A point-by-point response to the Referee #1 comments

Referee's comment 1:

First, the article claimed that in the abstract and throughout the paper (section 1: lines 29-30) that this is the first time that their method classified successfully NPF events. Of course this statement is not true. In fact, Junninen et al. (2007), Kulmala et al. (2012) and later Zaidan et al. (2017) also succeed classifying this automatically. Although they were using different methods and aiming slightly for different NPF classes, but the primary aim is the same: automatic NPF event classification. Zaidan et al. (2017) also obtained the accuracy about 84% in their recent study, which they may considered also a success. Therefore, the word "first time" here undermines the previous contributions. Instead, you should mention how different this method and/or how this method is better compared to the previous methods.

Authors' response:

We thank the referee to point out the valuable work of other researcher and the conference paper by Junninen et al., 2007, which we have not noticed before. In the revised version of MS, we will point out that this is the first time when a deep learning method, i.e. transfer learning of a deep neural network, has been successfully applied in NPF identification using the unprocessed data. In addition, we will compare in more detail our method with previous studies (Junninen et al., 2007; Kulmala et al., 2012; Zaidan et al., 2017). We would also note that the "protocol" for event analysis described by Kulmala et al. (2012) does not include the classification method by Junninen et al. (2007), even if Junninen is a co-author in the protocol paper.

Referee's comment 2:

In the lines 21-22 (section 1), the article said that "both of these studies were able to construct models to predict the probability of NPF occurrence with reasonable accuracy". This statement is not quite true. In fact, the paper by Hyvonen et al (2005) used data mining, such as ML classifier, to find relevant atmospheric variables to NPF. It is not to construct models to predict the probability of NPF. This statement should be revised.

Authors' response:

Hyvönen et al. states already in their Abstract "Using these two variables it was possible to derive a nucleation probability function." Therefore, even though the main purpose of the paper was to find variables related to NPF they still reported the probability function and thus the sentence in referred lines is correct. However, we will rephrase the sentence in the revised MS as follows: "Both of these studies were able to find the characteristic conditions for NPF event days in each site and it was seen that the conditions differ significantly. They were also able to construct models to predict the probability of NPF occurrence with reasonable accuracy, and this approach has also been used in the day-to-day planning of a complex airborne measurement campaign (Nieminen et al., 2015)"

Referee's comment 3:

The introduction section seems redundant and too long. There is a mixed reviews between the important of aerosol study, the use of data driven method and the algorithm review that is used in atmospheric study. This section should be narrowed down, by focusing only motivation of aerosol study and the use of data mining algorithms, such as CNN, in the field of atmospheric sciences and related discipline. General algorithm review can be pointed out to a specific reference sources.

Authors' response:

We agree that the introduction section is quite long. However, be believe that a general review of algorithms suitable for dataset classification would be interesting for audience of the journal. We will make the introduction section more compact and move details of CNN to an appendix in the revised version of MS.

Referee's comment 4:

In section 2 , lines 7-9, is it possible to get the latest statistics? The figure is 11 years old, this may be interesting to present the latest one, because the analyzed dataset was started from March 2017 (section 2: line 23).

Authors' response:

We have revised the statistic by the latest results by Nieminen et al. (2018).

Referee's comment 5:

Also, I may misunderstood your statement in section 2, lines: 23-25. Could you please clarify how the total days was 5534 from 24 March 2012 until 16 May 2017? (since in a year, there is only 365/366 days).

Authors' response:

The correct starting time is 24 March 2002, the error will be corrected in the revised version of MS.

Referee's comment 6:

For section 2.3, the first paragraph seems containing a lot of new details about the properties of CNN. This needs more clarification, for example, what are kernels, RELU and sofmax functions. Please also include other relevant mathematical details in the section 2.3? For example, in addition to Figure 2, it is good to include the mathematical representation of CNN.

Authors' response:

We will move detailed descriptions about CNN and training processes to the appendix in the revised MS including their short definitions. However, the detailed mathematical representation of a CNN will be quite long and include many variables and thus, we think, will most likely confuse the readers. The readers can read detailed description from textbooks of the subject, referred in the new version of the MS (e.g. Buduma and Locascio, 2017; Duda et al., 2012, Ch. 6.6.).

Referee's comment 7:

Any justification why you used standard procedure and options? (line 17, section 2.3)

Authors' response:

Our focus was to study suitability of a CNN based method for the event identification and we did not like to fine tune learning parameters. Therefore, we used procedures and options introduced in deep learning examples by Matlab because they worked well. In general, an optimization of the training parameters would be a very time consuming study. However, we will study an effect of training parameters in detail in the future.

Referee's comment 8:

Section 2.4, line 13, what happened to the first reference?

Authors' response:

We do not understand this comment. Both references (Venables and Ripley, 2002; Huber, 1981) are listed in the references. ANOVA is analysis of variance.

Referee's comment 9:

Another part of your data pre-processing is to uniform the pixel size of the images, could you explain how this has been done? Is it an automatic process?

Authors' response:

Resizing images to 227 x 227 pixels was done automatically by a standard imresize -function using default bicubic interpolation in the Matlab code (see https://se.mathworks.com/help/images/ref/imresize.html). This will be mentioned in the MS

Referee's comment 10:

What is the value to estimate class 0? because human might put some ambiguous days into this class., which they are not sure if the days are event and non-event. This class is very subjective, the CNN learning in this subjective class may confuse the model and bias the results.

Authors' response:

This is a very good question. In this study, we used the same classes than in the manual classification in order to compare human-made and CNN based results. In fact, we made some tests by ignoring Class BD but not Class 0. Ignoring Class 0 in the CNN analysis and classifying days that are not clearly event or non-event days (or bad data) into Class 0, it would be a very good idea to improve the model. We will test this in the future studies.

Referee's comment 11:

What do you think if you filter out the bad data before you feed this into CNN learning and analysis? In this case, the CNN learning can be simplified by reducing the number of classes.

Authors' response:

We made some tests by ignoring Class BD class but this did not significantly change the results (accuracy was practically the same). We feel that Class BD should be included to the CNN analysis because it is one way to recognize when the instrument was not working properly and also the accuracy to find those days is high (ca. 85 %). In fact, BD class helps us to ignore automatically days with a system malfunction during data analysis.
In general, reducing the number of classes is an interesting way of simplifying the learning process, but it means that we have less data in the training set, making the learning stage more prone to overfitting.

Referee's comment 12:

There is no method that is perfect. Describe the weakness of your method in the conclusion part?

Authors' response:

Of course, there are still some weaknesses in the developed CNN-based NPF identification method, e.g., quality and quantity of data is crucial in the training process, supervised learning is needed, the method needs some computing power (GPU), the identification is not perfect, particle formation and growth rates cannot be determined, etc. We will mention those weaknesses of the method in more detail in the revised MS.

References:

Buduma, N., and Locascio, N.: Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms, O'Reilly Media, Sebastopol, CA, USA, 298 pp., 2017.

Duda, R., Hart, P., and Stork, D.: Pattern classification, 2nd edition, John Wiley & Sons, Inc., 2012.

Hyvönen, S., Junninen, H., Laakso, L., Dal Maso, M., Grönholm, T., Bonn, B., Keronen, P., Aalto, P., Hiltunen, V., Pohja, T., Launiainen, S., Hari, P., Mannila, H., and Kulmala, M.: A look at aerosol formation using data mining techniques, Atmos. Chem. Phys., 5, 3345-3356, 10.5194/acp-5-3345-2005, 2005.

Junninen, Heikki, et al. "An Algorithm for Automatic Classification of Two dimensional Aerosol Data." Nucleation and Atmospheric Aerosols. Springer, Dordrecht, 2007. 957-961.

Kulmala, Markku, Tuukka Petäjä, Tuomo Nieminen, Mikko Sipilä, Hanna E Manninen, Katrianne Lehtipalo, Miikka Dal Maso, et al. 2012. "Measurement of the nucleation of atmospheric aerosol particles." Nature protocols 7 (9): 1651–1667.

Nieminen, T., Kerminen, V.-M., Petäjä, T., Aalto, P. P., Arshinov, M., Asmi, E., Baltensperger, U., Beddows, D. C. S., Beukes, J. P., Collins, D., Ding, A., Harrison, R. M., Henzing, B., Hooda, R., Hu, M., Hõrrak, U., Kivekäs, N., Komsaare, K., Krejci, R., Kristensson, A., Laakso, L., Laaksonen, A., Leaitch, W. R., Lihavainen, H., Mihalopoulos, N., Németh, Z., Nie, W., O'Dowd, C., Salma, I., Sellegri, K., Svenningsson, B., Swietlicki, E., Tunved, P., Ulevicius, V., Vakkari, V., Vana, M., Wiedensohler, A., Wu, Z., Virtanen, A., and Kulmala, M.: Global analysis of continental boundary layer new particle formation based on long-term measurements, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-304, in review, 2018.

Zaidan, M.A, V. Haapasilta, R. Relan, H. Junninen, P.P. Aalto, F.F. Canova, L. Laurson, and A.S. Foster. Neural network classifier on time series features for predicting atmospheric particle formation days. In The 20th International Conference on Nucleation and Atmospheric Aerosols, 2017

A point-by-point response to the Referee #3 comments

Referee's comment 1:

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 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.

Authors' response:

We have not yet tested the method with other sites. Basically, "banana type" events, non-event days and bad data should be recognized from other site data if pictures are plotted roughly in a similar way (one-day plot, size ranges, axes and color map). The method analyses features from size distribution plots, which are quite similar in many cases in different sites. However, we still think that the CNN should be transfer learned again for new sites in order to get best results, especially if the shapes of size distributions are completely different (e.g. low tide events in Mace Head in Ireland or rush hour episodes in urban environments). In other words, once learned CNN can be used in other sites but more precise results will be got if CNN has been transfer learned with data from the same site. In general, we think that anyone can already use the method because the basic concept is efficient enough. However, classification accuracy can be improved by testing different parameters and optimizing set of classes.

We will discuss this in more detail in the abstract and conclusions sections in the revised version of MS.

Referee's comment 2:

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.

Authors' response:

In general, the percentage of labeled data is not the most important parameter, but the number of images in relation to the size of the CNN is the relevant one. For instance, if the CNN is small – small number of layers and small number of neuron per layer – then, the minimum amount of images in the training pool is small, however, if the CNN is large, then the training pool should be large too. Furthermore, the complexity of the problem to be solved affects the number of images needed.

As a rule of thumb, if the data are similar enough from one site to another – particle concentration, time scale, measurement device, etc. – and the data are depicted with the same color map and log scale, then the classifier can be readily used. The color map plays a role only if it is not an "optimal" map; for instance, if the colormap distorts the data – e.g. a small variation in number concentration creates a big difference in color – then choosing another colormap will modify the performance of the classifier.

In our case, we have totally ca. 2000 days for training in the 50 % case and, e.g., Class 1 has only ca. 130 days. If you merge some classes together (e.g. Classes 1 and 2) and you have well-classified data, less training data is needed. Generally, you need some data for training. Once you have trained CNN, you can

used it for all new data from that site. Alternatively, simulated data could be used for training but we have not tested how well it works in practice.

We will discuss this in more detail in abstract and conclusions sections in the revised version of MS. In summary, we suggest that the CNN should likely be transfer-learned again for new sites to get the best results but in training, ca. 150 days per class should be enough to get a reasonable classification.

Referee's comment 3:

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.

Authors' response: We will only slightly modify abstract based on the Referee #1 comments.

Introduction and chapter 2.3

Referee's comment 4:

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.

Authors' response:

We have recognized this problem how to describe the method in a simple enough way but simultaneously theoretically enough. Based on the Referee #1 comments, we will move some of the most theoretical parts of the method description to an appendix to make the text more concise. Very practical descriptions and some examples (videos and codes) can be found, e.g. from Mathworks (Matlab) web pages: https://se.mathworks.com/discovery/deep-learning.html. Furthermore, more detailed description can be found from textbooks of the subject (e.g. Buduma and Locascio, 2017; Duda et al., 2012, Ch. 6.2.), which we now cite in the text.

Introduction

Referee's comment 5:

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.

Authors' response:

We will add to revised MS that passing on the manual classification method from researcher to researcher could lead an increasing systematic bias.

Referee's comment 6:

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)."

Authors' response: We will change text accordingly.

Referee's comment 7:

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

Authors' response: We will change text as suggested.

Referee's comment 8:

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

Authors' response: We will change text as suggested.

Referee's comment 9:

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.

Authors' response: This sentence have been deleted when we have reorganized MS.

Materials and Methods

Referee's comment 10:

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

Authors' response: We will change the title as suggested.

Referee's comment 11:

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.

Authors' response: We will change text as suggested.

Referee's comment 12:

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

Authors' response: We will change text accordingly.

Results and discussion

Referee's comment 13:

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

We have not studied this by statistical analysis and it is just a general statement based on randomly selected days. It would be very time-consuming if we check all misclassified days manually. In general, a part of misclassifications is pure booking errors made by researchers (wrong class written to database) and thus human-made errors seem to be more common. In addition, classification can be easily vary from researcher to researcher.

Tables

Referee's comment 13:

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.

Authors' response: We will change the caption text as suggested.

A list of all relevant changes made in the manuscript:

Abstract and conclusion:
- Pointed out that this is the first study when a deep learning method, i.e. transfer learning of a deep neural network, has been successfully applied in NPF identification using the unprocessed data
- More discussion about a suitable number of data needed for training and the use of the model in other sites

Introduction:
- We have made the introduction section more compact by moving details of CNN to an appendix
- Some smaller changes according to referees' comments.

Materials and methods
- We have moved some of the most theoretical parts of the method description to an appendix to make the text more concise.
- Some smaller changes according to referees' comments

Results and discussion
- Discussion about a previous study by Junninen et al., 2007
- Discussion about the use of the developed at other site data
- Some smaller changes according to referees' comments

Conclusions
- Summary of weaknesses of the method
- More discussion about a suitable number of data needed for training and the use of the model in other sites
- Some smaller changes according to referees' comments

Appendix A
- We have moved some of the most theoretical parts of the method description to an appendix:
    - General description of neural networks and training process
    - Convolutional neural network AlexNet and its transfer learning

References
- A few new reference
- Some smaller correction in references

Figures and tables
- Figure 2a modified (past and future text removed)
- Table 1 caption more informative

All relevant changes are marked in yellow in the MS text in next pages followed by a track change version of MS including all changes.

Identification of new particle formation events with deep learning

Jorma Joutsensaari[1], Matthew Ozon[1], Tuomo Nieminen[1], Santtu Mikkonen[1], Timo Lähivaara[1], Stefano Decesari[2], M. Cristina Facchini[2], Ari Laaksonen[1,3] and Kari E.J. Lehtinen[1,4]

[1]Department of Applied Physics, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland
[2]Institute of Atmospheric Sciences and Climate of the Italian National Research Council, Bologna, Italy
[3]Climate research Unit, Finnish Meteorological Institute, Helsinki, Finland
[4]Atmospheric Research Centre of Eastern Finland, Finnish Meteorological Institute, Kuopio, Finland

Correspondence to: Jorma Joutsensaari (jorma.joutsensaari@uef.fi)

Abstract. New particle formation (NPF) in the atmosphere is globally an important source of climate relevant aerosol particles. Occurrence of NPF events is typically analyzed manually by researchers from particle size distribution data day by day, which is time consuming and the classification of event types may be inconsistent. To get more reliable and consistent results, the NPF event analysis should be automatized. We have developed an automatic analysis method based on deep learning, a subarea of machine learning, for NPF event identification. To our knowledge, this is the first time when a deep learning method, i.e. transfer learning of a Convolutional Neural Networks (CNN), has successfully been used to classify automatically NPF events into different classes directly from particle size distribution images, similarly as the researchers do in the manual classification. The developed method is based on image analysis of particle size distributions using a pre-trained deep CNN, named AlexNet, which was transfer learned to recognize NPF event classes (six different types). In transfer learning, a partial set of particle size distribution images was used in the training stage of the CNN and the rest of images for testing the success of the training. The method was utilized for a 15-year long dataset measured at San Pietro Capofiume (SPC) in Italy. We studied the performance of the training with different training and testing image number ratios as well as with different regions of interest in the images. The results show that clear event (i.e., Classes 1 and 2) and non-event days can be identified with an accuracy of ca. 80 %, when the CNN classification is compared with that of an expert, which is a good first result for automatic NPF event analysis. In the event classification, the choice between different event classes is not an easy task even for trained researchers, thus overlapping or confusion between different classes occurs. Hence, we cross validated the learning results of CNN with the expert made classification. The results show that the overlapping occurs typically between the adjacent or similar type of classes, e.g., a manually classified Class 1 is categorized mainly into Classes 1 and 2 by CNN, indicating that the manual and CNN classifications are very consist for the most of the days. The classification would be more consistent, by both human and CNN, if only two different classes are used for event days instead of three classes. Thus, we recommend that in the future analysis, event days should be categorized into classes of "Quantifiable" (i.e. clear events, Classes 1 and 2) and "Non-Quantifiable" (i.e. weak events, Class 3). This would better describe the difference of those classes: both formation and growth rates can be determined for Quantifiable days but not both for Non-Quantifiable days. Furthermore, we investigated more deeply the days that are classified as clear events by experts and recognized as non-events by the CNN and vice versa. Clear

misclassifications seem to occur more commonly in manual analysis than in the CNN categorization, which is mostly due to the inconsistency in the human-made classification or errors in the booking of the event class. In general, the automatic CNN classifier has a better reliability and repeatability in NPF event classification than human-made classification and, thus, the transfer learned pre-trained CNNs are powerful tools to analyze long-term datasets. The developed NPF event classifier can be easily utilized to analyze any long-term datasets more accurately and consistently, which helps us to understand in detail aerosol-climate interactions and the long-term effects of climate change on NPF in the atmosphere. We encourage researchers to use the model in other sites. However, we suggest that the CNN should be transfer learned again for new site data with a minimum of ca. 150 figures per class to obtain good enough classification results, especially if the size distribution evolution differs from training data. In the future, we will utilize the method for data from other sites, develop it to analyze more parameters and evaluate how successfully CNN could be trained with synthetic NPF event data.

1 Introduction

Aerosol particles have various effects on air quality, human health and the global climate (Nel, 2005; WHO, 2013; IPCC, 2013). The air quality and health related problems are connected to each other, since in urban areas human exposure to elevated levels of particulate matter has been shown to cause respiratory problems and cardiovascular diseases (Brunekreef and Holgate, 2002), and eventually increase mortality (Samet et al., 2000). Very small particles like ultra-fine particles (smaller than 100 nm in diameter) can be particularly harmful because they can efficiently penetrate into the respiratory system and cause systemic effects (Nel, 2005). Air quality also affects visibility, for example, during smog episodes in large cities in Asia (Wang et al., 2013). On the global scale, aerosols affect the 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 radiation. Indirectly, aerosols affect the climate via their role in cloud formation as cloud condensation nuclei (CCN). The number concentration and chemical properties of CCN particles affect both the brightness of clouds (Twomey, 1974) and their lifetime (Albrecht, 1989). Increased number of CCN are associated to smaller cloud droplets, which can lead to brighter and longer-lived clouds (Andreae and Rosenfeld, 2008). It has been estimated that both the direct and indirect aerosol climate effects cause a net cooling of the climate, and can therefore cancel out part of the global warming caused by greenhouse gases (IPCC, 2013).

