computational cost likely did not.

30. L. 163-164.
“The spatial resolution of the models affects their ability to refer to the actual topography rather than a smoothed grid point.” Is this the reason that the ceilometer site is used as a parameter for the correction? If so, it should be clarified.

31. L. 164-165.
“the models' results were corrected by the actual ground base heights for each measurement site” - Unfortunately I cannot follow here. Presumably the correction meant here is not the correction presented in Sect. 6.3. Perhaps the authors mean that the model levels were adjusted based on the precise altitude of each ceilometer station? Clarification would be appreciated.

32. L. 144-162.
Considering that IFS provides boundary conditions for COSMO, and that the description of the COSMO model refers to IFS model parameterizations, the authors could consider switching the order of introducing the two models. e.g. move lines 156-165 before line 144.

33. L. 157.
It seems that the IFS has more vertical levels, but does it have better vertical resolution in the boundary layer? Information on vertical resolution should be added in Table 2.

34. L. 188-189.
“In order to derive the backscatter coefficient from ceilometer measurements, signal calibrations and water vapor corrections are necessary” - It is not clear if the corrections were done (presumably not), and should be clarified.

35. L. 193-194.
It could be mentioned that averaging multiple profiles improves the signal-to-noise ratio and thereby is likely to also improve the detection of the PBL height.

36. L. 197.
The overlap effect is a well known issue for lidar systems, however, the authors could provide a reference.

37. L. 215-217.
"the radiosonde's horizontal position is under 0.01° which is an order of magnitude from the models' grid resolution" - This is true for IFS but not for COSMO, which has a resolution of 0.025°. The authors should be more specific to avoid a misleading statement.

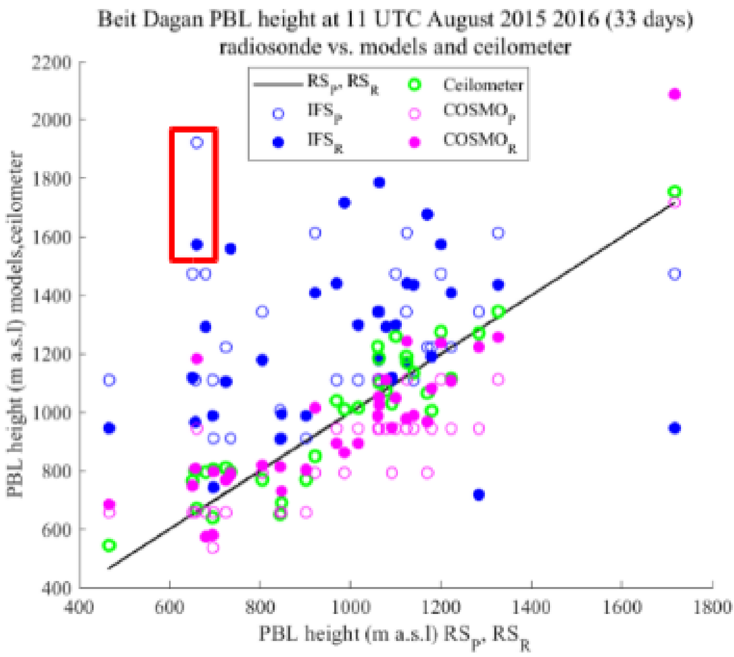
38. L. 239-241.
The method used for COSMO, why two different thresholds are needed, and how it differentiates from that used in for IFS or the radiosondes is not clear. What is the reason for applying a different criteria for COSMO than the IFS and soundings?

39. L. 282-283.
"This height indicates the entrainment zone rather than the actual cloud top." - For anything than the most optically thin clouds, the ceilometer signal attenuates before reaching the cloud top. Therefore, the ceilometer is very unlikely to be detecting cloud top.

40. L. 292-293.
Considering the change in IFS resolution between 2015 and 2016, is it appropriate to evaluate the IFS data together, or should data from 2015 be considered separately from 2016?

41. L. 302.
In the introduction it is mentioned that Ketterer et al. (2014) found poor correlation between ceilometer PBL height and the PBL height from COSMO. Why is their result so different from that found here?

42. L. 310 – 314.
As far as I can see in Fig. 2, the gap between IFS_P and RS is even larger for the data point indicated by the red rectangle in the figure below. I appreciate that the authors give an explanation to the anomalous PBL height on the 17 Aug 2016, but I'm concerned that this paragraph is slightly misleading. I'm not convinced that the difference between the IFS_P and RS is the largest on 17 Aug 2016.



