Review of "Multivariate statistical air mass discrimination for the high-alpine observatory at the Zugspitze mountain, Germany" by Sigmund et al. 2019 (acp-2019-211)

submitted to "Atmospheric Chemistry and Physics" "

May 21, 2019

In this study, the authors use a statistical approach to distinguish between different air masses at the high-alpine site Schneefernerhaus at Mt. Zugspitze in Germany. Based on different gas-phase and meteorological variables they performed a principal component analysis and define 9 air mass regimes. They validate the statistical approach with a mechanistic approach using the ceilometer-based mixing layer height and outline how this approach can be used for a real-time discrimination of air masses.

This study addresses a relevant topic in mountain meteorology and climate monitoring and falls into the scope of ACP. As far as I know, the method is novel and depicts a promising approach. It is clearly described and may thus be transferred to other high-alpine sites. Overall, the manuscript is well written and the outline is clear. Nevertheless, I have several detailed comments and suggestion which are given below and which should be considered by the authors before the manuscript is accepted for publication in ACP.

1 Specific comments

1. Introduction: The authors list studies done for many different high-altitude sites. However, one observatory which is not mentioned at all is Pic du Midi in the French Pyrenees. There was a recent study by Hulin et al. (2019) on atmospheric composition and the detection of thermally driven circulations with different methods, which should be cited as well.

2. p. 4, l. 22: "available at 1 min intervals". Are the values averaged over 1 min intervals or instantaneous?

3. p. 5, l. 2: What is the temporal resolution of the meteorological data?

4. p. 5, l. 3-5: How are the aerosol layer heights detected? With the manufacturer software or with a algorithm developed by the authors?

5. p. 5, l. 4: ">1": What is the unit? dB, B,...?

6. p. 5, l. 7: Is daylight saving time taken into account?

7. p. 5, l. 8: Why did the authors use a threshold of 66 % and not something else for the availability?

8. p. 6, l. 1ff: Why is there a time offset between the sites? Where the sites not synced to a time server? If the sites are not synced was there a shift of the time offset with time? E.g. what is the difference between Sept. 2013 and March 2014? How high is the correlation coefficient?

9. p. 6, l. 19: "exp" should not be italic.

10. p. 7, l. 16: Why only 89 % at the beginning? What happened to the remaining 11 %.

11. p. 8, l. 2: Why four classes? In the following lines only three air mass classes are defined?

12. p. 8, l. 24: What are "most suitable variables"? How are they defined? What are the criteria?

13. p. 9, l. 17: The figures should be referred to in the correct order, i.e. 8 not before 4.

14. p. 9, l. 19: What are the remaining measurements?

15. p. 9, l. 25: I was surprised to read about a marine boundary layer considering the location of Zugspitze. Where does the marine air mass come from? The authors speculate about that later in the manuscript, but I think a hint about its origin should already be given here.

16. p. 10, l. 8ff: What is the typical stability distribution? This could e.g. be checked with a histrogram.

17. p. 10, l. 12: GAP is a valley floor station. Why is this station used to detect strong synoptic forcing? Wouldn't it make more sense to use high-elevation sites?

18. p. 10, l. 16: Why does condition a) requires anabatic wind OR UFS below MLH_{GAP}? Why not AND?

19. p.10, l. 30-31: The authors state that PC1 was always a meaningful indicator while PC2 did not always allow for an unambiguous interpretation. On what is that assumption based?

20. Sect. 3.1: I found this part of the result section hard to follow and it might be difficult for readers not familiar with PCA to understand the interpretation of the results. It might be helpful to give a detailed example at the beginning on how to read and understand the loading diagrams in Fig. 6. What does a low score mean? Low absolute values or large negative values? In the text (p. 11, l. 1-6), the authors talk about scores while in Fig. 6 loadings are

shown. How does this relate?

21. p. 11, l. 32: "... air masses (Fig. 7a-g): compared ..." To make clear that the explanation why it was consistent is following.

22. p. 14, l. 15: This additional criteria of no clouds below 4 km, should be moved to Sect 2.5.

23. p. 14, l. 21: " .. this winds were not thermally induced BUT the MLH_{GAP} suggested non-ML air masses." BUT does not make sense. It should be AND.

24. p. 14, l. 28ff: What about LRMD and MBL/UFT air mass classes?

25. p. 15, l. 27: What differed between the six 2-month periods?

26. Fig. 7: Why not stick to the numbers I-IX for the regimes instead of introducing the long names for the regimes. This would make it more comparable with Figs. 3 and 8. The colours for the air mass classes should be brighter like in Fig. 8 and 9. Adjust the scales of the subplots to make them clearer to read (e.g. Figs. 7o, 7p, 7q, 7r).

27. Fig. 8: I probably understand it wrong, but how can regime VI belong to ML? In Fig. 3 there is no connector between ML and regime IV. Also, how can I and II belong to HYBRID? It would be good to enhance the boxes of the label to make the hatched areas better visible.

28. Fig. 9: Maybe add "UFS below MLH..." or "UFS above..." to the label and maybe refer to the text (p. 10, l. 14ff) where the 3 conditions are explained.

References

Hulin, M., Gheusi, F., Lothon, M., Pont, V., Lohou, F., Ramonet, M., Delmotte, M., Derrien, S., Athier, G., Meyerfeld, Y., Bezombes, Y., Augustin, P., and Ravetta, F.: Observations of Thermally Driven Circulations in the Pyrenees: Comparison of Detection Methods and Impact on Atmospheric Composition Measured at a Mountaintop, J. Appl. Meteor. Climatol., 58, 717–740, doi:10.1175/JAMC-D-17-0268.1, 2019.