Atmospheric new particle formation (NPF; formation and growth of secondary aerosol particles) is observed frequently in different environments in the planetary boundary layer (e.g. Kulmala et al., 2004). There are direct observations that NPF can increase the concentration of CCN particles regionally (Kerminen et al., 2012). Based on global aerosol model studies, it is estimated that 30–50% of global tropospheric CCN concentrations might be formed by atmospheric NPF (Spracklen et al., 2008; Merikanto et al., 2009; Yu and Luo, 2009). The longest continuous observational datasets of atmospheric NPF have been collected in Finland at the SMEAR stations in Hyytiälä and Värriö starting in 1996 and 1997, respectively (Kyrö et al., 2014; Nieminen et al., 2014), and at the GAW station in Pallas from 2000 onwards (Asmi et al., 2011). These three stations

are located in the Northern European boreal forest, and can be considered to represent rural and remote environments. In more anthropogenically influenced environments long-term NPF observations have been performed in Central Europe at Melpitz (Wang et al., 2017a) and in San Pietro Capofiume at the Po Valley basin in Northern Italy (Laaksonen et al., 2005; Hamed et al., 2007; Mikkonen et al., 2011). In high-altitude sites, which are at least part-time in the free troposphere, there are less and

5   shorter continuous measurements of NPF available. However, also in these high-altitude sites NPF has been observed to occur regularly (Kivekäs et al., 2009; Schmeissner et al., 2011; Herrmann et al., 2015). Recently, long-term NPF measurements have also been established in several new locations, e.g., in Beijing and Nanjing in China (Wang et al., 2017b; Kulmala et al., 2016; Qi et al., 2015) and in Korea (Kim et al., 2014; Kim et al., 2013).

10   Currently, all the NPF studies published in the literature have utilized visual-based methods to identify NPF events from the measurement data. Typically, these methods require 1–3 researchers to analyze periods of formation and growth of new modes in the size-distribution data. These methods were first introduced for analyzing data from the Finnish SMEAR stations by Dal Maso et al. (2005), and have been later slightly modified to suit analyzing data from different environments and measurement instruments (Hamed et al., 2007; Hirsikko et al., 2007; Vana et al., 2008). While the visual-based methods are in principle

15   simple and straightforward to apply, there are certain drawbacks in using them. First, they are very labor intensive, since the analysis of the aerosol size-distribution data is not automated. Second, these methods are somewhat subjective, i.e. different researchers might interpret the same datasets in slightly different ways. Finally, passing on the manual classification method from a researcher to a researcher could lead to an increasing systematic bias.

20   There have been attempts at improving NPF event classification methods and making them more automatic. In their comprehensive protocol article, Kulmala et al. (2012) introduced a concept for automatic detection of NPF events. This was based on identifying regions of interest (ROI) from the time-series of measured aerosol size-distribution data. These ROI were defined as time periods when elevated concentrations of sub-20 nm particles were observed compared to the concentration of larger particles. Developing the NPF classification to take into account also other data measured at the same site, such as

25   meteorological data and concentrations of trace gases, has allowed utilization of statistical methods to search for variables which could best explain and predict the occurrence of NPF. Hyvönen et al. (2005) and Mikkonen et al. (2006) applied discriminant analysis for multi-year datasets of aerosol size-distributions and several gas and meteorological parameters measured at Hyytiälä, Finland and San Pietro Capofiume, Italy, respectively. Both of these studies were able to find the characteristic conditions for NPF event days in each site and it was seen that the conditions differ significantly. They were also

30   able to construct models to predict the probability of NPF occurrence with reasonable accuracy, and this approach has also been used in the day-to-day planning of a complex airborne measurement campaign (Nieminen et al., 2015). Junninen et al. (2007) introduced an automatic algorithm based on self-organizing maps (SOM) and a decision tree to classify aerosol size distributions. More recently, preliminary attempt at utilizing machine learning on big datasets have been reported (Zaidan et al., 2017).

There exists a long list of algorithms to automate the classification of different type of datasets (Duda et al., 2012), such as k-means, Support Vector Machine (SVM), Boltzmann Machine (BM), decision trees, etc. Our method choice is the deep feedforward Neural Network (NN). The idea of NN is not new, as the idea was first brought to life by Hebb in 1949 (Hebb,

5 2005). Since then, this field has drawn the attention of many researchers (Farley and Clark, 1954; Widrow and Hoff, 1960; Schmidhuber, 2015) probably because of its apparent simplicity and versatility. In this study, we used a convolutional NN (CNN) because it mimics the visual cognition process of humans. One of the main bottlenecks in the use of the NNs is the learning stage of NN; NN must be trained before it can be used in image recognition or other applications. For instance, thousands of images are typically needed for learning in image recognition applications. To overcome this problem, we used

10 a pre-trained deep CNN, named AlexNet, which was originally trained to recognize different common items (e.g., pencils, cars and different animals) and we transfer learned it to recognize images of different NPF event types. The transfer learning reduces significantly the number of images needed in the training process, from thousands to hundreds. The CNN used in this study and transfer learning of the CNN are described in detail in the Appendix A and Fig. 2.

15 In atmospheric science, several studies have used deep learning or other novel machine learning methods in data analysis. Deep learning and other machine learning algorithms are commonly used in remote sensing (Zhang et al., 2016; Lary et al., 2016; Han et al., 2017; Hu et al., 2015). Remote sensing is very suitable for machine learning because large datasets are available and the theoretical knowledge is incomplete (Lary et al., 2016). For instance, Han et al. (2017) introduced a modified pre-trained AlexNet CNN and Hu et al. (2015) used several CNNs (e.g. AlexNet and VGGnets) for remote sensing image

20 classification. Ma et al. (2015) used transfer learning in a Support Vector Machine (SVM) approach for classification of dust and clouds from satellite data. Other applications in atmospheric science include, e.g., air quality predictions (Li et al., 2016; Ong et al., 2016), characterization of aerosol particles with an ultra violet-laser induced fluorescence (UV-LIF) spectrometer (Ruske et al., 2017) and aerosol retrievals from ground-based observations (Di Noia et al., 2015; Di Noia et al., 2017). In addition, we recently applied different machine learning methods (e.g. neural network and SVM) for aerosol optical depth

25 (AOD) retrieval from sun photometer data (Huttunen et al., 2016). Although machine and deep learning approaches have already been used in several applications in atmospheric science, the use of those novel artificial intelligence methods will expand rapidly in the future. Nowadays, those methods are more efficient and easier to use due to the development of user-friendly applications and increasing computing capacity (e.g. Graphics Processing Units, GPUs), and they can be applied in problems that are more complicated.

30

Here, we demonstrate that a novel deep learning-based method, i.e. transfer learning of a commonly used pre-trained CNN model (AlexNet), can be efficiently and accurately used in classifying of new particle formation (NPF) events. The method is based on image recognition of daily-measured particle size distribution data. To our knowledge, this is the first time when a deep learning method, i.e. transfer learning of a deep CNN, has been successfully used in an automatic NPF event analysis for

4

a long-term dataset. We will show that the deep learning-based method will increase the quality and reproducibility of event analysis compared to manual, human-made visual classification.

2 Materials and methods

2.1 Measurement site, instrumentation and dataset

5   In this study, we analyzed a long-term particle size distribution (PSD) dataset measured at the San Pietro Capofiume (SPC) measurement station (44° 39' N, 11° 37' E, 11 m a.s.l.), Italy. The PSD measurements started on 24 March 2002 at SPC and have been uninterrupted, except for occasional system malfunctions. The SPC station is located in a rural area in Po Valley about 30 km northeast from the city of Bologna. The Po Valley area is the largest industrial, trading and agricultural area in Italy with a high population density and hence it is one of the most polluted area in Europe. On the average at SPC, NPF events

10   occur on 36 % of the days whilst 33 % are clearly non-event days and the probability for NPF events is highest in spring and summer seasons (Hamed et al., 2007; Nieminen et al., 2018).

At SPC, PSDs are measured with a twin Differential Mobility Particle Sizer (DMPS) system; the first DMPS measures PSDs between 3 and 20 nm and the second one between 15 and 600 nm. The first DMPS consists of a 10.9 cm long Hauke-type

15   differential mobility analyzer (DMA)(Winklmayr et al., 1991) and an ultrafine condensation particle counter (CPC, TSI model 3025) whereas the second DMPS consists of a 28 cm long Hauke-type DMA and a standard CPC (TSI model 3010). One measurement cycle lasts for 10 min. The PSDs used in this study were calculated from the measured data using a Tichonov regularization method with a smoothness constraint (Voutilainen et al., 2001). The data from two different DMPS instruments was combined in the data inversion. The CPC counting efficiency and diffusional particle losses in the tubing and DMA

20   systems were taken into account in the data analysis. In addition to PSD measurements, several gas and meteorological parameters are continuously measured at the SPC station (e.g. SO_2, NO, NO_2, NO_x, O_3, temperature, relative humidity, wind direction, wind speed, global radiation, precipitation, and atmospheric pressure). The measurement site and instrumentation have been described in detail in previous studies (Laaksonen et al., 2005; Hamed et al., 2007).

25   The analyzed dataset covers 4177 days (files) from the start of the measurement in SPC on 24 March 2002 until 16 May 2017 (totally 5534 days). The number of days at the different NPF classes and division into training and testing categories are summarized in Table 1.

2.2 Classification of new particle formation events (traditional method)

Currently, a classification of NPF events is practically made manually, i.e., researchers visually inspect contour plots of time

30   series of aerosol size distribution data and time evolution of nucleation-mode particles (particle dimeter below ca. 50 nm) (Kulmala et al., 2012). For the dataset of this study, the manual classification of NPF events into different categories is based

on guidelines described by Mäkelä et al. (2000) and Hamed et al. (2007). Figure 1 shows examples of measured time series of PSDs (time in x-axis, particle diameter in y-axis and particle concentration presented by different colors) for different event classes.

5   In the first step of event analysis, data is classified into days with NPF events and days without particle formation (non-event days, NE). A day is considered as a NPF event day if formation of new aerosol particles starts in the nucleation mode size range (<25 nm) and subsequently grows, and the formation and growth is observed for several hours. The NPF event days are further classified according to the clarity and intensity of the events (Hamed et al., 2007):

- **Class 1 events** (Class_01) are characterized by high concentrations of 3–6 nm particles with only small fluctuations
10   of the size distribution and no or little pre-existing particles in the smallest size ranges. Class 1 events show an intensive and clear formation of small particles with continuous growth of particles for seven to ten hours.

- **Class 2 events** (Class_02) show the same behavior as Class 1 but with less clarity. The formation of new particles and their subsequent growth to larger particle sizes can be clearly observed but, e.g., fluctuations in the size distribution are larger. Furthermore, the growth lasts for a shorter time than for Class 1 being about five hours on
15   average. In event Classes 1 and 2, it is easy to follow the trend of the nucleation mode and, hence, the formation and growth rates of the formed particles can be determined confidently.

- **Class 3 events** (Class_03) include days when same evidence of new particle formation can be observed but growth is not clearly observed. For example, the formation of new particles and their growth to larger particle sizes may occur for a short time but is then interrupted (e.g. by a drop in the intensity of solar radiation, rain). In addition, the
20   days with weak growth are classified in that category

The classification of nucleation events is, however, subjective and overlapping or confusion within the classes may easily occur. To minimize the uncertainty of the classification method, Class 1 and Class 2 events are typically referred to as clear or intensive nucleation events (Clear Event class), where all classification stages were clearly fulfilled, whilst Class 3 events are referred to as Weak Events. Baranizadeh et al. (2014) named Clear Event and Weak Event days as "Quantifiable" and "Non-
25 Quantifiable" days, respectively, which describes better the difference of those classes. Both formation and growth rates can be determined for Quantifiable days but not both for Non-Quantifiable days.

Other classes of days are:

- **Non-Event days (NE)** (Class_NE): Days with no NPF in the nucleation mode particle size range are classified into
30   Non-Event days. These days are also interesting because, e.g., differences in conditions (meteorology, gas concentration, precipitation) during event and non-event periods are often studied in order to get better understanding of processes behind NPF.

- **Class 0** (Class_00): Some of days do not fulfil the criteria to be classified either event or non-event day and they are classified as Class 0. In that class, it is difficult to determine whether a nucleation event has actually occurred or not.

6

- **Bad Data (BD)** (Class_BD): Days with some malfunction in the measurement system (e.g. too high or low particle concentrations, missing data in part of the day) are classified as Bad Data. In the final data analysis, those days are typically ignored.

2.3 Event classification with a deep convolutional neural network (CNN)

5   In this study, we developed a novel method to analyze NPF events automatically. The schematic of the approach is shown in Fig. 2 and described in detail in the Appendix A. We used a large, deep convolutional neural network (CNN) named AlexNet (Krizhevsky et al., 2012, 2017), which has originally been trained on millions of images, as a subset of the ImageNet database (Deng et al., 2009), to classify images into 1000 different categories. The AlexNet was a breakthrough method in image analysis when it was introduced in 2012 and thus it is widely used in image recognition applications (LeCun et al., 2015).

10   Since the model has originally been trained to recognize images of very common objects, e.g., keyboards, mice, pencils, cars and many animals, it has to be fine-tuned to recognize other images.

The AlexNet itself cannot recognize NPF events so we used a transfer-learning technique for fine-tuning the model for PSD images (daily contour plots of time series of aerosol size distributions). In the fine-tuning of CNN by transfer-learning, the

15   new features are quickly learned by modifying a few of the last layers using a much smaller number of images than in the training of the original CNN (Weiss et al., 2016; Shin et al., 2016; Yosinski et al., 2014; Pan and Yang, 2010). This strategy is very efficient for images because of the structure of the CNN; the first layers process the image by extracting some features so that it becomes more abstract and the lasts layers attribute the probabilities of the classes. In our setup, we modified only a few of the last layers because of the low number of available training data.

20

In a transfer learning process of our data set, we categorized PSD images into six different classes based on the manual event classification (Class 1, Class 2, Class 3, NE, Class 0 and BD, see Fig. 1) and used a subset of images for training and the rest of the images for evaluation of the training (testing). Three different fractions of images were used in the training stage (80 %, 50 % and 20%, see Table 1) in order to study the effect of image set size on the NPF event classification. Images for training

25   and testing were selected randomly (certain percent of each category) and this was repeated ten times to evaluate the statistical accuracy of the training and data classification. The transfer learning process and data analysis were computed using a Matlab program (version R2016b) with the support of the package of Neural Network Toolbox Model for AlexNet Network (version 17.2.0.0) using a Linux server (CPU: 2x Intel Xeon E5-2630 v3, 2.40 GHz,16 cores; RAM 264.0 GB; GPU: Nvidia Tesla K40c, CUDA ver. 3.5, 12 GB). We used the standard procedures (trainNetwork, classify) and options (solver: sgdm, initial

30   learn rate: 0.001, max Epochs: 20, mini-batch size: 64) as described in an example for deep learning by Matlab (MathWorks, 2017). In the transfer learning of Matlab AlexNet, we changed two layers of the net to fit our data: the number of recognized classes was reduced to six in the last fully connected layer (fc8) and the category's names in the classification output layer (output) were charged to the names of our event classes. One training session lasted about from a half to one hour with the

used server and programs. The AlexNet CNN has been originally implemented in CUDA, but it is also available in Matlab. Other programing languages provide also good libraries for implementing the AlexNet such as Python or C++, and dedicated frameworks such as TensoFlow, Theano or Caffe are good candidates for an easy implementation of the NN and the learning stage.

5

To find out the best performance for the NPF event classification, we tested three different sets of particle size distribution images (Image Set, jpg file format with bit depth of 24 bits) with different plotting areas of the daily-measured size distributions:

Image Set 1. Original image (see Fig. 1) without the title, axis labels and numbers (i.e. axes of the plot and data inside
10   them), size 1801 x 951 pixels (file size ca. 180 kB).

Image Set 2. Only colored, measured part of original images (i.e. time 00:00 – 24:00, particle diameter 3 – 630 nm), size 1646 x 751 pixels (file size ca. 120 kB).

Image Set 3. As previous but only an active time for NPF events is considered (i.e. 06:00 – 18:00, diameter 3 – 630 nm), size 831 x 746 pixels (file size ca. 60 kB).

15   Image sets with the different size were tested because all images have to resize to 227 x 227 pixels, which is an input image size of the AlexNet. Resizing of images pixels was conducted by a Matlab imresize -function using default bicubic interpolation (MathWorks, 2018). Resizing of images can lost some information of measured PSDs needed in the NPF identification.

After transfer learning each dataset (three image sets with three different training-testing fractions of images, repeated 10
20   times), the accuracy and training success were evaluated. We calculated the average success rate (accuracy) of the transfer learned CNNs by comparing CNN-based and human-made classifications of the test images. We also combined some classes together in the result analysis since overlapping of the classes could have easily occurred in the classification, e.g., Class 1 and 2 were combined to a Clear Event class (Cl_1-2).

2.4 Statistical methods

25   The performance of the classification between different image sets and training rates in different event classes was compared with the Kruskal-Wallis test and with multivariate analysis of variance (ANOVA). The ANOVA was conducted with a robust fit function (rlm; Venables and Ripley, 2002; Huber, 1981) because the normality and homoscedasticity assumptions of ordinary least squares method were not completely fulfilled. All statistical analyses were performed in R-software (R Core Team, 2017).

3 Results and discussion

The success of the transfer learning process of CNN was evaluated by confusion matrices, which showed a classification accuracy of the CNN method over visual inspection, i.e., how a certain class classified by human (visually) were categorized in the CNN classification. An example of the confusion matrices for one training-testing run (Image Set 1, training-testing ratio of 50 %/50 %) is presented in Fig. 3. When the training data set was analyzed with the trained CNN (same data), the overall classification accuracy (i.e. a fraction of days with an equal classification) was ca. 98 % and accuracies at certain event classes varied between 94-100 % (see Fig. 3a). For all cases (all image sets and training ratios), mean accuracies of ten different runs varied from 93 % to 98 %. The results show that the method can easily classified training data sets with very high accuracy, indicating that the training process of CNN was very successful.

Table 2 shows a summary of the classification accuracy for all cases when the method was applied for testing data sets (different days). The overall accuracy (all classes) is about 63 % for all studied cases. If we consider individual classes in more detail, the highest accuracies are in classes of NE and BD (ca. 80 %) whereas the lowest accuracies are in Classes 2, 3 and 0 (ca. 45 %), followed by Class 1 (ca. 53 %). The highest accuracies in NE and BD classes are apparently due to easier classification compared to other classes, e.g., no particles at low particles size ranges, no intensive particle growth or absence of particle at all and unusual high particles concentration in a part of data (see Fig. 1). The classification of other days is more challenging because differences between classes are not so distinct and the choice between classes can be difficult. When the clear event classes (Class 1 and 2) are combined into one category (Clear Event class, Cl_01-02), the classification accuracy increased to ca. 75 %. Overall, a classification accuracy for Clear Events and Non-Event categories is ca. 75-80 %, which can be considered a very good first result for automatic NPF event classification. In general, the classification would be more consistent, by both human and CNN, if only two different classes are used for event days: Clear and Weak events or "Quantifiable" and "Non-Quantifiable" events as described by Baranizadeh et al. (2014). From "Quantifiable" days, it is possible to quantify basic parameters of NPF event, e.g., particle formation and growth rates. Thus, we recommend that NPF event days should be categorized into classes of "Quantifiable" (Q) and "Non-Quantifiable" (NQ) events, in addition to Non-Event (NE), Undefined (UD or Class 0) and Bad-Data (BD) classes in the future analysis. This would also increase the number of images of the event classes in the training of CNNs.

If we look at the results in more detail, we can see variation in results between different training fractions (number of days in training in different event classes are shown in Table 1). In the cases of the training fractions of 80 % and 50 %, the classification accuracy values are quite similar but the accuracy decreases when the training fraction is only 20 % (especially for clear event days). Statistical analysis between different training-testing ratios shows that the 50 % and 80 % training rates are equally precise in all comparisons. This means that we could get adequate classification performance already with 50% training fraction (i.e., 135-708 images per class). However, when the size of the training set was lowered to 20 % (54-283

images per class), the classification became more uncertain in many comparisons. At 20 %, statistically significant differences were found when all classes were analyzed together and with classes Cl-01-02, Class 00, Class 02 and Class BD, indicating that the number of training days was too low for precise classification (e.g. Class 1 had only 54 days for training). In summary, a training fraction of 50 % (minimum 135 images per class) is a good compromise between accuracy, reliability and number of training days for the used data set.