I suggest the authors re-formulate this paragraph with the emphasis on giving an explanation for the anomalous PBL on 17 Aug 2016, rather than claiming this is the day with largest discrepancies, or alternative provide an objective measure for a “largest gap” and an explanation why the large discrepancy in IFS_R is worth considering but the even larger discrepancy in IFS_P on another day is omitted. Based on the next section, I could guess that these data points indicated by the red box are from 10 Aug 2015 (Fig 4b). If so, please include this information in this section of the manuscript.

43. Sect 6.1.
No discussion about the differences between bulk Richardson and parcel method is included. From Tables 4 and 5 it seems like IFS results are more sensitive to the choice of method. Perhaps the authors could discus these results.

44. Sect 6.1.
As far as I can understand, the main purpose of this chapter is to demonstrate the feasibility of ceilometer measurements to use for model evaluation. The authors could consider using this 33 point data set to compare the model results to the ceilometer to see if the results are similar than those obtained in comparison with the radiosondes to give additional confidence.

45. L. 324-330.
If the 13 days evaluated in Sect 6.2. are also included in the analysis of Sect 6.1, this paragraph does not provide any new information. For the clarity of the manuscript, I would advice the authors to include all comparison of radiosonde with other data in Sect 6.1, and focus on the spatial analysis in Sect. 6.2, as indicated by the title.

46. L. 331.
“By and large, $COSMO_R$ achieved the best statistical results” - This statement seems overemphasized. In terms of root mean square error, $COSMO_P$ performed better on 4 of the 5 sites presented, and the mean error was better for 2 sites.

47. L. 336-349.
“These results emphasize the advantage of high-resolution regional models such as COSMO (~2.5 km resolution) over the IFS global model (resolution of ~13 km in 2015 and ~10 km in 2016) over a diverse area.” Although not necessarily surprising, this is one of the few clear results of the paper, and deserves to be discussed and possibly further analyzed. Is the poor performance of the IFS related to lacking representation of the sea breeze circulation or some local scale phenomena?

48. Sections 6.1 and 6.2.
Did the authors consider the differences between the bulk Richardson and parcel method, and whether it indicates certain shortcomings in the models description of the boundary layer structure or processes? Comparing the $COSMO_R$ and $COSMO_P$ mean errors presented in Table 5, it seems that the two methods produce more similar results more inland (Ramat David and Jerusalem) than closer to the coast (Tel Aviv, Beit Dagan, Weizmann). This seems to also hold for the IFS. Is this related to the meteorological conditions, or simply a coincidence? Again, a significantly larger data set would be desirable.

49. Section 6.2.
Why are only 5 sites included, if ceilometers are available at 8? No station with the description “South” is included in the analysis of spatial variability (Table 1, L. 320), do the included 5 ceilometer sites adequately represent the spatial variability of the studied region?

50. L. 342-344.
“Following the conclusions of previous stages, $COSMO_R$ was chosen as the model and method that achieved the best results.” In my opinion, this was not well demonstrated (see also comment 46).

51. L. 344.
I’m guessing that the time window chosen is somehow related to the diurnal PBL height cycle that was nicely described in Sect. 2. Please provide explanation for the time chosen.

52. Fig. 4 and Sect. 6.2.
How are daily values obtained? Is the procedure the same as in Sect. 6.1, e.g. estimating the PBL height at approximately 11 UTC? If so, it should be mentioned in the text.

53. L. 349-357.
I’m not sure I understand the correction procedure. First, the variables α, β and γ are obtained by using the mean error (ME) between model and ceilometer at each station, and the altitude and distance from shoreline as predictor variables. After α, β and γ are obtained, it is possible to estimate ME anywhere in the domain. The corrected PBL height is then the COSMOR PBL height + the ME that is computed using altitude, distance from shoreline and α, β and γ. The same procedure is repeated for each hour, resulting in a time dependent α, β and γ. Is this a correct interpretation? The authors should clarify the description of their method.

54. L. 349-357.
Could the authors report the values of α, β and γ? The choice of repeating the correction for each hour of the day suggest some dependence of the correction needed on the diurnal cycle, does that exist? Do α, β and γ vary from hour to hour? What is the role of γ in the equation, and is it really needed? Presenting α and β would show whether altitude (e.g. topography) or distance from the shoreline (e.g. sea breeze circulation?) contributes more to the model discrepancy.

