Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-211-SC1, 2019

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Multivariate statistical air mass discrimination for the high-alpine observatory at the Zugspitze mountain, Germany" *by* Armin Sigmund et al.

Armin Sigmund

armin.sigmund@epfl.ch

Received and published: 19 May 2019

My responses are marked in *italic and blue* and were directly inserted below each comment of the referee.

I thank Referee #1 for the valuable comments, which help to improve the manuscript. In the following, I reply to the general comments, which offer the largest potential for discussion.

Printer-friendly version

Discussion paper

General comments from Referee #1

(1) Pre-processing of the data before use in the statistical classification method is limited to standardization (that is, adjustment of the sample mean to 0 and of the sample variance to 1). I am wondering whether any slightly more sophisticated pre-processing could be beneficial.
For instance:
(a) Some of the variables in the data matrix have well-defined seasonal and diurnal cycles. Would it be possible to determine average annual and daily cycles, and to remove them from the data set? Performing the analysis on deviations from the average cycles might improve classification results.

In our analysis, the seasonal course was roughly removed by subtracting the 2-month mean in each of the 2-month periods when standardizing the data. In principle, it would be beneficial to remove the seasonal course more accurately so that the leading principal components would only depend on the shorter-term variability of the time series. In the beginning of the data analysis, we considered the determination and removal of the seasonal course with a spectral approach such as a wavelet filter. However, this idea was abandoned because it requires knowledge of the time series within a window centered at the time of interest and is thus not applicable to real-time operational mode. It would be possible to remove an average seasonal cycle from each variable based on a larger, multi-year data set. Due to interannual variability, however, this approach would only partly remove the seasonal course. For example, Scheel et al. *(1999) showed that the monthly mean* O_3 *mixing ratios at Mt. Zugspitze can strongly differ between individual years and the 10-year ensemble. Therefore and because a large fraction of the seasonal variability is already removed by subtracting the 2-month mean values, I doubt that the removal of an average seasonal cycle would substantially improve the classification results.*

The mean diurnal cycles of the analyzed variables reflect shifts between air masses *that we aim to identify. Afternoon maxima of atmospheric constituents such as water vapor,* CO, ^{222}Rn, *and* NO_y *at high-alpine sites have been explained by thermally induced uplift processes including anabatic winds (*Forrer et al. *(2000);* Zellweger et al. *(2003);* Griffiths et al. *(2014)). These conditions are typically accompanied by afternoon minima of atmospheric stability and air pressure. The latter is associated with plain to mountain winds in the Northern Alps (*Lugauer and Winkler, 2005*). Removing the average diurnal cycle would complicate the classification of air masses in the case of thermally induced vertical transport.*

(b) PCA does not require the data to follow multivariate normal distributions, but its results can often be interpreted more easily if they do. It strikes me that most variables are concentrations, therefore their PDF will certainly be markedly non-Gaussian. Would a cleverly designed variable transformation allow bringing more variance into the leading principal components?

Among the PCA input variables, especially CO, CH_4, *and in the winter half year also* CO_2 *tend to have right-skewed PDFs. A logarithmic transformation, for example, would reduce the influence of the "long tails" of the PDFs on the principal component loadings. However, a variable transformation would not completely remove the "long tails" because the PDFs rarely follow an idealized distribution such as a log-normal distribution and they vary with the 2-month periods. Additionally, a variable transformation has the following side effect. If a logarithmic transformation is used for example and the original variable increases by a certain amount of units, then the transformed variable will not increase by a certain amount of units but by a certain factor. Although a variable transformation might somewhat increase the fraction of variance explained by the leading principal components in some 2-months, I have the impression that it would not significantly improve the classification of the entire data set.*

(2) The matching between air-mass regimes (I-IX) and air-mass classes (ML, UFT/SIN, HYBRID) is different in each two-month period (see Figure 8). The manuscript text

contains little or no information about the overarching logic. Why was this necessary? What criteria were used to attribute regimes to classes, how did these criteria change with the season?

In my opinion, the ad-hoc tuning of the method is a serious shortcoming. It is clearly a subjective component of the classification, and as such it cannot be exported to other sites. The authors do not explain this point in a satisfactory manner, and they probably should. Why wasn't it possible to design a fully objective classification rule? Formal methods to identify classification rules exist and could be used (see for instance chapter 14 in Wilks, 2011, Statistical Methods in the Atmospheric Sciences. DOI: 10.1016/B978-0-12-385022-5.00014-2).

I realize that the mapping of air mass regimes to air mass classes is only shortly described in the manuscript. More details will be added. The loadings of the leading principal components changed with the 2-month periods (Fig. 6), which can be explained by seasonal changes in chemical processes (e.g. CO_2 *emissions and uptake, photochemical* O_3 *production) and atmospheric dynamics (e.g. thermally induced uplift). Therefore, the characteristics of the air mass regimes (I-IX) changed with the 2-month periods and required a separate interpretation in each 2-month period. The air mass regimes were assigned to the air mass classes by visually comparing boxplots of the original input variables between the air mass regimes (Fig. 7). In the winter half year, the class ML was assigned if* CO, CH_4, CO_2, *q, and* $\Delta\theta_v$ *were relatively high and* p_{GAP} *and* O_3 *were relatively low compared to the other regimes. The class UFT/SIN was assigned in the opposite case and the remaining regimes were assigned to the class HYBRID. Apart from* CO_2 *and* O_3*, the same criteria were used in the summer half year;* CO_2 *was required to be relatively low for the class ML and* O_3 *was not considered.*

I agree that this subjective mapping of regimes to classes is a shortcoming. Nevertheless, it is based on typical qualitative differences between lifted and subsided air masses. The present study has the character of a pilot study to develop a novel mul-

tivariate approach for air mass classification. In future studies, a more objective and robust mapping of regimes to classes could be achieved by using a metric such as the median of a regime to define a threshold for "relatively high/low" characteristics and by using data from multiple years or more observatories.

References

Forrer, J., Rüttimann, R., Schneiter, D., Fischer, A., Buchmann, B. and Hofer, P.: Variability of trace gases at the high-Alpine site Jungfraujoch caused by meteorological transport processes, Journal of Geophysical Research: Atmospheres, 105, 12241–12251, 2000.

Griffiths, A. D., Conen, F., Weingartner, E., Zimmermann, L., Chambers, S. D., Williams, A. G. and Steinbacher, M.: Surface-to-mountaintop transport characterised by radon observations at the Jungfraujoch, Atmospheric Chemistry and Physics, 14, 12763–12779, 2014.

Lugauer, M. and Winkler, P.: Thermal circulation in South Bavaria climatology and synoptic aspects, Meteorologische Zeitschrift, 14, 15–30, 2005.

Scheel, H., Sladkovic, R. and Kanter, H.: Ozone variations at the Zugspitze (2962 m asl) during 1996-1997, WIT Transactions on Ecology and the Environment, 35, 264–268, 10.2495/EURO990511, 1999.

Zellweger, C., Forrer, J., Hofer, P., Nyeki, S., Schwarzenbach, B., Weingartner, E., Ammann, M. and Baltensperger, U.: Partitioning of reactive nitrogen (NO_y) and dependence on meteorological conditions in the lower free troposphere, Atmospheric Chemistry and Physics, 3, 779–796, 2003.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-211, 2019.