We also studied the effect of image size (Image Sets 1-3) for classification accuracy. Only when all event classes were inspected together, there was a statistically significant difference between the image sets. Image Set 3 had a slightly lower performance rate than the two other sets. When limiting the classes to smaller sub groups the differences were not statistically significant anymore. The result indicates that image sets including all daily-measured size distribution data (Image Sets 1 and 2) are more suitable for CNN analysis than Image Set 3 with a reduced analysis period (6:00-18:00). Although images have to resize to the fixed input size (227 x 227 pixels) for CNN analysis, there is no need to reduce the analyzed period to cover only the active time for NPF. In fact, it is better to use all daily measured data in the analysis although resizing of images might lose a part of the information.

As described earlier, a choice between different event classes is somewhat arbitrary and not an easy task even for trained researchers and thus overlapping between different classes may occur. To analyze this overlapping in more detail, we plotted how a manually classified class is distributed into different classes by CNN classification. Figure 4 shows an example of CNN classification distributions for Image Set 1 with training-testing ratio of 80 %/20 %. Similar classification distribution into different classes can also be observed from the confusion matrix for Image Set 1 with training-testing ratio of 50 %/50 % (only one computing run) in Fig. 3b. The results show that an overlapping occurs typically between the adjacent or similar type of classes. For instance, a manually classified Class 1 is categorized mainly into Classes 1 and 2, Class 2 into Classes 1, 2 and 3, etc. The minimum overlapping is for classes NE and BD, which are the easiest classes to categorize by researchers. Similar overlapping is also observed in other analyzed cases (image sets and training-testing ratios).

To study overlapping or misclassification between different classes, we look in more detail at cases when Clear Event days (Classes 1 and 2) by human-made classification are categorized to Non-Event days by CNN-based classification and vice versa. Figure 5 shows examples of those days; the left-hand plots are categorized to Clear events by human and right-hand plots by CNN, the first row is examples of human made misclassifications, the second row CNN misclassifications and the third row difficult situations for classification. For instance, Fig. 5f shows a case when the initial stage of NPF has not been observed (probably due to change of wind or a mixing of boundary layer) but clear growth of particles is observed later. In that case, CNN-based classification does not "recognize" the missing of initial particle formation in the smallest particle sizes (ca. 3-4 nm) and therefore the day is classified as a Clear event day (Class 1 and 2). An opposite situation is shown in Fig. 5e where a Class 1 day (classified by human) was categorized two times into Non-Event and once into Class 2 day by CNN in

different computing runs. In that day, NPF is clear but the concentration of formed particles is lower than for typical event days and this probably affects the accuracy of the CNN-based classification. A similar misclassification can be seen in Fig 5c. In contrast, Figures 5a and 5b show examples of misclassifications, which are due to clear human errors, probably the researcher has just written down a wrong event class number. A general overview is that clear misclassifications seem to occur more commonly in human-made analysis than CNN-based categorization, which indicates that the developed CNN-based method has a better reliability and repeatability than manual human-made classification.

Only a few reports on automatic data analysis of NPF events have been published so far. In very recent conference proceedings, Zaidan et al. (2017) introduced a machine learning method based on a neural network. Their method utilized particle size distribution data pre-processing (fitting of lognormal distributions to the data), feature calculations and extractions (e.g. mean size, standard deviation, etc.) and principal component analysis (PCA) before the final classification by the neural network. Their preliminary results show a classification accuracy of ca. 83 % for Hyytiälä (Finland) data in 1996-2014 when only event and non-event days were considered. When compared to our method, they used several pre-processing steps before classification by the neural network and they did not use pre-trained CNNs or image recognition in their classifier. Kulmala et al. (2012) introduced a procedure for automatic detection of regional new particle formation. The procedure is based on monitoring the evolution of particle size distributions and it includes several steps, e.g., data noise cleaning and smoothing, excluding data larger than 20 nm and calculating regions of high particle concentration in particle size distributions. In the final step, the method automatically recognizes event regions from size distributions (i.e. regions of interest, ROI) and determines, e.g., NFP event start times as well as particle formation and growth rates. This method is very straightforward and does not include any data analysis methods based on artificial intelligence. To our knowledge, this method has not been used routinely in NPF event analysis for large datasets. Junninen et al. (2007) used an automatic algorithm based on self-organizing maps (SOM) and a decision tree to classify NPF events in Hyytiälä for 11-year data. They taught five SOMs by tuning them for specific event types and used a decision tree to get the probability for the day to belong to three different event classes (Event, Non-event and Undefined). The overall accuracy of the method was ca. 80 %. When comparing to our method, they pre-processed the data by decreasing the size resolution to 15 bins with variable width size bins in log-scale (higher resolution in smaller particles) and time resolution to 24 steps (1 h average). Furthermore, they weighted small particles (diameter < 20 nm) by a factor of 10 in one of the SOMs. To our knowledge, this method has not been used routinely in NPF event analysis. Hyvönen et al. (2005) used data mining techniques to analyze aerosol formation in Hyytiälä, Finland and Mikkonen et al. applied similar methods to datasets recorded in SPC (2006) and in Melpitz and Hohenpeissenberg, Germany (2011). They studied different variables and parameters that may be behind NPF but they did not make automatic classification for NPF events. Furthermore, Vuollekoski et al. (2012) introduced an idea on an eigenvector-based approach for automatic NPF event classification but they did not report any results because the method was still under development and its performance was uncertain.

We have not yet tested the method at other sites. Basically, "banana type" events, non-event days and bad data should be recognized from other site data if figures are plotted roughly in a similar way (one-day plot, same size ranges and axes, and color map). The method analyses features from size distribution plots, which are quite similar in many cases in different sites. However, we suggest that the CNN should be transfer learned again for new sites in order to get best results, especially if the

5 shapes of size distributions are different compared to those of training data, e.g., low tide events in coastal sites (O'Dowd et al., 2010; Vaattovaara et al., 2006) or rush hour episodes in urban environments (Jeong et al., 2004; Alam et al., 2003). We encourage researchers to use the method in other sites and report results in order to develop the method. The accuracy of classification could be improved, e.g., by tuning training parameters, optimizing a number of classes used in analysis (e.g. merging Classes 1 and 2) and using synthetic training data.

10 **4 Conclusions**

We have developed a novel method based on deep learning to analyze new particle formation (NPF) events. The method utilizes a commonly available pre-trained Convolutional Neural Network (AlexNet) that has been trained by transfer learning to classify particle size distribution images. To our knowledge, this is the first time when a deep leaning method, i.e. transfer learning of a deep Convolutional Neural Networks (CNN), has successfully been used to classify automatically NPF events

15 into different classes, including several event and non-event classes, directly from particle size distribution images, as the researchers do in a typical manual classification. Although there are general guidelines for human-made NPF event classification, the classification is always subjective and, therefore, it can vary between researchers or even within one researcher. In many ambiguous cases, it is not easy to attribute an event to the "correct" event class, even for an experienced researcher. The quality of the classification can especially vary for long-term data sets, which have been analyzed by several

20 researcher in different times. Furthermore, a wrong event class can be listed to database due to a human error, which reduces reliability of the classification. Therefore, an automatic method, which can manage the whole dataset at once with a high reproducibility, is desired for NPF event analysis.

Our results show that transfer learning of a pre-trained CNN to recognize images of particle size distributions is a very powerful

25 tool to analyze NPF events. The event classification can be done directly from existing data (figures) without any pre-processing of the data. Although an average classification accuracy of certain classes is ca. 65 %, the overall accuracy is ca. 80 % for Non-Event (NE) and Clear Event classes (Classes 1 and 2 combined), which is a good first result for automatic NPF event analysis. The most of miss-classified days have been categorized into the adjacent classes, which can be ambiguous to distinguish from each other. A compassion between CNN-based and human-made classification also showed that often the

30 difference in categorization is due to a wrong or an incorrectly listed classification by researcher. Human-made classification can easily vary by people to people and can change over time whereas CNN-based classification is consistent all times. The CNN based categorization seems, at least, to be as reliable as human-made and it could be even more reliable if training image

12

sets are selected carefully. Typically, an analysis of large size distribution dataset requires manual labor and training for several researchers, which is very time consuming and quality of analysis may vary. The developed automatic CNN-based NPF event analysis can be used to study long-term effects of climate change on NPF in more efficiently, accurately and consistently, which helps us to understand in detail aerosol-climate interactions.

5

The transfer learning of pre-trained CNNs (like Alexnet and GooLeNet) allows us to make automatic event classification systems effectively without long-lasting design, training and computing of CNNs. Typically, a training of a CNN needs from thousands up to a million images but in the transfer learning of a pre-trained CNN, a hundred images can be enough for a precise classification. The pre-trained CNNs, as well as other novel machine learning and artificial intelligence methods, and

10   the increased computing capacity due to Graphics Processing Units (GPU) enable us to analyze very complex and large datasets as is typically in atmospheric science.

Instead of the transfer learning of pre-trained CNN, several other artificial intelligence methods could also be utilized in NPF analysis. The Recurrent Neural Network (RNN) is a good candidate for the classification and predication of time series

15   (Pascanu et al., 2013), however, not viewed as surface plots but as a sequence of particle size distributions. This allows more variability in the NPF classes, e.g. a continuum of NPF event intensity, and determining time dependent processes like particle formation and growth. Furthermore, the underlying weights of NNs would give insights about the size distribution evolution and this could be potentially used as an evolution model by substituting or complementing the general dynamic equation (GDE). In addition, the use of unsupervised learning method can help us describing new NPF event classes or merge classes

20   based on humanly imperceptible features. Reinforcement learning (Kaelbling et al., 1996) is an interesting technique especially for the cases for which the classification is not well defined, e.g. if the classification varies from one human to another, because it only requires a rewarding system or policy. The rewards need to be positive if the prediction is satisfying and negative if not; hence, the reward values can be the difference between the number of person that classified in the predicted class and those who did not. However, it should be mentioned that this technique works best for prediction models, such as RNN.

25   Currently, the state-of-the-art in artificial intelligence are deep CNN based algorithms *AlphaGo Zero* and *AlphaZero* developed by DeepMind (Silver et al., 2017b; Silver et al., 2017a). Those algorithms achieved superhuman performance by *tabula rasa* reinforcement learning without human data or guidance and defeated human in the game of Go and the most dedicated program Stockfish in chess.

30   In summary, we encourage researchers to use the CNN-based NPF identification method to their own data because of the better reliability and repeatability compared to human-made classification. However, there are still some weaknesses that should be kept in mind, e.g., quality and quantity of data is crucial in the training process, supervised learning is needed, the method still needs quite a lot of computing power (GPU), the identification is not perfect and particle formation and growth rates cannot be determined using the current model. Furthermore, the CNN should likely be transfer learned again for new sites in order to

13

the get best results, especially if the size distribution evolution is very different compared to that of the site of the training data. We suggest that in training, ca. 150 days per class should be enough to get reasonable classification. Alternatively, simulated data could be used for training (Lähivaara et al., 2018) but we have not tested how well it works in practice. The method is, however, very easy to use and results are accurate and consistent, especially for long-term data, if the CNN has been trained

5   carefully with high quality and reasonable amount of data. Finally, experiences and data obtained from other sites can be used in the further development of the method, e.g. to find suitable learning parameters, more data for training and possibility for a cross-validation of the method.

In the near future, we will analyze long-term changes in NPF in San Pietro Capofiume and utilize the method for other field

10   stations (e.g. Puijo, Kuopio, Finland). We are including more parameters into automatized NPF analysis (e.g. particle formation and growth rates, event start and end times) and are developing methods to predict NPF events based on meteorological and other atmospheric data. In addition, the simulation based deep learning is a potential research topic in the future (Lähivaara et al., 2018).

5 Appendix A

15   *General description of neural networks and training process*

We used a deep feedforward Neural Network (NN) to analyze NPF events in this study. Figure 2 shows a schematic of the used method and detailed descriptions of the method can be found in the textbooks of the subject (e.g. Buduma and Locascio, 2017; Duda et al., 2012). In general, the NNs are built by stacking interconnected layers of atomic units, or neurons. In each

20   layer, each neuron computes one very simple operation – often named activation – which consists of computing the weighted sum of the input variables plus a threshold and then passing the results to a function – e.g. tanh -function or Rectified Linear Unit (ReLU, i.e. $\max(0, x)$ -function)(Buduma and Locascio, 2017). To put it simply, a neuron emits a signal if the excitation signal – the input variable – reaches a threshold otherwise it is inactive. A single neuron cannot do much on its own, merely a linear discrimination, however, a set of connected and trained neurons, as a NN, can mimic human cognition abilities – the

25   emergence property. The depth – or number of layers – and the topology of a NN will determine what it can be used for. For instance, the recurrent NNs (RNN) are well designed for speech, text and time series classification or prediction (Pascanu et al., 2013) and convolutional NNs (CNN) outperform any other architecture at classifying visual data, i.e. images (Krizhevsky et al., 2012).

30   Before it can start classifying, a NN – of any kind – must be trained for the specific task, e.g. to recognize a car or a keyboard in an image. While the topology is, for most cases, assigned a priori, the parameters – weights and thresholds – are to be learned. The learning stage is crucial because it determines the efficiency of the classifier and, therefore, the machine learning

14

community has dedicated much effort to provide solutions for the issue – e.g. the backpropagation algorithm (Rumelhart et al., 1986) – and is still focusing on it decades after it all started (LeCun et al., 2015; Schmidhuber, 2015). There are as many learning methods as there are NN topologies, however they can be sorted in a few categories. For instance, the learning method may require a labeled training set, falling into the supervised learning (SL) category (Duda et al., 2012; Amari, 1998), or it

5   may learn by itself without an already classified set, in which case the method is referred to as unsupervised learning (UL) (Le, 2013; Radford et al., 2016). Another relevant classification of the learning methods is the depth of the NN. Even though there is no common consensus to define the limit between shallow and deep NN (DNN), it is commonly accepted to refer to deep NN as NN having at least two layers and very deep those with more than ten layers (Schmidhuber, 2015).

10   The learning stage is one of the main bottlenecks in an artificial intelligence, especially for NN. For instance, thousands of images are typically needed for learning in image recognition applications. Therefore, instead of relearning the complete structure if a new class is introduced or merged with another one, a less expensive strategy has emerged, namely transfer learning; it consists in using part of the existing learned parameters and learning only a subset of the structure. The pre-training method consists in first training the network with an UL algorithm and then continue the training with SL (Bengio et al., 2006).

15   Another solution is reinforcement learning, which is a technique involving an agent that learns policies based on interaction with its environment using trial-and-error. For this method, the correct states are not known – therefore it is an UL – but a system of rewards gives hints whether the predictions are correct or not (Mnih et al., 2013; Sutton, 1988; Kaelbling et al., 1996; Schmidhuber, 2015). Finally, the transfer learning method, which we have used in this study, consists of learning the weights of an NN that contains a lot of classes (see "General" frame of Fig. 2b) and then using those weights either as a starting

20   point for other learning sets with less classes or using partially the weights – e.g. for the first layers of the NN – and train only one part in order to specialize the NN as it is depicted the frame "Specialization" of Fig. 2b. (Yosinski et al., 2014; Cireşan et al., 2012; Pan and Yang, 2010; Mesnil et al., 2012; Krizhevsky et al., 2012; Weiss et al., 2016). For all kinds of NNs, the learning process is prone to learning too much detail – missing the generality of a semantic class – related to the examples. This phenomenon is known as overfitting and several methods have been developed to overcome this such as dropout (i.e.

25   ignoring randomly some neurons by setting them to zero during training)(Srivastava et al., 2014; Hinton et al., 2012).

Convolutional neural network AlexNet and its transfer learning

In this study, we used a CNN because it mimics the visual cognition process of the human. Figure 2c illustrates the structure

30   of the whole CNN and Fig. 2d and 2e some layers of the CNN in detail. Instead of being fully connected, the CNN is only locally connected, in other words, a neuron of a layer is connected to a compact subset of neurons of the previous layer. For every subset, the neurons compute the same operation – the parameters are shared across the neurons of a layer – resulting in a convoluted version of the input signal; hence the name. It is a good model for human vision because it applies the convolution operation throughout the image field – thus it has drawn a lot of attention in the image processing community (e.g. LeCun et

al., 1998; Cireşan et al., 2011; Chellapilla et al., 2006). CNN became popular when the large, deep CNN named AlexNet outperformed by far all the other pattern recognition algorithms during the ImageNet Large Scale Visual Recognition Challenge, ILSVRC 2012 (Krizhevsky et al., 2012). Since then, CNN has been improved for each classification challenge, starting by the ZFnet (Zeiler and Fergus, 2014) then by the deeper GoogleNet (Szegedy et al., 2015) and finally by the deepest
5   Microsoft ResNet (He et al., 2016).

AlexNet was used to identify NPF events in this study. The architecture of the AlexNet, shown in Fig 2c, consists of five convolutional layers (CL, Fig. 2d), some of which are followed by max-pooling (MP) layers (a down sampling layer that extracts the maximum values of predefined subregion) and three fully connected layers (DNN, Fig. 2e) with a final 1000-way
10   softmax (an output layer), which assigns a probability ($P(\ell)$) to each of the classes' label (ℓ) for the input image (using a normalized exponential function, softmax). The first five layers of the AlexNet extract abstract features that are easier to classify than the original input image. This abstraction is computed by chaining up CL and MP layers. The CL (depicted in Fig. 2d) computes several convolution of the image by several kernels (i.e. matrices used in the convolution) – that must be learnt – and generates a multidimensional output – one dimension per kernel. For instance, in Fig. 2d the two kernels v_1 and
15   v_2 are applied to two compact sub regions of the data (*field 1* and *field 2*) by two units (*unit 1* and *unit 2*), which generate two 2D outputs ($y_1 = (x_1^1, x_1^2)$ and $y_2 = (x_2^1, x_2^2)$). The CL is what makes the CNN similar to the human visualization system because it applies the same transformation throughout the field of view – making the classification shift invariant. The MP layer merely computes a smaller image from each output image of the CL, e.g., an image of 256 x 256 pixels generated by CL will be reduced to smaller image of 64 x 64 by calculating maximum values of the 4 x 4 patches of the original image. The
20   last layers of the AlexNet is a DNN as depicted in Fig. 2e , i.e. a NN whose input and output layers are connected by hidden layers of neurons. Contrary to the CL, each neuron of a DNN's layer is connected to all the neurons of the previous layer and the parameters (weights and threshold) are not shared across the layer; this is the most general setting for a feedforward NN. In the AlexNet, the output DNN is the decision making center, which would be the analog to the abstraction stage created by the brain as the output of the visual system of a human. To speed up the learning process and reduce overfitting in the fully
25   connected layers (DNN), batch normalization and dropout regularization are employed in the AlexNet as well as the activation function ReLU.

The transfer learning process of the AlexNet (see Fig. 2b) is a two-stage optimization method. At first, the network is trained for a learning dataset (S_g) composed of the data itself (x_k) and its labels (d_k). This is achieved by solving an optimization
30   problem for all the NN's parameters that are represented by the vector sets V and W, the first layers and the last layers, respectively. The optimization problem is defined so that the cost function $\mathcal{E}$ reaches a minimum for some optimal sets of parameters $\hat{V}$ and $\hat{W}$. Once the optimal parameters are known, for the most of elements $x_k \in S_g$, the output of the NN – $J(x_k|\hat{V}, \hat{W})$ – predicts the actual label d_k, which means that the NN is ready to classify data of the same semantic field as this

of the learning set. The second stage of transfer learning consists of optimizing only a few layers amongst the last ones – or possibly all the layers – using as an initial guess the optimal sets of the general learning problem. For instance, the cost function $\mathcal{E}$ in Fig 2b may be the same for the general and specialized learning step, but for the first stage, all the parameters are allow to vary while only the W may vary during the specialization. This strategy is very efficient for images because of the structure of the CNN; the first layers process the image by extracting some features so that it becomes more abstract and the last layers attribute the probabilities of belonging to the classes. In our setup, we modified only a few of the last layers because of the low number of available training data S_s, and used original weights and thresholds in the other layers.

References

Alam, A., Shi Ji, P., and Harrison Roy, M.: Observations of new particle formation in urban air, J. Geophys. Res. Atmos., 108, 10.1029/2001JD001417, 2003.

Albrecht, B. A.: Aerosols, Cloud Microphysics, and Fractional Cloudiness, Science, 245, 1227-1230, 10.1126/science.245.4923.1227, 1989.