55. L. 358
Is the cross-section along a fixed longitude?

56. L. 369-370.
“The lowest value was corrected from 09 UTC (11 LST) to 14 UTC (16 LST)” - The way I understand this sentence is that the the lowest value was before the correction at 9 UTC, and after the correction it was at 14 UTC. This seems to contradict Fig. 5, which shows the opposite. Comparing Figures 5 a and b, it seems that the uncorrected data had the lowest PBL height at 14 UTC (independent of longitude). After the correction, at longitudes eastward of 35.1° (where Jerusalem lies) the lowest PBL height is found at 9 UTC. It would be advisable for the authors to clarify their statement.

57. L. 403.
“which improved the description of the diurnal PBL heights” - Unfortunately, there is no evidence presented that the model performance would have improved. See comment 6.

58. Conclusions.
The authors could discuss how the results obtained for daytime in a summer month might compare to other seasons.

59. Table 1.
Height limit is given as 7.7 or 15.4 km, but the footnote states that the data acquisition was limited to 4.5 km. It is not clear what is the vertical extent of the measurement. Although it is not that important for the study, the presentation is confusing and could be clarified.

60. Table 1.
The table includes specifications for the sites such as “north”, “south”, “inland”, “mountain”, but these do not seem to be defined or used elsewhere in the manuscript. Perhaps the regions could

provisionally be indicated on a map, and used in the discussion of the results.

61. Table 3.
For completeness, the table could include the mean and standard deviation also from the radiosonde used as a reference.

70. Table 4.
“The PBL heights were compared to the heights measured by the Beit Dagan ceilometer.” The text states (lines 321-322) “the models' results were compared to the ceilometers' measurements in each site”. These two statements seem to contradict each other, and I would ask the authors to correct one of them, or to clarify why different comparison measurements are considered in the text and in the table.

71. Tables 4 and 5.
It would be interesting to also see the mean PBL height of the ceilometer (the reference) at each site.

72. Figures 1 and 6.
Considering the political situation in some areas of Western Asia, the authors should carefully consult the journals guidelines regarding maps.

73. Fig. 3a.
The figure could contain the PBL height estimated by the two methods. It would be helpful to demonstrate the performance of the two methods.

74. Fig 3b.
It does not look like the data has been averaged for 30 min. Is the data presented at original 15 sec resolution? Please clarify in the caption.

75. Fig. 3b.
The authors should consider showing the time series of ceilometer and model based PBL height in this figure. It would be interesting to see 1) how the wavelet covariance transformation method is performing on the time series presented, 2) how the models predict the temporal development of the PBL height, and 3) whether the difference between model and ceilometer is random or the two models and two methods are consistently over or underestimating the PBL height during this one day. Although it might seem trivial to the authors, this helps the reader to gain confidence in the methods and helps with the understanding of the diurnal cycle of the PBL that is described in Sect. 2.

76. Fig. 3c.
The results presented here are not discussed. A description of the results presented here, and the ways they help to interpret Fig. 3 a and b or other results should be added. Furthermore, the wind direction figure could be improved by shifting the x-axis so that it is centered around North (e.g. scale from 180 to 360/0 to 180 degrees).

77. Fig. 4.
Figure 4 is hardly mentioned in the manuscript (it is referred to in the caption of Table 4, and Fig 4b is mentioned on line 326). Consequently, it is not clear what this figure is communicating. What is the additional information provided that is not already presented in Fig. 2? The better performance of COSMO compared to IFS, and the good agreement of ceilometer and radiosonde (Fig. 4b) are already demonstrated in Sect. 6.1.

78. Fig. 5
Figure 5 could indicate the locations of the Tel Aviv and Jerusalem ceilometer stations, as well as the mean (and standard deviation) of the PBL height estimated at these sites.

79. Figure 5 and 6.
I don't think it is necessary to list the sites and number of days used for the analysis in each figure caption. In my opinion simply a reference to the text for more details would do.

80. Fig 6.
Figure 6 could include the information of the mean PBL height at the stations.

81. Fig. 6b.
It is not clear what variable is presented in Fig 6b. Is it the ME estimated based on Equation 6, or one of the fitted parameters (α, β, γ)?

82. Citations.
The authors should check their citations and list of references list. For example, Uzan et al. (2012) and Uzan et al (2018) are cited but missing from the the reference list

83. Figures.
The authors should pay attention to the quality of figures. The font size could be increased in almost all figures (especially hard to read is Fig. 3), and use of color-blind friendly colors should be considered.