Atmospheric
Chemistry
and Physics
Discussions

ACPD

Interactive
comment

Interactive comment on "Identification of new particle formation events with deep learning" *by* Jorma Joutsensaari et al.

Anonymous Referee #3

Received and published: 7 May 2018

General comments

The approach to use a Neural Networks (NN) method to automatize classification of new particle formation events is very useful in aerosol research at hundreds of previous, current or future atmospheric sites, where particle number size distributions are measured. It facilitates objective classification and faster classification, where you have long data sets. Hence, based on the research needs, this article deserves publication.

I therefore recommend the paper to be accepted subject to only minor revisions.

The minor concerns that must be addressed are:

You need to explain what happens when you have a different site: do you need to remake the training and testing with the new subset of images? Or can you use the

Printer-friendly version

Discussion paper

developed recognition in this study without any changes? And what happens when you have a site that has completely different shape of the size distribution compared to San Petro Capofiume? Or size range of the size distribution? In other words, what are your instructions and recommendations of how to proceed with your results and your methods when you do the same analysis at a different site? This needs to be clearly explained in abstract and conclusion sections. And there is no mentioning if you really recommend the method to be used already, or if you would like to develop it further before anyone should use it.

In relation to this: If you have to remake the training at each measurement site, do you always need to train the dataset with 50 % of the pictures? Does this mean that you have to select 50 % of your data at a new site already classified manually to be able to do the classification with NN? If this is the case, it is a severe drawback of the method. If you have to classify 50 % of the days manually on each new site, then then there is little point of doing the NN classification. If this is the case, then please write it clearly in the abstract and conclusions.

Even if these severe drawbacks occur, I still think the paper deserves publication, since a disadvantegeous result is important to convey to the science audience wanting to automatize classification of new particle formation events.

Some misspellings and grammatical errors. Please correct accordingly.

Thereafter, I go through some other minor issues below:

Abstract

The abstract is unusually long and has a very long description of the deep learning process. However, this is justified in this case, since aerosol researchers are normally not working with deep learning and a longer description is useful. So, the abstract should not be shortened.

Introduction and chapter 2.3

To be able to understand the NN methods, one way is to either, be very theoretically minded with an ability to understand abstract concepts and base your knowledge on this paper and other articles describing the methods, or you have to be practically oriented and learn by doing and be shown practical examples. As an extremely practically oriented person without an abstract mind set, I have no chance of understanding the methods based on reading. However, this does not automatically disqualify your text. After all, the abstract thinking might understand it. Hence, we have to accept that some people will understand the text, and some will not. Those that will not understand, will have to be learned by extensive simple examples, or by a teacher with a few practical examples, or at specialized workshops, and maybe with support from pedagogical video clips. Since it is not your task to develop extensive pedagogical descriptive examples (which is beyond the page limit of normal scientific papers), we have to accept this pedagogical problem and leave it as it is without further changes. Introduction

Page 2, row 12: Please add that passing on the method of classifying new particle formation events to a second person(s) might lead to systematic bias. If the second person passes on the knowledge to a third person(s), the systematic bias could increase further, and so on. I have experienced this problem previously, and it is a serious problem with the manual classification, and gives further motivation to develop automatic methods.

Page 2, rows 13-15: Wrong referencing to effects. That aerosols affect radiative balance does not automatically mean that they influence the climate via the direct and indirect effect. Please rephrase into something like this: "...radiative balance of the Earth and therefore the climate. They affect the climate directly by either scattering incoming solar radiation back to space or by absorbing it. Indirectly, aerosols affect the climate via their role in cloud formation as cloud condensation nuclei (CCN)."

Page 2, rows 18-19: Please add to the text that also the direct effect is leading to a cooling.

Page 2, rows 19: Please avoid using everyday language like "some". Please write "part" instead.

Page 4, rows 30-31: "Therefore, the idea of reusing what is already known instead of re-learning from scratch every time a new class has come up.". This sentence is not grammatically correct. Materials and Methods

Page 6, row 26: Please write that it is the "traditional method" in the title. Otherwise it can be confused with your new method.

Page 9, row 28 until Page 10, row 2. You mention that you "also tested three different sets of particle size distribution images". The reader might understand that these 3 methods are additional ones to the original method, while I think you mean that these are all the three methods that you have. Please rephrase to make it clearer, maybe by avoiding using "also" in the first sentence.

Page 9, rows 12-13 and Page 10, row 6. Please refer to Table 1 when mentioning the training/testing procedure.

Results and discussion

Page 12, rows 24-25. Do you have statistics to support your claim?

Tables

Table 1 caption text does not make sense when reading for the first time without studying the manuscript in detail. Please explain shortly in the caption text what you mean with training and testing. I am aware that this is explained later in the text (chapter 2.3), but needs a short explanation also when you mention Table 1 for the first time. Alternatively, you can write in the Table caption that this training and testing is explained in chapter 2.3.