Amari, S.-i.: Natural Gradient Works Efficiently in Learning, Neural Comput., 10, 251-276, 10.1162/089976698300017746, 1998.

Andreae, M. O., and Rosenfeld, D.: Aerosol–cloud–precipitation interactions. Part 1. The nature and sources of cloud-active aerosols, Earth-Sci. Rev., 89, 13-41, 10.1016/j.earscirev.2008.03.001, 2008.

Asmi, E., Kivekäs, N., Kerminen, V. M., Komppula, M., Hyvärinen, A. P., Hatakka, J., Viisanen, Y., and Lihavainen, H.: Secondary new particle formation in Northern Finland Pallas site between the years 2000 and 2010, Atmos. Chem. Phys., 11, 12959-12972, 10.5194/acp-11-12959-2011, 2011.

Baranizadeh, E., Arola, A., Hamed, A., Nieminen, T., Mikkonen, S., Virtanen, A., Kulmala, M., Lehtinen, K., and Laaksonen, A.: The effect of cloudiness on new-particle formation: investigation of radiation levels, Boreal Environ. Res., 19, 343-354, 2014.

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H.: Greedy layer-wise training of deep networks, Proceedings of the 19th International Conference on Neural Information Processing Systems, Canada, 153-160, 2006.

17

Brunekreef, B., and Holgate, S. T.: Air pollution and health, Lancet, 360, 1233-1242, 10.1016/S0140-6736(02)11274-8, 2002.

Buduma, N., and Locascio, N.: Fundamentals of deep learning: Designing next-generation machine intelligence algorithms, O'Reilly Media, Sebastopol, CA, USA, 298 pp., 2017.

5   Chellapilla, K., Puri, S., and Simard., P.: High performance convolutional neural networks for document processing, Tenth International Workshop on Frontiers in Handwriting Recognition, La Baule (France), Suvisoft, <inria-00112631>, 2006.

Cireşan, D. C., Meier, U., Masci, J., Gambardella, L. M., and Schmidhuber, J.: Flexible, high performance convolutional neural networks for image classification, Proceedings of the Twenty-Second international joint conference on artificial Intelligence, Barcelona, Catalonia, Spain, 1237-1242, 10.5591/978-1-57735-516-8/ijcai11-210, 2011.

10   Cireşan, D. C., Meier, U., and Schmidhuber, J.: Transfer learning for Latin and Chinese characters with Deep Neural Networks, The 2012 International Joint Conference on Neural Networks (IJCNN), 2012, 1-6,

Dal Maso, M., Kulmala, M., Riipinen, I., Wagner, R., Hussein, T., Aalto, P. P., and Lehtinen, K. E. J.: Formation and growth of fresh atmospheric aerosols: Eight years of aerosol size distribution data from SMEAR II, Hyytiälä, Finland, Boreal Environ. Res., 10, 323-336, 2005.

15   Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 248-255, 10.1109/CVPR.2009.5206848, 2009.

Di Noia, A., Hasekamp, O. P., van Harten, G., Rietjens, J. H. H., Smit, J. M., Snik, F., Henzing, J. S., de Boer, J., Keller, C. U., and Volten, H.: Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations, Atmos. Meas. Tech., 8, 281-299, 10.5194/amt-8-281-2015, 2015.

20   Di Noia, A., Hasekamp, O. P., Wu, L., van Diedenhoven, B., Cairns, B., and Yorks, J. E.: Combined neural network/Phillips-Tikhonov approach to aerosol retrievals over land from the NASA Research Scanning Polarimeter, Atmos. Meas. Tech., 2017, 4235–4252, 10.5194/amt-10-4235-2017, 2017.

Duda, R., Hart, P., and Stork, D.: Pattern classification, 2nd edition, John Wiley & Sons, Inc., 2012.

25   Farley, B., and Clark, W.: Simulation of self-organizing systems by digital computer, Transactions of the IRE Professional Group on Information Theory, 4, 76-84, 10.1109/TIT.1954.1057468, 1954.

Hamed, A., Joutsensaari, J., Mikkonen, S., Sogacheva, L., Dal Maso, M., Kulmala, M., Cavalli, F., Fuzzi, S., Facchini, M. C., Decesari, S., Mircea, M., Lehtinen, K. E. J., and Laaksonen, A.: Nucleation and growth of new particles in Po Valley, Italy, Atmos. Chem. Phys., 7, 355-376, 2007.

30   Han, X., Zhong, Y., Cao, L., and Zhang, L.: Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification, Remote Sens., 9, 848, 10.3390/rs9080848, 2017.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778, 2016.

Hebb, D. O.: The organization of behavior: A neuropsychological theory, Psychology Press, 2005.

18

Herrmann, E., Weingartner, E., Henne, S., Vuilleumier, L., Bukowiecki, N., Steinbacher, M., Conen, F., Collaud Coen, M., Hammer, E., Jurányi, Z., Baltensperger, U., and Gysel, M.: Analysis of long-term aerosol size distribution data from Jungfraujoch with emphasis on free tropospheric conditions, cloud influence, and air mass transport, J. Geophys. Res. Atmos., 120, 9459-9480, 10.1002/2015JD023660, 2015.

5 Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R.: Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint, arXiv:1207.0580, 2012.

Hirsikko, A., Bergman, T., Laakso, L., Dal Maso, M., Riipinen, I., Hõrrak, U., and Kulmala, M.: Identification and classification of the formation of intermediate ions measured in boreal forest, Atmos. Chem. Phys., 7, 201-210, 10.5194/acp-7-201-2007, 2007.

10 Hu, F., Xia, G.-S., Hu, J., and Zhang, L.: Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery, Remote Sens., 7, 14680, 2015.

Huber, P. J.: Robust Statistics, John Wiley, New York, 1981.

Huttunen, J., Kokkola, H., Mielonen, T., Mononen, M. E. J., Lipponen, A., Reunanen, J., Lindfors, A. V., Mikkonen, S., Lehtinen, K. E. J., Kouremeti, N., Bais, A., Niska, H., and Arola, A.: Retrieval of aerosol optical depth from surface solar

15 radiation measurements using machine learning algorithms, non-linear regression and a radiative transfer-based look-up table, Atmos. Chem. Phys., 16, 8181-8191, 10.5194/acp-16-8181-2016, 2016.

Hyvönen, S., Junninen, H., Laakso, L., Dal Maso, M., Grönholm, T., Bonn, B., Keronen, P., Aalto, P., Hiltunen, V., Pohja, T., Launiainen, S., Hari, P., Mannila, H., and Kulmala, M.: A look at aerosol formation using data mining techniques, Atmos. Chem. Phys., 5, 3345-3356, 10.5194/acp-5-3345-2005, 2005.

20 IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., 2013.

Jeong, C.-H., Hopke, P. K., Chalupa, D., and Utell, M.: Characteristics of Nucleation and Growth Events of Ultrafine Particles Measured in Rochester, NY, Environ. Sci. Technol., 38, 1933-1940, 10.1021/es034811p, 2004.

25 Junninen, H., Riipinen, I., Dal Maso, M., and Kulmala, M.: An algorithm for automatic classification of two-dimensional aerosol data, in: Nucleation and Atmospheric Aerosols, Springer, Dordrecht, 957-961, 2007.

Kaelbling, L. P., Littman, M. L., and Moore, A. W.: Reinforcement learning: a survey, J. Artif. Int. Res., 4, 237-285, 10.1613/jair.301, 1996.

Kerminen, V. M., Paramonov, M., Anttila, T., Riipinen, I., Fountoukis, C., Korhonen, H., Asmi, E., Laakso, L., Lihavainen,

30 H., Swietlicki, E., Svenningsson, B., Asmi, A., Pandis, S. N., Kulmala, M., and Petäjä, T.: Cloud condensation nuclei production associated with atmospheric nucleation: a synthesis based on existing literature and new results, Atmos. Chem. Phys., 12, 12037-12059, 10.5194/acp-12-12037-2012, 2012.

Kim, Y., Yoon, S.-C., Kim, S.-W., Kim, K.-Y., Lim, H.-C., and Ryu, J.: Observation of new particle formation and growth events in Asian continental outflow, Atmos. Environ., 64, 160-168, 10.1016/j.atmosenv.2012.09.057, 2013.

Kim, Y., Kim, S.-W., and Yoon, S.-C.: Observation of new particle formation and growth under cloudy conditions at Gosan Climate Observatory, Korea, Meteorol. Atmos. Phys., 126, 81-90, 10.1007/s00703-014-0336-2, 2014.

Kivekäs, N., Sun, J., Zhan, M., Kerminen, V. M., Hyvärinen, A., Komppula, M., Viisanen, Y., Hong, N., Zhang, Y., Kulmala, M., Zhang, X. C., Deli, G., and Lihavainen, H.: Long term particle size distribution measurements at Mount Waliguan, a high-altitude site in inland China, Atmos. Chem. Phys., 9, 5461-5474, 10.5194/acp-9-5461-2009, 2009.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems 25, edited by: Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., Curran Associates, Inc., 1097-1105, 2012.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, Commun. ACM, 60, 84-90, 10.1145/3065386, 2017.

Kulmala, M., Vehkamäki, H., Petäjä, T., Dal Maso, M., Lauri, A., Kerminen, V. M., Birmili, W., and McMurry, P. H.: Formation and growth rates of ultrafine atmospheric particles: a review of observations, J. Aerosol Sci., 35, 143-176, 2004.

Kulmala, M., Petäjä, T., Nieminen, T., Sipilä, M., Manninen, H. E., Lehtipalo, K., Dal Maso, M., Aalto, P. P., Junninen, H., Paasonen, P., Riipinen, I., Lehtinen, K. E. J., Laaksonen, A., and Kerminen, V.-M.: Measurement of the nucleation of atmospheric aerosol particles, Nat. Protoc., 7, 1651-1667, 10.1038/nprot.2012.091, 2012.

Kulmala, M., Petäjä, T., Kerminen, V.-M., Kujansuu, J., Ruuskanen, T., Ding, A., Nie, W., Hu, M., Wang, Z., Wu, Z., Wang, L., and Worsnop, D. R.: On secondary new particle formation in China, Front. Env. Sci. Eng., 10, 8, 10.1007/s11783-016-0850-1, 2016.

Kyrö, E. M., Väänänen, R., Kerminen, V. M., Virkkula, A., Petäjä, T., Asmi, A., Dal Maso, M., Nieminen, T., Juhola, S., Shcherbinin, A., Riipinen, I., Lehtipalo, K., Keronen, P., Aalto, P. P., Hari, P., and Kulmala, M.: Trends in new particle formation in eastern Lapland, Finland: effect of decreasing sulfur emissions from Kola Peninsula, Atmos. Chem. Phys., 14, 4383-4396, 10.5194/acp-14-4383-2014, 2014.

Laaksonen, A., Hamed, A., Joutsensaari, J., Hiltunen, L., Cavalli, F., Junkermann, W., Asmi, A., Fuzzi, S., and Facchini, M. C.: Cloud condensation nucleus production from nucleation events at a highly polluted region, Geophys. Res. Lett., 32, L06812, 10.1029/2004gl022092, 2005.

Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L.: Machine learning in geosciences and remote sensing, Geosci. Front., 7, 3-10, 10.1016/j.gsf.2015.07.003, 2016.

Le, Q. V.: Building high-level features using large scale unsupervised learning, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, 8595-8598,

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86, 2278-2324, 10.1109/5.726791, 1998.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436-444, 10.1038/nature14539, 2015.

Li, X., Peng, L., Hu, Y., Shao, J., and Chi, T.: Deep learning architecture for air quality predictions, Environ. Sci. Pollut. Res. Int., 23, 22408-22417, 10.1007/s11356-016-7812-9, 2016.

Lähivaara, T., Kärkkäinen, L., Huttunen, J. M. J., and Hesthaven, J. S.: Deep convolutional neural networks for estimating porous material parameters with ultrasound tomography, The Journal of the Acoustical Society of America, 143, 1148-5   1158, 10.1121/1.5024341, 2018.

Ma, Y., Gong, W., and Mao, F.: Transfer learning used to analyze the dynamic evolution of the dust aerosol, J. Quant. Spectrosc. Radiat. Transfer, 153, 119-130, 10.1016/j.jqsrt.2014.09.025, 2015.

MathWorks: Deep Learning: Transfer Learning in 10 lines of MATLAB Code: https://se.mathworks.com/matlabcentral/fileexchange/61639-deep-learning--transfer-learning-in-10-lines-of-matlab-code, 10   last access: 29.11.2017, 2017.

MathWorks: MathWorks Documentation: imresize: https://se.mathworks.com/help/images/ref/imresize.html, last access: 14.6.2018, 2018.

Merikanto, J., Spracklen, D. V., Mann, G. W., Pickering, S. J., and Carslaw, K. S.: Impact of nucleation on global CCN, Atmos. Chem. Phys., 9, 8601-8616, 10.5194/acp-9-8601-2009, 2009.

15   Mesnil, G., Dauphin, Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I., Lavoie, E., Muller, X., Desjardins, G., Warde-Farley, D., Vincent, P., Courville, A., and Bergstra, J.: Unsupervised and Transfer Learning Challenge: a Deep Learning Approach, Proceedings of Machine Learning Research, Proceedings of ICML Workshop on Unsupervised and Transfer Learning, 27, 97-110, 2012.

Mikkonen, S., Lehtinen, K. E. J., Hamed, A., Joutsensaari, J., Facchini, M. C., and Laaksonen, A.: Using discriminant 20   analysis as a nucleation event classification method, Atmos. Chem. Phys., 6, 5549-5557, 10.5194/acp-6-5549-2006, 2006.

Mikkonen, S., Korhonen, H., Romakkaniemi, S., Smith, J. N., Joutsensaari, J., Lehtinen, K. E. J., Hamed, A., Breider, T. J., Birmili, W., Spindler, G., Plass-Duelmer, C., Facchini, M. C., and Laaksonen, A.: Meteorological and trace gas factors affecting the number concentration of atmospheric Aitken (Dp = 50 nm) particles in the continental boundary layer: 25   parameterization using a multivariate mixed effects model, Geosci. Model Dev., 4, 1-13, 10.5194/gmd-4-1-2011, 2011.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M.: Playing atari with deep reinforcement learning., arXiv preprint, arXiv:1312.5602, 2013.

Mäkelä, J. M., Dal Maso, M., Pirjola, L., Keronen, P., Laakso, L., Kulmala, M., and Laaksonen, A.: Characteristics of the atmospheric particle formation events observed at a borel forest site in southern Finland, Boreal Environ. Res., 5, 299-30   313, 2000.

Nel, A.: Air Pollution-Related Illness: Effects of Particles, Science, 308, 804-806, 10.1126/science.1108752, 2005.

Nieminen, T., Asmi, A., Dal Maso, M., Aalto, P. P., Keronen, P., Petaja, T., Kulmala, M., and Kerminen, V. M.: Trends in atmospheric new-particle formation: 16 years of observations in a boreal-forest environment, Boreal Environ. Res., 19, 191-214, 2014.

Nieminen, T., Yli-Juuti, T., Manninen, H. E., Petäjä, T., Kerminen, V. M., and Kulmala, M.: Technical note: New particle formation event forecasts during PEGASOS–Zeppelin Northern mission 2013 in Hyytiälä, Finland, Atmos. Chem. Phys., 15, 12385-12396, 10.5194/acp-15-12385-2015, 2015.

5   Nieminen, T., Kerminen, V. M., Petäjä, T., Aalto, P. P., Arshinov, M., Asmi, E., Baltensperger, U., Beddows, D. C. S., Beukes, J. P., Collins, D., Ding, A., Harrison, R. M., Henzing, B., Hooda, R., Hu, M., Hõrrak, U., Kivekäs, N., Komsaare, K., Krejci, R., Kristensson, A., Laakso, L., Laaksonen, A., Leaitch, W. R., Lihavainen, H., Mihalopoulos, N., Németh, Z., Nie, W., O'Dowd, C., Salma, I., Sellegri, K., Svenningsson, B., Swietlicki, E., Tunved, P., Ulevicius, V., Vakkari, V., Vana, M., Wiedensohler, A., Wu, Z., Virtanen, A., and Kulmala, M.: Global analysis of continental boundary layer new particle formation based on long-term measurements, Atmos. Chem. Phys. Discuss., 2018, 1-34,

10   10.5194/acp-2018-304, 2018.

O'Dowd, C., Monahan, C., and Dall'Osto, M.: On the occurrence of open ocean particle production and growth events, Geophys. Res. Lett., 37, doi:10.1029/2010GL044679, 2010.

Ong, B. T., Sugiura, K., and Zettsu, K.: Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM2.5, Neural Comput. Appl., 27, 1553-1566, 10.1007/s00521-015-1955-3, 2016.

15   Pan, S. J., and Yang, Q.: A Survey on Transfer Learning, IEEE T. Knowl. Data Eng., 22, 1345-1359, 10.1109/TKDE.2009.191, 2010.

Pascanu, R., Mikolov, T., and Bengio, Y.: On the difficulty of training recurrent neural networks, Proceedings of the 30th International Conference on Machine Learning, Proceedings of Machine Learning Research, 28, 1310-1318, 2013.

Qi, X. M., Ding, A. J., Nie, W., Petäjä, T., Kerminen, V. M., Herrmann, E., Xie, Y. N., Zheng, L. F., Manninen, H., Aalto,

20   P., Sun, J. N., Xu, Z. N., Chi, X. G., Huang, X., Boy, M., Virkkula, A., Yang, X. Q., Fu, C. B., and Kulmala, M.: Aerosol size distribution and new particle formation in the western Yangtze River Delta of China: 2 years of measurements at the SORPES station, Atmos. Chem. Phys., 15, 12445-12464, 10.5194/acp-15-12445-2015, 2015.

R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria: https://www.R-project.org/, 2017.

25   Radford, A., Metz, L., and Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint, arXiv:1511.06434v2, 2016.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning internal representations by error propagation, in: Parallel distributed processing: explorations in the microstructure of cognition, vol. 1, edited by: David, E. R., James, L. M., and Group, C. P. R., MIT Press, 318-362, 1986.

30   Ruske, S., Topping, D. O., Foot, V. E., Kaye, P. H., Stanley, W. R., Crawford, I., Morse, A. P., and Gallagher, M. W.: Evaluation of machine learning algorithms for classification of primary biological aerosol using a new UV-LIF spectrometer, Atmos. Meas. Tech., 10, 695-708, 10.5194/amt-10-695-2017, 2017.

Samet , J. M., Dominici , F., Curriero , F. C., Coursac , I., and Zeger , S. L.: Fine Particulate Air Pollution and Mortality in 20 U.S. Cities, 1987–1994, New Engl. J. Med., 343, 1742-1749, 10.1056/nejm200012143432401, 2000.

Schmeissner, T., Krejci, R., Ström, J., Birmili, W., Wiedensohler, A., Hochschild, G., Gross, J., Hoffmann, P., and Calderon, S.: Analysis of number size distributions of tropical free tropospheric aerosol particles observed at Pico Espejo (4765 m a.s.l.), Venezuela, Atmos. Chem. Phys., 11, 3319-3332, 10.5194/acp-11-3319-2011, 2011.

Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Networks, 61, 85-117, 10.1016/j.neunet.2014.09.003, 2015.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M.: Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, IEEE T. Med. Imaging, 35, 1285-1298, 10.1109/TMI.2016.2528162, 2016.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., and Hassabis, D.: Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm, arXiv preprint, arXiv:1712.01815, 2017a.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., and Hassabis, D.: Mastering the game of Go without human knowledge, Nature, 550, 354-359, 10.1038/nature24270, 2017b.

Spracklen, D. V., Carslaw, K. S., Kulmala, M., Kerminen, V. M., Sihto, S. L., Riipinen, I., Merikanto, J., Mann, G. W., Chipperfield, M. P., Wiedensohler, A., Birmili, W., and Lihavainen, H.: Contribution of particle formation to global cloud condensation nuclei concentrations, Geophys. Res. Lett., 35, L06808, 10.1029/2007gl033038, 2008.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res., 15, 1929-1958, 2014.

