Atmos. Chem. Phys. Discuss., 6, S3770–S3773, 2006
www.atmos-chem-phys-discuss.net/6/S3770/2006/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Using discriminant analysis as a nucleation event classification method" *by* S. Mikkonen et al.

Anonymous Referee #1

Received and published: 13 October 2006

General comments:

In this paper, the authors apply a statistical analysis technique, known as 'Discriminant Analysis' (DA) to analyse a 3 years long dataset of nucleation events in the Po Valley (N.-Italy). The subject is on the border between atmospheric sciences and data mining techniques; to my feeling the emphasis in the paper is more on the statistical technique and it may therefore be less suited for the ACPD readership. The study however is an interesting exercise, even if a general application as parameterization technique in the end seems not immediately feasible (a negative result is a result anyhow). The paper is rather well written, although the section on the kernel estimation is probably too technical for non-specialists in DA. As a non-specialist I can only assume that the technique has been applied correctly technically speaking, although I have strong doubts that nu-

cleation events can be adequately described using this approach. Anyhow, I feel that the paper could be published (after having addressed the comments below) as it may stimulate an interesting discussion on statistical techniques and DA in particular as a tool for parameterization development in atmospheric chemistry models.

Specific comments:

The DA method is used to identify the set of meteorological and physico-chemical processes and parameters which determine whether or not a nucleation event would occur. The outcome of such an analysis could be one or more of the following:

- provide new insights on processes from an existing dataset

- simplify data analysis and data mining

- lead to a general parameterization, in order to simplify model calculations.

The authors should in their conclusions be more specific if (and how far) any of these objectives has been met by this study. In my opinion, they have not:

- The insights on indicator variables for nucleation eventrs were already obtained in the more straight-forward approach by Hamed et al. (analyzing the same dataset).

- Although the current analysis made it possible to 'fit' a model based on 3 parameters (relative humidity, O3 and radiation) to reproduce relatively well the event and non-event days in the observational set, clearly this parameterization is too coarse to be useful as a predictive tool under different circumstances. This is not surprising, as nucleation is a very critical phenomenon, depending in a non-linear way on source and sink rates of the nucleating species (apart from other environmental parameters like temperature and humidity).

My major critique to the study is that it is applying a technique which is basically not suitable for describing such a complex phenomenon (at least not based on the parameter set used). If DA is applied, it should start in the first place from a profound insight

ACPD
6, S3770–S3773, 2006

EGU

in the processes that affect nucleation, and I have the feeling that this insight is lacking. For instance, a parameter representing the formation rate or concentration level of the condensing species is missing. O3 may be a far proxy but probably better results would be obtained including SO2, which is not part of the data set, and therefore the analysis is flawed from the start. Although the authors have considered SO2 (but could not apply it as a variable due to lacking data) they should realize that probably they are missing one of the key parameters in the analysis which seriously reduces the applicability of the method. Also, using 24 hour averages of the variables seems to make little sense, as nucleation is mostly a day-time event. A seasonal separation of data may also have made a classification more straightforward.

If anything can be concluded from the study, it is that discriminant analysis (with the limited set of variables used in this study) is not a suitable technique for the prediction of nucleation events. At most it can be used to identify indicators (not necessarily causes) for nucleation events in an existing observational dataset, although it does not seem to yield more information than the analysis by Hamed et al. Hence, the conclusion of the authors that the model is easy to implement and apply in an atmospheric model is somewhat overkill.

Technical corrections:

It would be useful to describe in the introduction (or when presenting the scatter plots in Fig 1) the key parameters which affect nucleation, and from there justifying the selected variables. In classical nucleation theory, RH is a parameter which strongly promotes multi-component species nucleation. It is interesting to point out that atmospheric humidity also leads to 'swelling' of pre-existing particles, which act as a scavenging surface for the nucleating species (H2SO4, organics,. . .), hence RH also inhibits nucleation by increasing the condensational sink. The observed negative correlation between RH and nucleation indicates that the latter effect dominates the former.

P 8487 L 14: 'Materials' is probably not the appropriate term. Mabye just use 'Methods'

ACPD
6, S3770–S3773, 2006

EGU

P 8487 L 16: Pietro i.s.o Piero

P 8489 L7-8-9: Should be expressed more accurately. It is because you 'assume' that O3 and 'condensational sink' are high on pollution days that a positive correlation is obvious.

I presume the authors mean: ". . . ., which can be expected because high levels of particulate matter and O3 are commonly observed during pollution events".

Section 2.1.1 is rather technical and hard to understand for non-specialists (while it is probably trivial for specialists).

P 8494 L 12: "As the unclassifiedclass is not exactly independent. . . " please formulate more accurately what you mean by this.

Section 4, Discussion: I would actually like to see the 'function' or 'recipe' used to classify the data, i.e. a function in terms of RH, O3 and radiation.

P 8494 line 26: It is not proven that O3 has a positive effect on nucleation or helps particles to grow. The only thing one can say is that O3 is an indicator for nucleation events. Both O3 formation and nucleation may be caused by similar conditions (presence of radiation, presence of precursor trace gases,. . .) but this does not imply that there is a causal relationship between the two. (Also in conclusions, P 8496)

ACPD

6, S3770–S3773, 2006