Sutton, R. S.: Learning to Predict by the Methods of Temporal Differences, Mach. Learn., 3, 9-44, 10.1023/a:1022633531479, 1988.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7-12 June 2015, 1-9, 10.1109/CVPR.2015.7298594, 2015.

Twomey, S.: Pollution and the planetary albedo, Atmos. Environ., 8, 1251-1256, doi.org/10.1016/0004-6981(74)90004-3, 1974.

Vaattovaara, P., Huttunen, P. E., Yoon, Y. J., Joutsensaari, J., Lehtinen, K. E. J., O'Dowd, C. D., and Laaksonen, A.: The composition of nucleation and Aitken modes particles during coastal nucleation events: evidence for marine secondary organic contribution, Atmos. Chem. Phys., 6, 4601-4616, 2006.

Vana, M., Ehn, M., Petäjä, T., Vuollekoski, H., Aalto, P., de Leeuw, G., Ceburnis, D., O'Dowd, C. D., and Kulmala, M.: Characteristic features of air ions at Mace Head on the west coast of Ireland, Atmos. Res., 90, 278-286, 10.1016/j.atmosres.2008.04.007, 2008.

Wang, J.-F., Hu, M.-G., Xu, C.-D., Christakos, G., and Zhao, Y.: Estimation of Citywide Air Pollution in Beijing, PLOS ONE, 8, e53400, 10.1371/journal.pone.0053400, 2013.
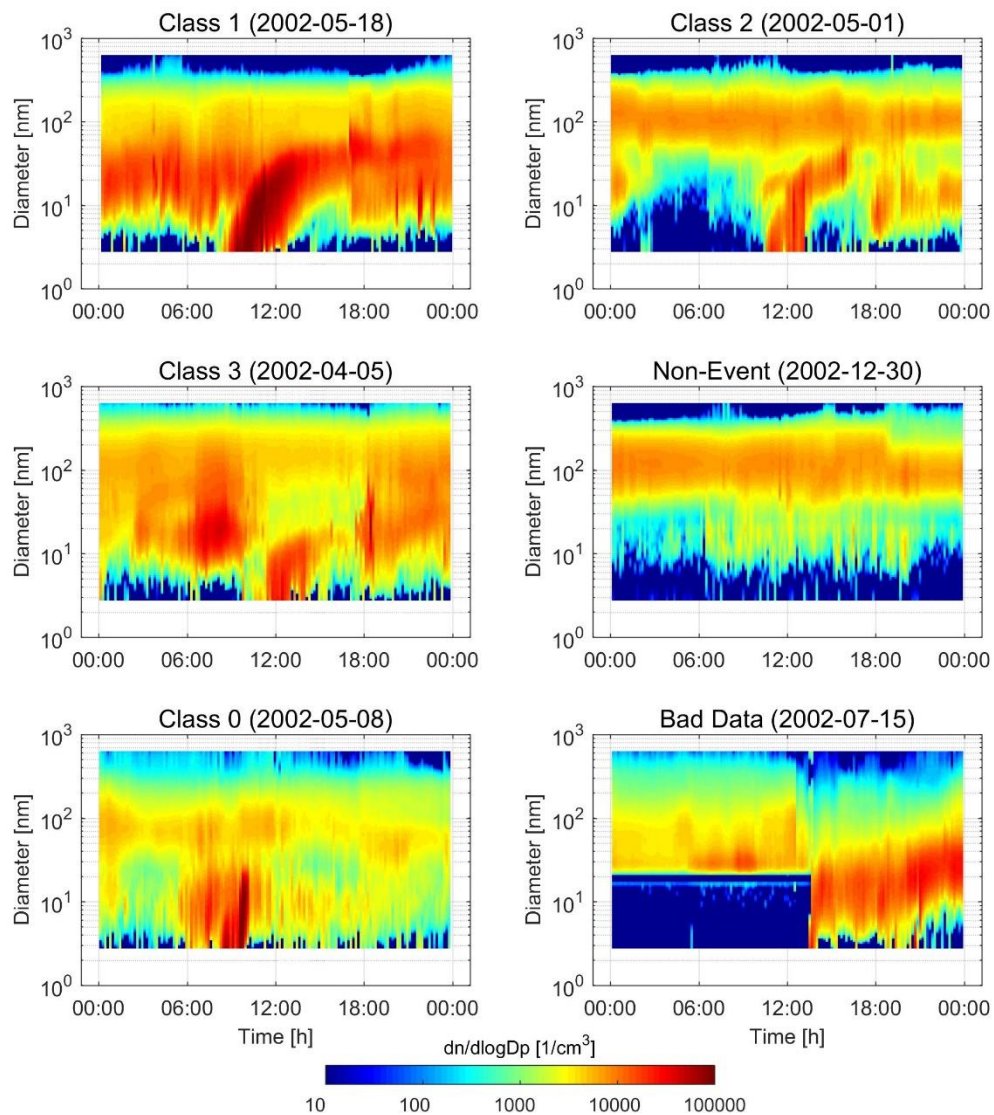
Wang, Z., Birmili, W., Hamed, A., Wehner, B., Spindler, G., Pei, X., Wu, Z., Cheng, Y., Su, H., and Wiedensohler, A.: Contributions of volatile and nonvolatile compounds (at 300°C) to condensational growth of atmospheric nanoparticles: An assessment based on 8.5 years of observations at the Central Europe background site Melpitz, J. Geophys. Res. Atmos., 122, 485-497, 10.1002/2016JD025581, 2017a.

5 Wang, Z., Wu, Z., Yue, D., Shang, D., Guo, S., Sun, J., Ding, A., Wang, L., Jiang, J., Guo, H., Gao, J., Cheung, H. C., Morawska, L., Keywood, M., and Hu, M.: New particle formation in China: Current knowledge and further directions, Sci. Total Environ., 577, 258-266, 10.1016/j.scitotenv.2016.10.177, 2017b.

Weiss, K., Khoshgoftaar, T. M., and Wang, D.: A survey of transfer learning, J. Big Data, 3, 9, 10.1186/s40537-016-0043-6, 2016.

10 Venables, W. N., and Ripley, B. D.: Modern Applied Statistics with S, Fourth Edition, Springer, New York, 2002.

WHO: Review of evidence on health aspects of air pollution - REVIHAAP Project: Technical Report, World Health Organization, WHO Regional Office for Europe, Copenhagen, 2013.

Widrow, B., and Hoff, M. E.: Adaptive Switching Circuits, Technical Report No. 1553-1, Stanford Electronics Laboratories, Stanford University, CA, 1960.

15 Winklmayr, W., Reischl, G. P., Lindner, A. O., and Berner, A.: A new electromobility spectrometer for the measurement of aerosol size distributions in the size range from 1 to 1000 nm, J. Aerosol Sci., 22, 289-296, 1991.

Voutilainen, A., Kolehmainen, V., and Kaipio, J. P.: Statistical inversion of aerosol size measurement data, Inverse Probl. Eng., 9, 67-94, 2001.

Vuollekoski, H., Junninen, H., Dal Maso, M., Kerminen, V.-M., and Kulmala, M.: An eigenvector-based approach for

20 automatic classification of new particle formation events, in: Proceedings of Finnish Center of Excellence in 'Physics, Chemistry, Biology and Meteorology of Atmospheric Composition and Climate Change', and Nordic Center of Excellence in 'Cryosphere-Atmosphere Interactions in a Changing Arctic Climate', Annual Meetings 2012, edited by: Kulmala, M., Lappalainen, H. K., Boy, M., Brus, M., and Nieminen, T., Report Series in Aerosol Science, N:o 134, Helsinki, Finland, 653-654, 2012. http://www.atm.helsinki.fi/FAAR/reportseries/rs-134.pdf.

25 Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.: How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems 27, edited by: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., 3320-3328, 2014.

Yu, F., and Luo, G.: Simulation of particle size distribution with a global aerosol model: contribution of nucleation to aerosol and CCN number concentrations, Atmos. Chem. Phys., 9, 7691-7710, 10.5194/acp-9-7691-2009, 2009.

30 Zaidan, M. A., Haapasilta, V., Relan, R., Junninen, H., Aalto, P. P., Canova, F. F., Laurson, L., and Foster, A. S.: Neural network classifier on time series features for predicting atmospheric particle formation days, in: The 20th International Conference on Nucleation and Atmospheric Aerosols, edited by: Halonen, R., Nikandrova, A., Kontkanen, J., Enroth, J. A., and Vehkamäki, H., Report Series in Aerosol Science, N:o 200, Helsinki, Finland, 687-690, 2017. http://www.atm.helsinki.fi/FAAR/reportseries/rs-200.pdf.
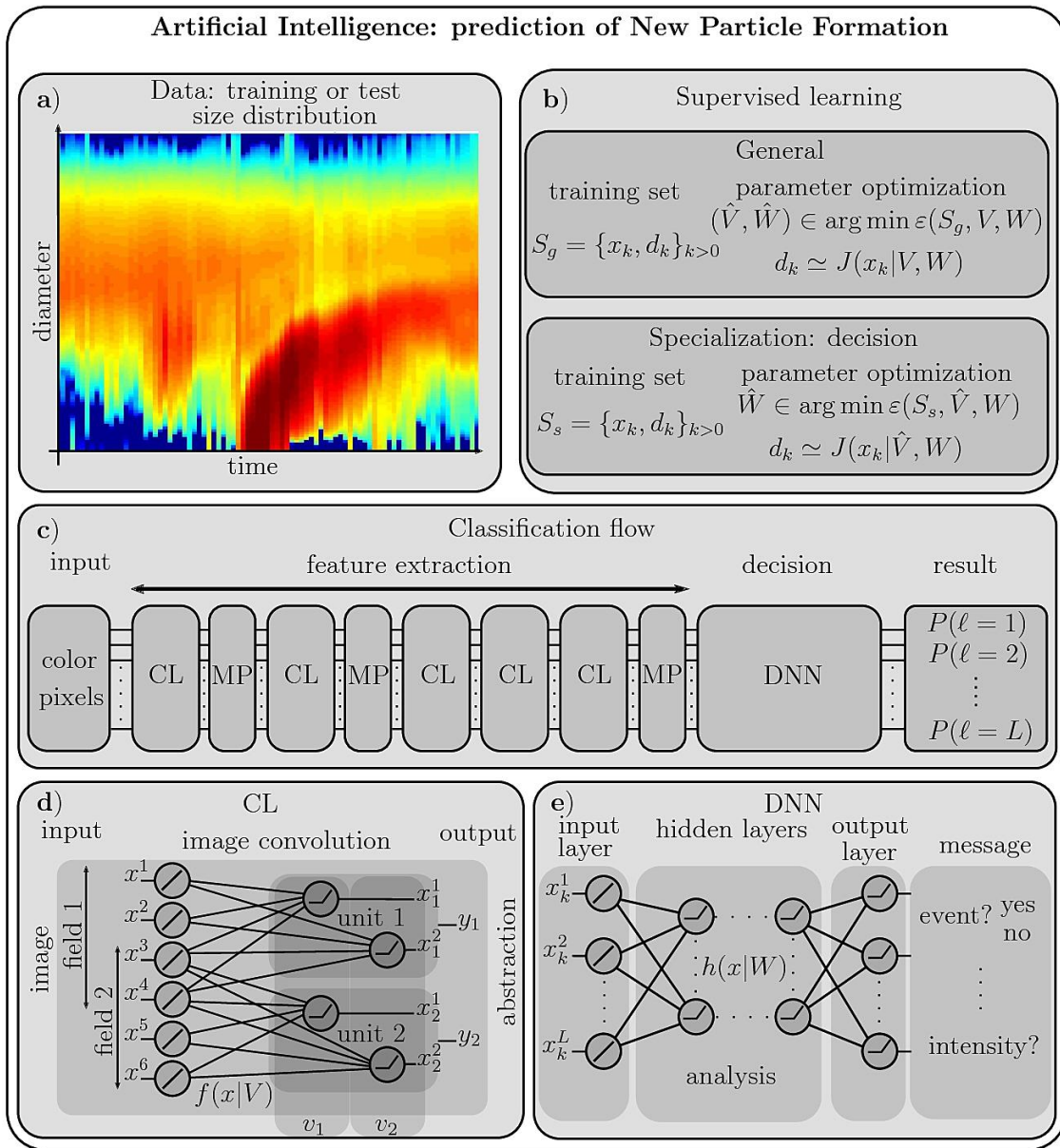
Zeiler, M. D., and Fergus, R.: Visualizing and Understanding Convolutional Networks, in: Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I, edited by: Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., Springer International Publishing, Cham, 818-833, 2014. https://doi.org/10.1007/978-3-319-10590-1_53.

5   Zhang, L. P., Zhang, L. F., and Du, B.: Deep learning for remote sensing data: A technical tutorial on the state of the art, IEEE Geoscience and Remote Sensing Magazine, 4, 22-40, 10.1109/mgrs.2016.2540798, 2016.

Figures:

Figure 1: Examples of particle size distributions (time in x-axis, particle diameter in y-axis and particle concentration presented by different colors) for different NPF event classes. Name of event class is indicated in plot title and date of measurement in brackets.

5

Figure 2: Visual summary of the classification method: a) a typical dataset with a NPF event, b) the learning process of the optimization problem, c) the classification flow of the Convolutional Neural Network (CNN) and the two types of layers of the CNN, d) the Convolutional Layer (CL) and e) the Deep Neural Network (DNN), involved in the total CNN. The classification flow (c) is from the AlexNet CNN (Krizhevsky et al., 2012) composed of 5 CL intertwined with Max Pooling (MP) layers followed by a fully connected DNN of 3 layers. The method and variables in the figure are described in detail in Chapter 2.3 and Appendix A.

Figure 3: Confusion matrices of a) training and b) testing data sets from one run of Image Set 1 with training-testing ratio of 50 %/50 %. Numbers (percent and absolute number of days) in rows indicate how a certain class classified by human (visually) were categorized in the CNN classification.
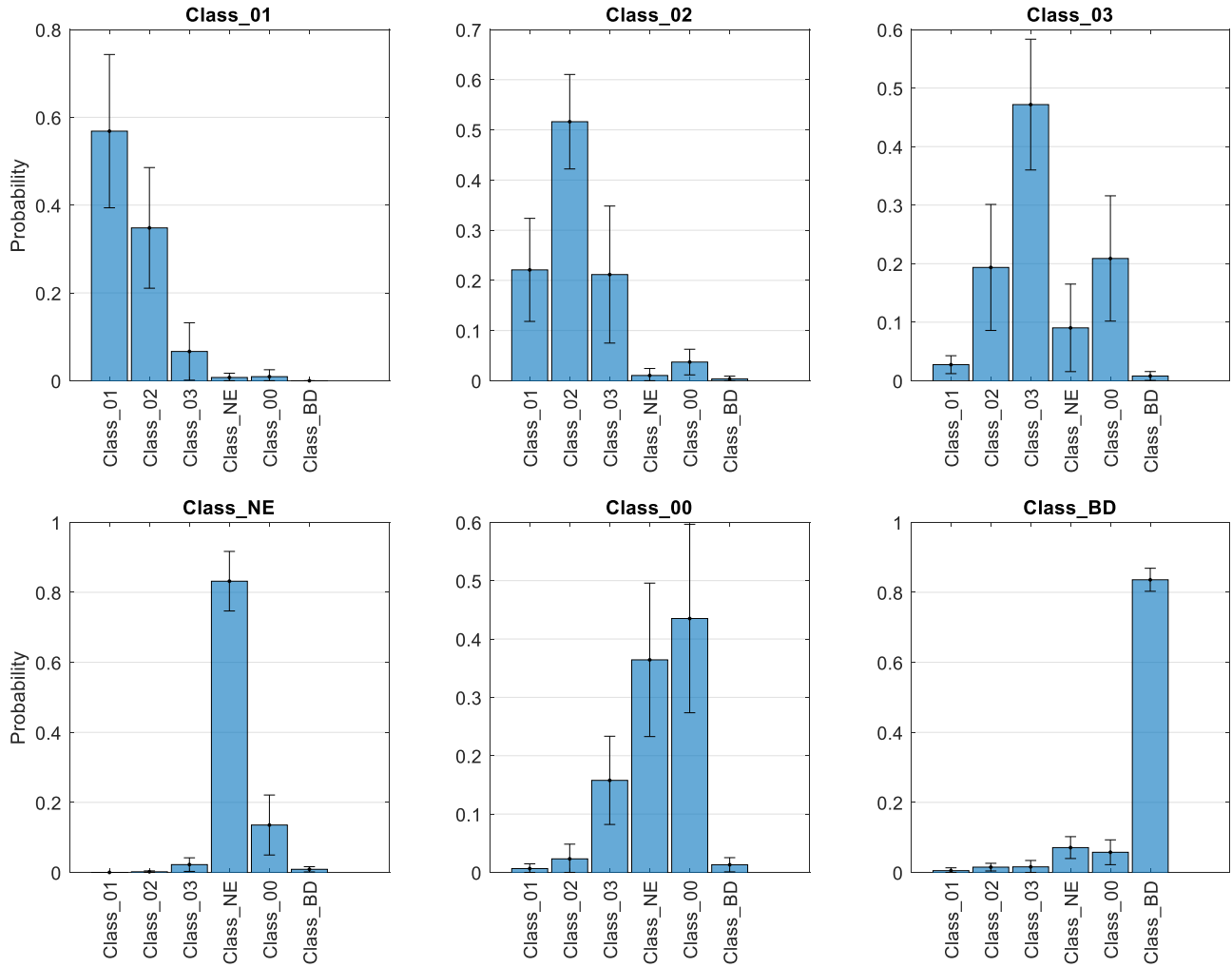
Figure 4: Probability plots of distribution of CNN classification (x-axis labels) for different human-made classifications (indicated in a title of the plot), for Image Set 1 with training-testing ratio of 80 %/20 %. Histograms are mean values of ten different training-testing runs and error bars indicate standard deviations of the results.

Figure 5: Examples of particle size distributions for different days when visually categorized clear event days (Classes 1 and 2) are classified Non-Event days by CNN classification or vice versa. In title, numbers after *Obs* and *Pred* show NPF event classes determined with visual and CNN-based classification, respectively (date of measurement in brackets). Several event class numbers in Pred-cases are results from different calculation runs. The first row (a and b) contains examples of human made misclassifications, the second row (c and d) CNN misclassifications and the third row (e and f) ambiguous situations (low particle concentrations and change in air masses during day, respectively). CNN classification is from Image Set 1 with training-testing ratio of 80 %/20 %.

Tables:

Table 1. <mark>Summary of number of days in different NPF event classes based on manual classification and division of days to three different training (used in CNN learning) and testing (used in testing success of training) ratio categories (80 %/20 %, 50 %/50% and 20 %/80 %).</mark>

Event class	Days Total	Days %	Training 80 %	Testing 20 %	Training 50 %	Testing 50 %	Training 20 %	Testing 80 %
Class_01	269	6 %	215	54	135	134	54	215
Class_02	431	10 %	345	86	216	215	86	345
Class_03	619	15 %	495	124	310	309	124	495
Class_NE	1416	34 %	1133	283	708	708	283	1133
Class_00	841	20 %	673	168	421	420	168	673
Class_BD	601	14 %	481	120	301	300	120	481
Total	4177	100 %	3342	835	2091	2086	835	3342

Table 2. Summary of classification accuracy (%) of the transfer learned CNN when applied for test datasets (mean value ± standard deviation from ten model simulations) for three image sets with three different percentage of training images. Some classes have been merged together: CL_01-02 is a combination of Class_01 and Class_02, etc.

Event class	Image Set 1 80 %	Image Set 1 50 %	Image Set 1 20 %	Image Set 2 80 %	Image Set 2 50 %	Image Set 2 20 %	Image Set 3 80 %	Image Set 3 50 %	Image Set 3 20 %	Overall
All_Class	65 ± 1	64 ± 1	61 ± 1	65 ± 1	64 ± 1	60 ± 3	62 ± 3	63 ± 1	60 ± 2	63 ± 2
Class_01	57 ± 17	54 ± 20	51 ± 15	63 ± 11	38 ± 10	53 ± 27	54 ± 17	54 ± 10	47 ± 22	52 ± 7
Class_02	52 ± 9	51 ± 15	31 ± 17	40 ± 17	57 ± 14	27 ± 19	47 ± 20	45 ± 14	45 ± 21	44 ± 10
Class_03	47 ± 11	32 ± 11	50 ± 14	46 ± 12	42 ± 14	48 ± 14	40 ± 10	44 ± 10	40 ± 16	43 ± 6
Class_NE	83 ± 9	80 ± 8	81 ± 13	80 ± 10	80 ± 10	81 ± 18	77 ± 12	81 ± 10	79 ± 11	80 ± 2
Class_00	44 ± 16	53 ± 13	38 ± 18	51 ± 13	48 ± 12	36 ± 28	49 ± 12	43 ± 14	41 ± 14	45 ± 6
Class_BD	84 ± 3	85 ± 3	83 ± 3	85 ± 4	86 ± 2	80 ± 2	83 ± 4	84 ± 2	80 ± 2	83 ± 2
Cl_01-02	81 ± 11	85 ± 9	62 ± 19	76 ± 11	78 ± 12	62 ± 19	81 ± 11	77 ± 13	71 ± 24	75 ± 8
Cl_01-02-03	84 ± 6	79 ± 6	81 ± 6	82 ± 3	81 ± 5	80 ± 8	83 ± 6	82 ± 5	82 ± 3	81 ± 2
Cl_03-NE	77 ± 10	69 ± 10	78 ± 12	74 ± 8	73 ± 7	77 ± 18	70 ± 10	75 ± 10	72 ± 10	74 ± 3
Cl_03-00	63 ± 10	60 ± 7	61 ± 13	66 ± 10	63 ± 13	60 ± 18	62 ± 10	60 ± 9	58 ± 14	61 ± 2
Cl_NE-00	90 ± 5	93 ± 3	90 ± 4	92 ± 2	91 ± 3	91 ± 5	91 ± 3	90 ± 4	90 ± 2	91 ± 1
Cl_03-NE-00	93 ± 4	91 ± 4	94 ± 4	93 ± 3	93 ± 3	95 ± 3	92 ± 4	93 ± 3	91 ± 6	93 ± 1

Identification of new particle formation events with deep learning

Jorma Joutsensaari[1], Matthew Ozon[1], Tuomo Nieminen[1], Santtu Mikkonen[1], Timo Lähivaara[1], Stefano Decesari[2], M. Cristina Facchini[2], Ari Laaksonen[1,3] and Kari E.J. Lehtinen[1,4]

[1]Department of Applied Physics, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland
5 [2]Institute of Atmospheric Sciences and Climate of the Italian National Research Council, Bologna, Italy
[3]Climate research Unit, Finnish Meteorological Institute, Helsinki, Finland
[4]Atmospheric Research Centre of Eastern Finland, Finnish Meteorological Institute, Kuopio, Finland

Correspondence to: Jorma Joutsensaari (jorma.joutsensaari@uef.fi)

Abstract. New particle formation (NPF) in the atmosphere is globally an important source of climate relevant aerosol particles.
10 Occurrence of NPF events is typically analyzed manually by researchers from particle size distribution data day by day, which is time consuming and the classification of event types may be inconsistent. To get more reliable and consistent results, the NPF event analysis should be automatized. We have developed an automatic analysis method based on deep learning, a subarea of machine learning, for NPF event identification. To our knowledge, this is the first time when ~~NPF events have been~~ a deep learning method, i.e. transfer learning of a Convolutional Neural Networks (CNN), has successfully ~~classified~~been used to 15 classify automatically NPF events into different classes directly from particle size distribution images, similarly as the researchers do in the manual classification. The developed method is based on image analysis of particle size distributions using a pre-trained deep ~~Convolutional Neural Networks (CNN),~~CNN, named AlexNet, which was transfer learned to recognize NPF event classes (six different types). In transfer learning, a partial set of particle size distribution images ~~were~~was used in the training stage of the CNN and the rest of images for testing the success of the training. The method was utilized 20 for a 15-year long dataset measured at San Pietro Capofiume (SPC) in Italy. We studied the performance of the training with different training and testing image number ratios as well as with different regions of interest in the images. The results show that clear event (i.e., Classes 1 and 2) and non-event days can be identified with an accuracy of ca. 80 %, when the CNN classification is compared with that of an expert, which is a good first result for automatic NPF event analysis. In the event classification, the choice between different event classes is not an easy task even for trained researchers, thus overlapping or 25 confusion between different classes occurs. Hence, we cross validated the learning results of CNN with the expert made classification. The results show that the overlapping occurs typically between the adjacent or similar type of classes, e.g., a manually classified Class 1 is categorized mainly into Classes 1 and 2 by CNN, indicating that the manual and CNN classifications are very consist for the most of the days. The classification would be more consistent, by both human and CNN, if only two different classes are used for event days instead of three classes. Thus, we recommend that in the future analysis, 30 event days should be categorized into classes of "Quantifiable" (i.e. clear events, Classes 1 and 2) and "Non-Quantifiable" (i.e. weak events, Class 3). This would better describe the difference of those classes: both formation and growth rates can be determined for Quantifiable days but not both for Non-Quantifiable days. Furthermore, we investigated more deeply the days

that are classified as clear events by experts and recognized as non-events by the CNN and vice versa. Clear misclassifications seem to occur more commonly in manual analysis than in the CNN categorization, which is mostly due to the inconsistency in the human-made classification or errors in the booking of the event class. In general, the automatic CNN classifier has a better reliability and repeatability in NPF event classification than human-made classification and, thus, the transfer learned pre-trained CNNs are powerful tools to analyze long-term datasets. The developed NPF event classifier can be easily utilized to analyze any long-term datasets more accurately and consistently, which helps us to understand in detail aerosol-climate interactions and the long-term effects of climate change on NPF in the atmosphere. We encourage researchers to use the model in other sites. However, we suggest that the CNN should be transfer learned again for new site data with a minimum of ca. 150 figures per class to obtain good enough classification results, especially if the size distribution evolution differs from training data. In the future, we will utilize the method for data from other sites, develop it to analyze more parameters and evaluate how successfully CNN could be trained with synthetic NPF event data.

1 Introduction

Aerosol particles have various effects on air quality, human health and the global climate (Nel, 2005; WHO, 2013; IPCC, 2013). The air quality and health related problems are connected to each other, since in urban areas human exposure to elevated levels of particulate matter has been shown to cause respiratory problems and cardiovascular diseases (Brunekreef and Holgate, 2002), and eventually increase mortality (Samet et al., 2000). Very small particles like ultra-fine particles (smaller than 100 nm in diameter) can be particularly harmful because they can efficiently penetrate into the respiratory system and cause systemic effects (Nel, 2005). Air quality also affects visibility, for example, during smog episodes in large cities in Asia (Wang et al., 2013). On the global scale, aerosols affect the radiative balance of the Earth and therefore the climate both. They affect the climate directly and indirectly. The direct effect is caused by the aerosol particles by either scattering incoming solar radiation back to space or by absorbing it radiation. Indirectly aerosol, aerosols affect the climate via their role in cloud formation as cloud condensation nuclei (CCN). The number concentration and chemical properties of CCN particles affect both the brightness of clouds (Twomey, 1974) and their lifetime (Albrecht, 1989). Increased number of CCN are associated to smaller cloud droplets, which can lead to brighter and longer-lived clouds (Andreae and Rosenfeld, 2008). It has been estimated that both the direct and indirect aerosol climate effects cause a net cooling of the climate, and can therefore cancel out some part of the global warming caused by greenhouse gases (IPCC, 2013).

Atmospheric new particle formation (NPF; formation and growth of secondary aerosol particles) is observed frequently in different environments in the planetary boundary layer (e.g. Kulmala et al., 2004). There are direct observations that NPF can increase the concentration of CCN particles regionally (Kerminen et al., 2012). Based on global aerosol model studies, it is estimated that 30–50% of global tropospheric CCN concentrations might be formed by atmospheric NPF (Spracklen et al., 2008; Merikanto et al., 2009; Yu and Luo, 2009). The longest continuous observational datasets of atmospheric NPF have

2

been collected in Finland at the SMEAR stations in Hyytiälä and Värriö starting in 1996 and 1997, respectively (Kyrö et al., 2014; Nieminen et al., 2014), and at the GAW station in Pallas from 2000 onwards (Asmi et al., 2011). These three stations are located in the Northern European boreal forest, and can be considered to represent rural and remote environments. In more anthropogenically influenced environments long-term NPF observations have been performed in Central Europe at Melpitz

5 (Wang et al., 2017a) and in San Pietro Capofiume at the Po Valley basin in Northern Italy (Laaksonen et al., 2005; Hamed et al., 2007; Mikkonen et al., 2011). In high-altitude sites, which are at least part-time in the free troposphere, there are less and shorter continuous measurements of NPF available. However, also in these high-altitude sites NPF has been observed to occur regularly (Kivekäs et al., 2009; Schmeissner et al., 2011; Herrmann et al., 2015). Recently, long-term NPF measurements have also been established in several new locations, e.g., in Beijing and Nanjing in China (Wang et al., 2017b; Kulmala et al., 2016;

10 Qi et al., 2015) and in Korea (Kim et al., 2014; Kim et al., 2013).

Currently, all the NPF studies published in the literature have utilized visual-based methods to identify NPF events from the measurement data. Typically, these methods require 1–3 researchers ~~analyzing the data~~ to ~~go through the data and look for~~analyze periods of formation and growth of new modes in the size-distribution data. These methods were first introduced

15 for analyzing data from the Finnish SMEAR stations by Dal Maso et al. (2005), and have been later slightly modified to suit analyzing data from different environments and measurement instruments (Hamed et al., 2007; Hirsikko et al., 2007; Vana et al., 2008). While the visual-based methods are in principle simple and straightforward to apply, there are certain drawbacks in using them. First, they are very labor intensive, since the analysis of the aerosol size-distribution data is not automated. Second, these methods are somewhat subjective, i.e. different researchers might interpret the same datasets in slightly different ways.

20 Finally, passing on the manual classification method from a researcher to a researcher could lead to an increasing systematic bias.

There have been attempts at improving ~~these~~NPF event classification methods and making them more automatic. In their comprehensive protocol article, Kulmala et al. (2012) introduced a concept for automatic detection of NPF events. This was

25 based on identifying regions of interest (ROI) from the time-series of measured aerosol size-distribution data. These ROI were defined as time periods when elevated concentrations of sub-20 nm particles were observed compared to the concentration of larger particles. Developing the NPF classification to take into account also other data measured at the same site, such as meteorological data and concentrations of trace gases, has allowed utilization of statistical methods to search for variables which could best explain and predict the occurrence of NPF. Hyvönen et al. (2005) and Mikkonen et al. (2006) applied

30 discriminant analysis for multi-year datasets of aerosol size-distributions and several gas and meteorological parameters measured at Hyytiälä, Finland and San Pietro Capofiume, Italy~~. Both of these studies were~~, respectively. Both of these studies were able to find the characteristic conditions for NPF event days in each site and it was seen that the conditions differ significantly. They were also able to construct models to predict the probability of NPF occurrence with reasonable accuracy, and this approach has also been used in the day-to-day planning of a complex airborne measurement campaign (Nieminen et

al., 2015). ~~Junninen et al. (2007) introduced an automatic algorithm based on self-organizing maps (SOM) and a decision tree to classify aerosol size distributions.~~ More recently, preliminary attempt at utilizing machine learning on big datasets have been reported (Zaidan et al., 2017).

5   There exists a long list of algorithms to automate the classification of different type of datasets (Duda et al., 2012), such as k-means, Support Vector Machine (SVM), Boltzmann Machine (BM), decision trees, etc. Our method choice is the deep feedforward Neural Network (NN). The idea of NN is not ~~young~~new, as the idea was first brought to life by Hebb in 1949 (Hebb, 2005). Since then, this field has drawn the attention of many researchers (Farley and Clark, 1954; Widrow and Hoff, 1960; Schmidhuber, 2015) probably because of its apparent simplicity and versatility. ~~The NNs are built by stacking~~
10 ~~interconnected layers of atomic units, or neurons, as depicted in Fig. 2~~In this study, we used a convolutional NN (CNN) because it mimics the visual cognition process of humans. One of the main bottlenecks in the use of the NNs is the learning stage of NN; NN must be trained before it can be used in image recognition or other applications. ~~In each layer, each neuron computes one very simple operation — often named activation — which consists of computing the weighted sum of the input variables plus a threshold and then passing the results to a function — e.g. tanh or Rectified Linear Unit (ReLU)(Buduma and~~
15 ~~Lacascio, 2017). To put it simply, a neuron emits a signal if the excitation signal — the input variable — reaches a threshold otherwise it is inactive. A single neuron cannot do much on its own, merely a linear discrimination, however, a set of connected and trained neurons, as a NN, can mimic human cognition abilities — the emergence property. The depth — or number of layers — and the topology of a NN will determine what it can be used for. For instance, the recurrent NNs (RNN) are well designed for speech, text and time series classification or prediction (Pascanu et al., 2013) and convolutional NNs (CNN) outperform~~
20 ~~any other architecture at classifying visual data, i.e. images or any 2D data (Krizhevsky et al., 2012). Before it can start classifying, a NN — of any kind — must be trained for the specific task it is assigned, e.g. recognize a car or a keyboard in a color image. While the topology is, for most cases, assigned a priori, the parameters — weights and thresholds — are to be learned. The learning stage is crucial because it determines the efficiency of the classifier and, therefore, the machine learning community has dedicated much effort to provide solutions of the issue — e.g. the backpropagation algorithm (Rumelhart et al.,~~
25 ~~1986) — and is still focusing on it decades after it all started (LeCun et al., 2015; Schmidhuber, 2015). There are as many learning methods as there are NN topologies, however they can be sorted in a few categories. For instance, the learning method may require a labeled training set, falling into the supervised learning (SL) category (Duda et al., 2012; Amari, 1998), or it may learn by itself without an already classified set, in which case the method is referred to as unsupervised learning (UL) (Le, 2013; Radford et al., 2016). Another relevant classification of the learning methods is the depth of the NN. Even though~~
30 ~~there is no common consensus to define the limit between shallow and deep NN (DNN), it is commonly accepted to refer to deep NN as NN having at least two layers and very deep those with more than ten layers (Schmidhuber, 2015).~~

~~In this study, we use a CNN because it mimics the visual cognition process of the human; Fig. 2c illustrates the structure of the whole CNN and Fig. 2d and 2e some layers of the CNN in detail. Instead of being fully connected, the CNN is only locally~~

connected, in other words, a neuron of a layer is connected to a compact subset of neurons of the previous layer. For every subset, the neurons compute the same operation – the parameters are shared across the neurons of a layer – resulting in a convoluted version of the input signal; hence the name. It is a good model for human vision because it applies the convolution operation throughout the image field – thus it has drawn a lot of attention in the image processing community (e.g. LeCun et

5   al., 1998; Cireşan et al., 2011; Chellapilla et al., 2006). However, CNN only became popular when the AlexNet outperformed by far all the other pattern recognition algorithms during the ImageNet Large Scale Visual Recognition Challenge, ILSVRC 2012 (Krizhevsky et al., 2012). Since then, it has been improved for each classification challenge, starting by the ZFnet (Zeiler and Fergus, 2014) then by the deeper GoogleNet (Szegedy et al., 2015) and finally by the deepest Microsoft ResNet (He et al., 2016).

10

One of the main bottlenecks with an artificial intelligence is the learning stage, especially for NN. For instance, thousands of images are typically needed for learning in image recognition applications. Therefore, the idea of reusing what is already known instead of re-learning from scratch every time a new class has come up. To overcome this problem, we used a pre-trained deep CNN, named AlexNet, which was originally trained to recognize different common items (e.g., pencils, cars and

15   different animals) and we transfer learned it to recognize images of different NPF event types. The transfer learning reduces significantly the number of images needed in the training process, from thousands to hundreds. The CNN used in this study and transfer learning of the CNN are described in detail in the Appendix A and Fig. 2.

The pre-training method consists in first training the network with an UL algorithm and then continue the training with SL (Bengio et al., 2006). Another solution is reinforcement learning, which is a technique involving an agent that learns policies

20   based on interaction with its environment using trial and error. For this method, the correct states are not known – therefore it is an UL – but a system of rewards gives hints whether the predictions are correct or not (Mnih et al., 2013; Sutton, 1988; Kaelbling et al., 1996; Schmidhuber, 2015). Finally, the transfer learning method, which we have used in this study, consists of learning the weights of an NN that contains a lot of classes (see "General" frame of Fig. 2b) and then using those weights either as a starting point for other learning sets with less classes or using partially the weights – e.g. for the first layers of the

25   NN – and train only one part in order to specialize the NN as it is depicted the frame "Specialization" of Fig. 2b. (Yosinski et al., 2014; Cireşan et al., 2012; Pan and Yang, 2010; Dauphin et al., 2012; Krizhevsky et al., 2012; Weiss et al., 2016). For all kinds of NNs, the learning process is prone to learning too much detail – missing the generality of a semantic class – related to the examples. This phenomenon is known as overfitting and several methods have been developed to overcome this such as dropout (Srivastava et al., 2014; Hinton et al., 2012).

30

In atmospheric science, several studies have used deep learning or other novel machine learning methods in data analysis. Deep learning and other machine learning algorithms are commonly used in remote sensing (Zhang et al., 2016; Lary et al., 2016; Han et al., 2017; Hu et al., 2015). Remote sensing is very suitable for machine learning because large datasets are available and the theoretical knowledge is incomplete (Lary et al., 2016). For instance, Han et al. (2017) introduced a modified

pre-trained AlexNet CNN and Hu et al. (2015) used several CNNs (e.g. AlexNet and VGGnets) for remote sensing image classification. Ma et al. (2015) used transfer learning in a Support Vector Machine (SVM) approach for classification of dust and clouds from satellite data. Other applications in atmospheric science include, e.g., air quality predictions (Li et al., 2016; Ong et al., 2016), characterization of aerosol particles with an ultra violet-laser induced fluorescence (UV-LIF) spectrometer

5 (Ruske et al., 2017) and aerosol retrievals from ground-based observations (Di Noia et al., 2015; Di Noia et al., 2017). In addition, we recently applied different machine learning methods (e.g. neural network and SVM) for aerosol optical depth (AOD) retrieval from sun photometer data (Huttunen et al., 2016). Although machine and deep learning approaches have already been used in several applications in atmospheric science, the use of those novel artificial intelligence methods will expand rapidly in the future. Nowadays, those methods are more efficient and easier to use due to the development of user-

10 friendly applications and increasing computing capacity (e.g. Graphics Processing Units, GPUs), and they can be applied in problems that are more complicated.

Here, we demonstrate that a novel deep learning-based method, i.e. transfer learning of a commonly used pre-trained CNN model (AlexNet), can be efficiently and accurately used in classifying of new particle formation (NPF) events. The method is

15 based on image recognition of daily-measured particle size distribution data. To our knowledge, this is the first time when ~~an automatic~~ a deep learning method, i.e. transfer learning of a deep CNN, has been successfully used in an automatic NPF event analysis for a long-term dataset. We will show that the deep learning-based method will increase the quality and reproducibility of event analysis compared to manual, human-made visual classification.

2 Materials and methods

20 ### 2.1 Measurement site, instrumentation and dataset

In this study, we analyzed a long-term particle size distribution (PSD) dataset measured at the San Pietro Capofiume (SPC) measurement station (44° 39' N, 11° 37' E, 11 m a.s.l.), Italy. The PSD measurements started on 24 March 2002 at SPC and have been uninterrupted, except for occasional system malfunctions. The SPC station is located in a rural area in Po Valley about 30 km northeast from the city of Bologna. The Po Valley area is the largest industrial, trading and agricultural area in

25 Italy with a high population density and hence it is one of the most polluted area in Europe. On the average at SPC, NPF events occur on ~~ca. 30~~36 % of the days whilst ~~ca. 30~~33 % are clearly non-event days and the probability for NPF events is highest in spring and summer seasons (Hamed et al., 2007; Nieminen et al., 2018).

At SPC, PSDs are measured with a twin Differential Mobility Particle Sizer (DMPS) system; the first DMPS measures PSDs

30 between 3 and 20 nm and the second one between 15 and 600 nm. The first DMPS consists of a 10.9 cm long Hauke-type differential mobility analyzer (DMA)(Winklmayr et al., 1991) and an ultrafine condensation particle counter (CPC, TSI model 3025) whereas the second DMPS consists of a 28 cm long Hauke-type DMA and a standard CPC (TSI model 3010). One

6

measurement cycle lasts for 10 min. The PSDs used in this study were calculated from the measured data using a Tichonov regularization method with a smoothness constraint (Voutilainen et al., 2001). The data from two different DMPS instruments was combined in the data inversion. The CPC counting efficiency and diffusional particle losses in the tubing and DMA systems were taken into account in the data analysis. In addition to PSD measurements, several gas and meteorological parameters are continuously measured at the SPC station (e.g. SO_2, NO, NO_2, NO_x, O_3, temperature, relative humidity, wind direction, wind speed, global radiation, precipitation, and atmospheric pressure). The measurement site and instrumentation have been described in detail in previous studies (Laaksonen et al., 2005; Hamed et al., 2007).

The analyzed dataset covers 4177 days (files) from the start of the measurement in SPC on 24 March ~~2012~~2002 until 16 May 2017 (totally 5534 days). The number of days at the different NPF classes and division into training and testing categories are summarized in Table 1.

2.2 Classification of new particle formation events (~~current~~traditional method)

Currently, a classification of NPF events is practically made manually, i.e., researchers visually inspect contour plots of time series of aerosol size distribution data and time evolution of nucleation-mode particles (particle dimeter below ca. 50 nm) (Kulmala et al., 2012). For the dataset of this study, the manual classification of NPF events into different categories is based on guidelines described by Mäkelä et al. (2000) and Hamed et al. (2007). Figure 1 shows examples of measured time series of PSDs (time in x-axis, particle diameter in y-axis and particle concentration presented by different colors) for different event classes.

In the first step of event analysis, data is classified into days with NPF events and days without particle formation (non-event days, NE). A day is considered as a NPF event day if formation of new aerosol particles starts in the nucleation mode size range (<25 nm) and subsequently grows, and the formation and growth is observed for several hours. The NPF event days are further classified according to the clarity and intensity of the events (Hamed et al., 2007):

- **Class 1 events** (Class_01) are characterized by high concentrations of 3–6 nm particles with only small fluctuations of the size distribution and no or little pre-existing particles in the smallest size ranges. Class 1 events show an intensive and clear formation of small particles with continuous growth of particles for seven to ten hours.
- **Class 2 events** (Class_02) show the same behavior as Class 1 but with less clarity. The formation of new particles and their subsequent growth to larger particle sizes can be clearly observed but, e.g., fluctuations in the size distribution are larger. Furthermore, the growth lasts for a shorter time than for Class 1 being about five hours on average. In event Classes 1 and 2, it is easy to follow the trend of the nucleation mode and, hence, the formation and growth rates of the formed particles can be determined confidently.
- **Class 3 events** (Class_03) include days when same evidence of new particle formation can be observed but growth is not clearly observed. For example, the formation of new particles and their growth to larger particle sizes may

7

occur for a short time but is then interrupted (e.g. by a drop in the intensity of solar radiation, rain). In addition, the days with weak growth are classified in that category

The classification of nucleation events is, however, subjective and overlapping or confusion within the classes may easily occur. To minimize the uncertainty of the classification method, Class 1 and Class 2 events are typically referred to as clear or

5 intensive nucleation events (Clear Event class), where all classification stages were clearly fulfilled, whilst Class 3 events are referred to as Weak Events. Baranizadeh et al. (2014) named Clear Event and Weak Event days as "Quantifiable" and "Non-Quantifiable" days, respectively, which describes better the difference of those classes. Both formation and growth rates can be determined for Quantifiable days but not both for Non-Quantifiable days.

10 Other classes of days are:

- **Non-Event days (NE)** (Class_NE): Days with no NPF in the nucleation mode particle size range are classified into Non-Event days. These days are also interesting because, e.g., differences in conditions (meteorology, gas concentration, precipitation) during event and non-event periods are often studied in order to get better understanding of processes behind NPF.

15 - **Class 0** (Class_00): Some of days do not fulfil the criteria to be classified either event or non-event day and they are classified as Class 0. In that class, it is difficult to determine whether a nucleation event has actually occurred or not.

- **Bad Data (BD)** (Class_BD): Days with some malfunction in the measurement system (e.g. too high or low particle concentrations, missing data in part of the day) are classified as Bad Data. In the final data analysis, those days are typically ignored.

20 **2.3 Event classification with a deep convolutional neural network (CNN)**

In this study, we developed a novel method to analyze NPF events automatically. The schematic of the approach is shown in Fig. 2. and described in detail in the Appendix A. We ~~use~~used a large, deep convolutional neural network (CNN) named AlexNet~~, which has originally been trained to classify images into 1000 different categories~~ (Krizhevsky et al., 2012, 2017)~~. The AlexNet's architecture, shown in Fig 2c, consists of five convolutional layers (CL, Fig. 2d), some of which are followed~~

25 ~~by max-pooling (MP) layers, and three fully connected layers (DNN, Fig. 2e) with a final 1000-way softmax, which assigns a probability ($P(\ell)$) to each of the classes' label (ℓ) for the input image.~~, which has originally been trained on millions of images, as a subset of the ImageNet database (Deng et al., 2009), to classify images into 1000 different categories. ~~The first five layers of the AlexNet extract abstract features that are easier to classify than the original input image. This abstraction is computed by chaining up CL and MP layers. The CL (depicted in Fig. 2d) computes several convolution of the image by several kernels~~

30 ~~– that must be learnt – and generates a multidimensional output – one dimension per kernel. For instance, in Fig. 2d the two kernels v_1 and v_2 are applied to two compact sub-regions of the data (field 1 and field 2) by two units (unit 1 and unit 2), which generate two 2D outputs ($y_1 = (x_1^1, x_1^2)$ and $y_2 = (x_2^1, x_2^2)$). The CL is what makes the CNN similar to the human visualization system because it applies the same transformation throughout the field of view. The MP layer merely computes a smaller~~

8

~~image from each output image of the CL, e.g., an image of 256 x 256 pixels generated by CL will be reduced to smaller image of 64 x 64 by calculating maximum values of the 4 x 4 patches of the original image. The last layers of the AlexNet is a DNN as depicted in Fig. 2c , i.e. a NN whose input and output layers are connected by hidden layers of neurons. Contrary to the CL, each neuron of a DNN's layer is connected to all the neurons of the previous layer and the parameters (weights and threshold) are not shared across the layer; this is the most general setting for a feedforward NN. In the AlexNet, the output DNN is the decision making center, which would be the analog of the brain in the visual system of a human. To speed up the learning process and reduce overfitting in the fully connected layers (DNN), the batch normalization, the dropout regularization and the Rectified Linear Unit (ReLU, $\max(0, x)$) methods are employed in the AlexNet.~~ The AlexNet was a breakthrough method in image analysis when it was introduced in 2012 and thus it is widely used in image recognition applications (LeCun et al., 2015). ~~This model has originally been trained on millions of images of very common objects, e.g., keyboards, mice, pencils, cars and many animals, as a subset of the ImageNet database (Deng et al., 2009) and thus it must be fine tuned to recognize other images.~~ Since the model has originally been trained to recognize images of very common objects, e.g., keyboards, mice, pencils, cars and many animals, it has to be fine-tuned to recognize other images.

The AlexNet itself cannot recognize NPF events so we used a transfer-learning technique for fine-tuning the model for PSD images (daily contour plots of time series of aerosol size distributions). In the fine-tuning of CNN by transfer-learning, the new features are quickly learned by modifying a few of the last layers using a much smaller number of images than in the training of the original CNN (Weiss et al., 2016; Shin et al., 2016; Yosinski et al., 2014; Pan and Yang, 2010). ~~The transfer learning process (see Fig.~~This strategy is very efficient for images because of the structure of the CNN; the first layers process the image by extracting some features so that it becomes more abstract and the ~~2b) is a two-stage optimization method. At first, the network is trained for a learning dataset (S_g) including the data itself (x_k) and its labels (d_k). This is achieved by solving an optimization problem for all the NN's parameters that are represented by the vector sets V and W, the first layers and the last layers, respectively. The optimization problem is defined so that the cost function $\mathcal{E}$ reaches a minimum for optimal sets of parameters $\hat{V}$ and $\hat{W}$. Once the optimal parameters are known, for almost all the elements $x_k \in S_g$, the first output of the NN predicts the actual label $d_k \simeq J(x_k | \hat{V}, \hat{W})$, which means that the NN is ready to classify data of the same semantic field as this of the learning set. The second stage of transfer learning consists of optimizing only a few layers amongst the lasts – or possibly all the layers – using as an initial guess the optimal sets of the general learning problem. For instance, the cost function $\mathcal{E}$ in Fig 2b may be the same for the general and specialized learning step, but for the first stage, all the parameters are allow to vary while only the W may vary during the specialization. This strategy is very efficient for images because of the structure of the CNN; the first layers process the image by extracting some features so that it becomes more abstract and the~~ lasts layers attribute the probabilities of the classes. In our setup, we modified only a few of the last layers because of the low number of available training data ~~S_s~~.

In a transfer learning process of our data set, we categorized PSD images into six different classes based on the manual event classification (Class 1, Class 2, Class 3, NE, Class 0 and BD, see Fig. 1) and used a subset of images for training and the rest of the images for evaluation of the training (testing). Three different fractions of images were used in the training stage (80 %, 50 % and 20~~%)~~%, see Table 1) in order to study the effect of image set size on the NPF event classification. Images for training

5   and testing were selected randomly (certain percent of each category) and this was repeated ten times to evaluate the statistical accuracy of the training and data classification. The transfer learning process and data analysis were computed using a Matlab program (version R2016b) with the support of the package of Neural Network Toolbox Model for AlexNet Network (version 17.2.0.0) using a Linux server (CPU: 2x Intel Xeon E5-2630 v3, 2.40 GHz,16 cores; RAM 264.0 GB; GPU: Nvidia Tesla K40c, CUDA ver. 3.5, 12 GB). We used the standard procedures (trainNetwork, classify) and options (solver: sgdm, initial

10   learn rate: 0.001, max Epochs: 20, mini-batch size: 64) as described in an example for deep learning by Matlab (MathWorks, 2017). In the transfer learning of Matlab AlexNet, we changed two layers of the net to fit our data: the number of recognized classes was reduced to six in the last fully connected layer (fc8) and the category's names in the classification output layer (output) were charged to the names of our event classes. One training session lasted about from a half to one hour with the used server and programs. The AlexNet CNN has been originally implemented in CUDA, but it is also available in Matlab.

15   Other programing languages provide also good libraries for implementing the AlexNet such as Python or C++, and dedicated frameworks such as TensoFlow, Theano or Caffe are good candidates for an easy implementation of the NN and the learning stage.

20   To find out the best performance for the NPF event classification, we ~~also~~ tested three different sets of particle size distribution images (Image Set, jpg file format with bit depth of 24 bits) with different plotting areas of the daily-measured size distributions:

> **Image Set 1.** Original image (see Fig. 1) without the title, axis labels and numbers (i.e. axes of the plot and data inside them), size 1801 x 951 pixels (file size ca. 180 kB).

25   **Image Set 2.** Only colored, measured part of original images (i.e. time 00:00 – 24:00, particle diameter 3 – 630 nm), size 1646 x 751 pixels (file size ca. 120 kB).

> **Image Set 3.** As previous but only an active time for NPF events is considered (i.e. 06:00 – 18:00, diameter 3 – 630 nm), size 831 x 746 pixels (file size ca. 60 kB).

Image sets with the different size were tested because all images have to resize to 227 x 227 pixels, which is an input image

30   size of the AlexNet. Resizing of images pixels was conducted by a Matlab imresize -function using default bicubic interpolation (MathWorks, 2018). Resizing of images can lost some information of measured PSDs needed in the NPF identification.

After transfer learning each dataset (three image sets with three different training-testing fractions of images, repeated 10 times), the accuracy and training success were evaluated. We calculated the average success rate (accuracy) of the transfer

learned CNNs by comparing CNN-based and human-made classifications of the test images. We also combined some classes together in the result analysis since overlapping of the classes could have easily occurred in the classification, e.g., Class 1 and 2 were combined to a Clear Event class (Cl_1-2).

2.4 Statistical methods

5   The performance of the classification between different image sets and training rates in different event classes was compared with the ~~Kruskall~~Kruskal-Wallis test and with multivariate analysis of variance (ANOVA). The ANOVA was conducted with a robust fit function ~~(rlm; Venables and Ripley, 2002; Huber, 1981)~~(rlm; Venables and Ripley, 2002; Huber, 1981) because the normality and homoscedasticity assumptions of ordinary least squares method were not completely fulfilled. All statistical analyses were performed in R-software (R Core Team, 2017).

10 **3 Results and discussion**

The success of the transfer learning process of CNN was evaluated by confusion matrices, which showed a classification accuracy of the CNN method over visual inspection, i.e., how a certain class classified by human (visually) were categorized in the CNN classification. An example of the confusion matrices for one training-testing run (Image Set 1, training-testing ratio of 50 %/50 %) is presented in Fig. 3. When the training data set was analyzed with the trained CNN (same data), the

15 overall classification accuracy (i.e. a fraction of days with an equal classification) was ca. 98 % and accuracies at certain event classes varied between 94-100 % (see Fig. 3a). For all cases (all image sets and training ratios), mean accuracies of ten different runs varied from 93 % to 98 %. The results show that the method can easily classified training data sets with very high accuracy, indicating that the training process of CNN was very successful.

20 Table 2 shows a summary of the classification accuracy for all cases when the method was applied for testing data sets (different days). The overall accuracy (all classes) is about 63 % for all studied cases. If we consider individual classes in more detail, the highest accuracies are in classes of NE and BD (ca. 80 %) whereas the lowest accuracies are in Classes 2, 3 and 0 (ca. 45 %), followed by Class 1 (ca. 53 %). The highest accuracies in NE and BD classes are apparently due to easier classification compared to other classes, e.g., no particles at low particles size ranges, no intensive particle growth or absence

25 of particle at all and unusual high particles concentration in a part of data (see Fig. 1). The classification of other days is more challenging because differences between classes are not so distinct and the choice between classes can be difficult. When the clear event classes (Class 1 and 2) are combined into one category (Clear Event class, Cl_01-02), the classification accuracy increased to ca. 75 %. Overall, a classification accuracy for Clear Events and Non-Event categories is ca. 75-80 %, which can be considered a very good first result for automatic NPF event classification. In general, the classification would be more

30 consistent, by both human and CNN, if only two different classes are used for event days: Clear and Weak events or "Quantifiable" and "Non-Quantifiable" events as described by Baranizadeh et al. (2014). From "Quantifiable" days, it is

possible to quantify basic parameters of NPF event, e.g., particle formation and growth rates. Thus, we recommend that NPF event days should be categorized into classes of "Quantifiable" (Q) and "Non-Quantifiable" (NQ) events, in addition to Non-Event (NE), Undefined (UD or Class 0) and Bad-Data (BD) classes in the future analysis. This would also increase the number of images of the event classes in the training of CNNs.

5

If we look at the results in more detail, we can see variation in results between different training fractions (number of days in training in different event classes are shown in Table 1). In the cases of the training fractions of 80 % and 50 %, the classification accuracy values are quite similar but the accuracy decreases when the training fraction is only 20 % (especially for clear event days). Statistical analysis between different training-testing ratios shows that the 50 % and 80 % training rates

10 are equally precise in all comparisons. This means that we could get adequate classification performance already with 50% training fraction (i.e., 135-708 images per class). However, when the size of the training set was lowered to 20 % (54-283 images per class), the classification became more uncertain in many comparisons. At 20 %, statistically significant differences were found when all classes were analyzed together and with classes Cl-01-02, Class 00, Class 02 and Class BD, indicating that the number of training days was too low for precise classification (e.g. Class 1 had only 54 days for training). In summary,

15 a training fraction of 50 % (minimum 135 images per class) is a good compromise between accuracy, reliability and number of training days for the used data set.

We also studied the effect of image size (Image Sets 1-3) for classification accuracy. Only when all event classes were inspected together, there was a statistically significant difference between the image sets. Image Set 3 had a slightly lower

20 performance rate than the two other sets. When limiting the classes to smaller sub groups the differences were not statistically significant anymore. The result indicates that image sets including all daily-measured size distribution data (Image Sets 1 and 2) are more suitable for CNN analysis than Image Set 3 with a reduced analysis period (6:00-18:00). Although images have to resize to the fixed input size (227 x 227 pixels) for CNN analysis, there is no need to reduce the analyzed period to cover only the active time for NPF. In fact, it is better to use all daily measured data in the analysis although resizing of images might

25 lose a part of the information.

As described earlier, a choice between different event classes is somewhat arbitrary and not an easy task even for trained researchers and thus overlapping between different classes may occur. To analyze this overlapping in more detail, we plotted how a manually classified class is distributed into different classes by CNN classification. Figure 4 shows an example of CNN

30 classification distributions for Image Set 1 with training-testing ratio of 80 %/20 %. Similar classification distribution into different classes can also be observed from the confusion matrix for Image Set 1 with training-testing ratio of 50 %/50 % (only one computing run) in Fig. 3b. The results show that an overlapping occurs typically between the adjacent or similar type of classes. For instance, a manually classified Class 1 is categorized mainly into Classes 1 and 2, Class 2 into Classes 1, 2 and 3,

etc. The minimum overlapping is for classes NE and BD, which are the easiest classes to categorize by researchers. Similar overlapping is also observed in other analyzed cases (image sets and training-testing ratios).

To study overlapping or misclassification between different classes, we look in more detail at cases when Clear Event days (Classes 1 and 2) by human-made classification are categorized to Non-Event days by CNN-based classification and vice versa. Figure 5 shows examples of those days; the left-hand plots are categorized to Clear events by human and right-hand plots by CNN, the first row is examples of human made misclassifications, the second row CNN misclassifications and the third row difficult situations for classification. For instance, Fig. 5f shows a case when the initial stage of NPF has not been observed (probably due to change of wind or a mixing of boundary layer) but clear growth of particles is observed later. In that case, CNN-based classification does not "recognize" the missing of initial particle formation in the smallest particle sizes (ca. 3-4 nm) and therefore the day is classified as a Clear event day (Class 1 and 2). An opposite situation is shown in Fig. 5e where a Class 1 day (classified by human) was categorized two times into Non-Event and once into Class 2 day by CNN in different computing runs. In that day, NPF is clear but the concentration of formed particles is lower than for typical event days and this probably affects the accuracy of the CNN-based classification. A similar misclassification can be seen in Fig 5c. In contrast, Figures 5a and 5b show examples of misclassifications, which are due to clear human errors, probably the researcher has just written down a wrong event class number. A general overview is that clear misclassifications seem to occur more commonly in human-made analysis than CNN-based categorization, which indicates that the developed CNN-based method has a better reliability and repeatability than manual human-made classification.

Only a few reports on automatic data analysis of NPF events have been published so far. In very recent conference proceedings, Zaidan et al. (2017) introduced a machine learning method based on a neural network. Their method utilized particle size distribution data pre-processing (fitting of lognormal distributions to the data), feature calculations and extractions (e.g. mean size, standard deviation, etc.) and principal component analysis (PCA) before the final classification by the neural network. Their preliminary results show a classification accuracy of ca. 83 % for Hyytiälä (Finland) data in 1996-2014 when only event and non-event days were considered. When compared to our method, they used several pre-processing steps before classification by the neural network and they did not use pre-trained CNNs or image recognition in their classifier. Kulmala et al. (2012) introduced a procedure for automatic detection of regional new particle formation. The procedure is based on monitoring the evolution of particle size distributions and it includes several steps, e.g., data noise cleaning and smoothing, excluding data larger than 20 nm and calculating regions of high particle concentration in particle size distributions. In the final step, the method automatically recognizes event regions from size distributions (i.e. regions of interest, ROI) and determines, e.g., NFP event start times as well as particle formation and growth rates. This method is very straightforward and does not include any data analysis methods based on artificial intelligence. To our knowledge, this method has not been used routinely in NPF event analysis for large datasets. Junninen et al. (2007) used an automatic algorithm based on self-organizing maps (SOM) and a decision tree to classify NPF events in Hyytiälä for 11-year data. They taught five SOMs by tuning them

13

for specific event types and used a decision tree to get the probability for the day to belong to three different event classes (Event, Non-event and Undefined). The overall accuracy of the method was ca. 80 %. When comparing to our method, they pre-processed the data by decreasing the size resolution to 15 bins with variable width size bins in log-scale (higher resolution in smaller particles) and time resolution to 24 steps (1 h average). Furthermore, they weighted small particles (diameter < 20 nm) by a factor of 10 in one of the SOMs. To our knowledge, this method has not been used routinely in NPF event analysis. Hyvönen et al. (2005) used data mining techniques to analyze aerosol formation in Hyytiälä, Finland and Mikkonen et al. applied similar methods to datasets recorded in SPC (2006) and in Melpitz and Hohenpeissenberg, Germany (2011). They studied different variables and parameters that may be behind NPF but they did not make automatic classification for NPF events. Furthermore, Vuollekoski et al. (2012) introduced an idea on an eigenvector-based approach for automatic NPF event classification but they did not report any results because the method was still under development and its performance was uncertain.

We have not yet tested the method at other sites. Basically, "banana type" events, non-event days and bad data should be recognized from other site data if figures are plotted roughly in a similar way (one-day plot, same size ranges and axes, and color map). The method analyses features from size distribution plots, which are quite similar in many cases in different sites. However, we suggest that the CNN should be transfer learned again for new sites in order to get best results, especially if the shapes of size distributions are different compared to those of training data, e.g., low tide events in coastal sites (O'Dowd et al., 2010; Vaattovaara et al., 2006) or rush hour episodes in urban environments (Jeong et al., 2004; Alam et al., 2003). We encourage researchers to use the method in other sites and report results in order to develop the method. The accuracy of classification could be improved, e.g., by tuning training parameters, optimizing a number of classes used in analysis (e.g. merging Classes 1 and 2) and using synthetic training data.

4 Conclusions

We have developed a novel method based on deep learning to analyze new particle formation (NPF) events. The method utilizes a commonly available pre-trained Convolutional Neural Network (AlexNet) that has been trained by transfer learning to classify particle size distribution images. To our knowledge, this is the first time when NPF events have been a deep leaning method, i.e. transfer learning of a deep Convolutional Neural Networks (CNN), has successfully classifiedbeen used to classify automatically NPF events into different classes, including several event and non-event classes, directly from particle size distribution data.images, as the researchers do in a typical manual classification. Although there are general guidelines for human-made NPF event classification, the classification is always subjective and, therefore, it can vary between researchers or even within one researcher. In many ambiguous cases, it is not easy to attribute an event to the "correct" event class, even for an experienced researcher. The quality of the classification can especially vary for long-term data sets, which have been analyzed by several researcher in different times. Furthermore, a wrong event class can be listed to database due to a human

error, which reduces reliability of the classification. Therefore, an automatic method, which can manage the whole dataset at once with a high reproducibility, is desired for NPF event analysis.

Our results show that transfer learning of a pre-trained CNN to recognize images of particle size distributions is a very powerful

5 tool to analyze NPF events. The event classification can be done directly from existing data (figures) without any pre-processing of the data. Although an average classification accuracy of certain classes is ca. 65 %, the overall accuracy is ca. 80 % for Non-Event (NE) and Clear Event classes (Classes 1 and 2 combined), which is a good first result for automatic NPF event analysis. The most of miss-classified days have been categorized into the adjacent classes, which can be ambiguous to distinguish from each other. A compassion between CNN-based and human-made classification also showed that often the

10 difference in categorization is due to a wrong or an incorrectly listed classification by researcher. Human-made classification can easily vary by people to people and can change over time whereas CNN-based classification is consistent all times. The CNN based categorization seems, at least, to be as reliable as human-made and it could be even more reliable if training image sets are selected carefully. Typically, an analysis of large size distribution dataset requires manual labor and training for several researchers, which is very time consuming and quality of analysis may vary. The developed automatic CNN-based NPF event

15 analysis can be used to study long-term effects of climate change on NPF in more efficiently, accurately and consistently, which helps us to understand in detail aerosol-climate interactions.

The transfer learning of pre-trained CNNs (like Alexnet and GooLeNet) allows us to make automatic event classification systems effectively without long-lasting design, training and computing of CNNs. Typically, a training of a CNN needs from

20 thousands up to a million images but in the transfer learning of a pre-trained CNN, a hundred images can be enough for a precise classification. The pre-trained CNNs, as well as other novel machine learning and artificial intelligence methods, and the increased computing capacity due to Graphics Processing Units (GPU) enable us to analyze very complex and large datasets as is typically in atmospheric science.

25 Instead of the transfer learning of pre-trained CNN, several other artificial intelligence methods could also be utilized in NPF analysis. The Recurrent Neural Network (RNN) is a good candidate for the classification and predication of time series (Pascanu et al., 2013), however, not viewed as surface plots but as a sequence of particle size distributions. This allows more variability in the NPF classes, e.g. a continuum of NPF event intensity, and determining time dependent processes like particle formation and growth. Furthermore, the underlying weights of NNs would give insights about the size distribution evolution

30 and this could be potentially used as an evolution model by substituting or complementing the general dynamic equation (GDE). In addition, the use of unsupervised learning method can help us describing new NPF event classes or merge classes based on humanly imperceptible features. Reinforcement learning ~~(Kaelbling et al., 1996)~~(Kaelbling et al., 1996) is an interesting technique especially for the cases for which the classification is not well defined, e.g. if the classification varies from one human to another, because it only requires a rewarding system or policy. The rewards need to be positive if the

prediction is satisfying and negative if not; hence, the reward values can be the difference between the number of person that classified in the predicted class and those who did not. However, it should be mentioned that this technique works best for prediction models, such as RNN. Currently, the state-of-the-art in artificial intelligence are deep CNN based algorithms *AlphaGo Zero* and *AlphaZero* developed by DeepMind (Silver et al., 2017b; Silver et al., 2017a). Those algorithms achieved

5 superhuman performance by *tabula rasa* reinforcement learning without human data or guidance and defeated human in the game of Go and the most dedicated program Stockfish in chess.

In summary, we encourage researchers to use the CNN-based NPF identification method to their own data because of the better reliability and repeatability compared to human-made classification. However, there are still some weaknesses that should be

10 kept in mind, e.g., quality and quantity of data is crucial in the training process, supervised learning is needed, the method still needs quite a lot of computing power (GPU), the identification is not perfect and particle formation and growth rates cannot be determined using the current model. Furthermore, the CNN should likely be transfer learned again for new sites in order to the get best results, especially if the size distribution evolution is very different compared to that of the site of the training data. We suggest that in training, ca. 150 days per class should be enough to get reasonable classification. Alternatively, simulated

15 data could be used for training (Lähivaara et al., 2018) but we have not tested how well it works in practice. The method is, however, very easy to use and results are accurate and consistent, especially for long-term data, if the CNN has been trained carefully with high quality and reasonable amount of data. Finally, experiences and data obtained from other sites can be used in the further development of the method, e.g. to find suitable learning parameters, more data for training and possibility for a cross-validation of the method.

20

In the near future, we will analyze long-term changes in NPF in San Pietro Capofiume and utilize the method for other field stations (e.g. Puijo, Kuopio, Finland). We are including more parameters into automatized NPF analysis (e.g. particle formation and growth rates, event start and end times) and are developing methods to predict NPF events based on meteorological and other atmospheric data. In addition, the simulation based deep learning is a potential research topic in the future (~~see, e.g.,~~

25 Lähivaara et al., ~~2017~~2018).

5 Appendix A

General description of neural networks and training process

We used a deep feedforward Neural Network (NN) to analyze NPF events in this study. Figure 2 shows a schematic of the

30 used method and detailed descriptions of the method can be found in the textbooks of the subject (e.g. Buduma and Locascio, 2017; Duda et al., 2012). In general, the NNs are built by stacking interconnected layers of atomic units, or neurons. In each layer, each neuron computes one very simple operation – often named activation – which consists of computing the weighted

16

sum of the input variables plus a threshold and then passing the results to a function – e.g. tanh -function or Rectified Linear Unit (ReLU, i.e. $\max(0, x)$ -function)(Buduma and Locascio, 2017). To put it simply, a neuron emits a signal if the excitation signal – the input variable – reaches a threshold otherwise it is inactive. A single neuron cannot do much on its own, merely a linear discrimination, however, a set of connected and trained neurons, as a NN, can mimic human cognition abilities – the emergence property. The depth – or number of layers – and the topology of a NN will determine what it can be used for. For instance, the recurrent NNs (RNN) are well designed for speech, text and time series classification or prediction (Pascanu et al., 2013) and convolutional NNs (CNN) outperform any other architecture at classifying visual data, i.e. images (Krizhevsky et al., 2012).

Before it can start classifying, a NN – of any kind – must be trained for the specific task, e.g. to recognize a car or a keyboard in an image. While the topology is, for most cases, assigned a priori, the parameters – weights and thresholds – are to be learned. The learning stage is crucial because it determines the efficiency of the classifier and, therefore, the machine learning community has dedicated much effort to provide solutions for the issue – e.g. the backpropagation algorithm (Rumelhart et al., 1986) – and is still focusing on it decades after it all started (LeCun et al., 2015; Schmidhuber, 2015). There are as many learning methods as there are NN topologies, however they can be sorted in a few categories. For instance, the learning method may require a labeled training set, falling into the supervised learning (SL) category (Duda et al., 2012; Amari, 1998), or it may learn by itself without an already classified set, in which case the method is referred to as unsupervised learning (UL) (Le, 2013; Radford et al., 2016). Another relevant classification of the learning methods is the depth of the NN. Even though there is no common consensus to define the limit between shallow and deep NN (DNN), it is commonly accepted to refer to deep NN as NN having at least two layers and very deep those with more than ten layers (Schmidhuber, 2015).

The learning stage is one of the main bottlenecks in an artificial intelligence, especially for NN. For instance, thousands of images are typically needed for learning in image recognition applications. Therefore, instead of relearning the complete structure if a new class is introduced or merged with another one, a less expensive strategy has emerged, namely transfer learning; it consists in using part of the existing learned parameters and learning only a subset of the structure. The pre-training method consists in first training the network with an UL algorithm and then continue the training with SL (Bengio et al., 2006). Another solution is reinforcement learning, which is a technique involving an agent that learns policies based on interaction with its environment using trial-and-error. For this method, the correct states are not known – therefore it is an UL – but a system of rewards gives hints whether the predictions are correct or not (Mnih et al., 2013; Sutton, 1988; Kaelbling et al., 1996; Schmidhuber, 2015). Finally, the transfer learning method, which we have used in this study, consists of learning the weights of an NN that contains a lot of classes (see "General" frame of Fig. 2b) and then using those weights either as a starting point for other learning sets with less classes or using partially the weights – e.g. for the first layers of the NN – and train only one part in order to specialize the NN as it is depicted the frame "Specialization" of Fig. 2b. (Yosinski et al., 2014; Cireşan et al., 2012; Pan and Yang, 2010; Mesnil et al., 2012; Krizhevsky et al., 2012; Weiss et al., 2016). For all kinds of NNs, the

17

learning process is prone to learning too much detail – missing the generality of a semantic class – related to the examples. This phenomenon is known as overfitting and several methods have been developed to overcome this such as dropout (i.e. ignoring randomly some neurons by setting them to zero during training)(Srivastava et al., 2014; Hinton et al., 2012).

5   *Convolutional neural network AlexNet and its transfer learning*

In this study, we used a CNN because it mimics the visual cognition process of the human. Figure 2c illustrates the structure of the whole CNN and Fig. 2d and 2e some layers of the CNN in detail. Instead of being fully connected, the CNN is only locally connected, in other words, a neuron of a layer is connected to a compact subset of neurons of the previous layer. For

10   every subset, the neurons compute the same operation – the parameters are shared across the neurons of a layer – resulting in a convoluted version of the input signal; hence the name. It is a good model for human vision because it applies the convolution operation throughout the image field – thus it has drawn a lot of attention in the image processing community (e.g. LeCun et al., 1998; Cireşan et al., 2011; Chellapilla et al., 2006). CNN became popular when the large, deep CNN named AlexNet outperformed by far all the other pattern recognition algorithms during the ImageNet Large Scale Visual Recognition

15   Challenge, ILSVRC 2012 (Krizhevsky et al., 2012). Since then, CNN has been improved for each classification challenge, starting by the ZFnet (Zeiler and Fergus, 2014) then by the deeper GoogleNet (Szegedy et al., 2015) and finally by the deepest Microsoft ResNet (He et al., 2016).

AlexNet was used to identify NPF events in this study. The architecture of the AlexNet, shown in Fig 2c, consists of five

20   convolutional layers (CL, Fig. 2d), some of which are followed by max-pooling (MP) layers (a down sampling layer that extracts the maximum values of predefined subregion) and three fully connected layers (DNN, Fig. 2e) with a final 1000-way softmax (an output layer), which assigns a probability ($P(\ell)$) to each of the classes' label (ℓ) for the input image (using a normalized exponential function, softmax). The first five layers of the AlexNet extract abstract features that are easier to classify than the original input image. This abstraction is computed by chaining up CL and MP layers. The CL (depicted in

25   Fig. 2d) computes several convolution of the image by several kernels (i.e. matrices used in the convolution) – that must be learnt – and generates a multidimensional output – one dimension per kernel. For instance, in Fig. 2d the two kernels v_1 and v_2 are applied to two compact sub regions of the data (*field 1* and *field 2*) by two units (*unit 1* and *unit 2*), which generate two 2D outputs ($y_1 = (x_1^1, x_1^2)$ and $y_2 = (x_2^1, x_2^2)$). The CL is what makes the CNN similar to the human visualization system because it applies the same transformation throughout the field of view – making the classification shift invariant. The MP

30   layer merely computes a smaller image from each output image of the CL, e.g., an image of 256 x 256 pixels generated by CL will be reduced to smaller image of 64 x 64 by calculating maximum values of the 4 x 4 patches of the original image. The last layers of the AlexNet is a DNN as depicted in Fig. 2e , i.e. a NN whose input and output layers are connected by hidden layers of neurons. Contrary to the CL, each neuron of a DNN's layer is connected to all the neurons of the previous layer and the parameters (weights and threshold) are not shared across the layer; this is the most general setting for a feedforward NN.

18

In the AlexNet, the output DNN is the decision making center, which would be the analog to the abstraction stage created by the brain as the output of the visual system of a human. To speed up the learning process and reduce overfitting in the fully connected layers (DNN), batch normalization and dropout regularization are employed in the AlexNet as well as the activation function ReLU.

5

The transfer learning process of the AlexNet (see Fig. 2b) is a two-stage optimization method. At first, the network is trained for a learning dataset (S_g) composed of the data itself (x_k) and its labels (d_k). This is achieved by solving an optimization problem for all the NN's parameters that are represented by the vector sets V and W, the first layers and the last layers, respectively. The optimization problem is defined so that the cost function $\mathcal{E}$ reaches a minimum for some optimal sets of

10 parameters $\hat{V}$ and $\hat{W}$. Once the optimal parameters are known, for the most of elements $x_k \in S_g$, the output of the NN – $J(x_k|\hat{V}, \hat{W})$ – predicts the actual label d_k, which means that the NN is ready to classify data of the same semantic field as this of the learning set. The second stage of transfer learning consists of optimizing only a few layers amongst the last ones – or possibly all the layers – using as an initial guess the optimal sets of the general learning problem. For instance, the cost function $\mathcal{E}$ in Fig 2b may be the same for the general and specialized learning step, but for the first stage, all the parameters are allow

15 to vary while only the W may vary during the specialization. This strategy is very efficient for images because of the structure of the CNN; the first layers process the image by extracting some features so that it becomes more abstract and the last layers attribute the probabilities of belonging to the classes. In our setup, we modified only a few of the last layers because of the low number of available training data S_s, and used original weights and thresholds in the other layers.

20

References

Alam, A., Shi Ji, P., and Harrison Roy, M.: Observations of new particle formation in urban air, J. Geophys. Res. Atmos., 108, 10.1029/2001JD001417, 2003.

Albrecht, B. A.: Aerosols, Cloud Microphysics, and Fractional Cloudiness, Science, 245, 1227-1230,

30 10.1126/science.245.4923.1227, 1989.

Amari, S.-i.: Natural Gradient Works Efficiently in Learning, Neural Comput., 10, 251-276, 10.1162/089976698300017746, 1998.

Andreae, M. O., and Rosenfeld, D.: Aerosol–cloud–precipitation interactions. Part 1. The nature and sources of cloud-active aerosols, Earth-Sci. Rev., 89, 13-41, 10.1016/j.earscirev.2008.03.001, 2008.

Asmi, E., Kivekäs, N., Kerminen, V. M., Komppula, M., Hyvärinen, A. P., Hatakka, J., Viisanen, Y., and Lihavainen, H.: Secondary new particle formation in Northern Finland Pallas site between the years 2000 and 2010, Atmos. Chem. Phys., 11, 12959-12972, 10.5194/acp-11-12959-2011, 2011.

Baranizadeh, E., Arola, A., Hamed, A., Nieminen, T., Mikkonen, S., Virtanen, A., Kulmala, M., Lehtinen, K., and Laaksonen, A.: The effect of cloudiness on new-particle formation: investigation of radiation levels, Boreal Environ. Res., 19, 343-354, 2014.

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H.: Greedy layer-wise training of deep networks, Proceedings of the 19th International Conference on Neural Information Processing Systems, Canada, 153-160, 2006.

Brunekreef, B., and Holgate, S. T.: Air pollution and health, Lancet, 360, 1233-1242, 10.1016/S0140-6736(02)11274-8, 2002.

Buduma, N., and ~~Lacascio~~Locascio, N.: Fundamentals of deep learning: Designing next-generation machine intelligence algorithms, O'Reilly Media, Sebastopol, CA, USA, 298 pp., 2017.

Chellapilla, K., Puri, S., and Simard., P.: High performance convolutional neural networks for document processing, Tenth International Workshop on Frontiers in Handwriting Recognition, La Baule (France), Suvisoft, <inria-00112631>, 2006.

Cireşan, D. C., Meier, U., Masci, J., Gambardella, L. M., and Schmidhuber, J.: Flexible, high performance convolutional neural networks for image classification, Proceedings of the Twenty-Second international joint conference on artificial Intelligence, Barcelona, Catalonia, Spain, 1237-1242, 10.5591/978-1-57735-516-8/ijcai11-210, 2011.

Cireşan, D. C., Meier, U., and Schmidhuber, J.: Transfer learning for Latin and Chinese characters with Deep Neural Networks, The 2012 International Joint Conference on Neural Networks (IJCNN), 2012, 1-6,

Dal Maso, M., Kulmala, M., Riipinen, I., Wagner, R., Hussein, T., Aalto, P. P., and Lehtinen, K. E. J.: Formation and growth of fresh atmospheric aerosols: Eight years of aerosol size distribution data from SMEAR II, Hyytiälä, Finland, Boreal Environ. Res., 10, 323-336, 2005.

~~Dauphin, G. M. Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I., Lavoie, E., Muller, X., Desjardins, G., Warde-Farley, D., Vincent, P., Courville, A., and Bergstra, J.: Unsupervised and Transfer Learning Challenge: a Deep Learning Approach, Proceedings of ICML Workshop on Unsupervised and Transfer Learning, Proceedings of Machine Learning Research, 97-110, 2012.~~

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 248-255, 10.1109/CVPR.2009.5206848, 2009.

Di Noia, A., Hasekamp, O. P., van Harten, G., Rietjens, J. H. H., Smit, J. M., Snik, F., Henzing, J. S., de Boer, J., Keller, C. U., and Volten, H.: Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations, Atmos. Meas. Tech., 8, 281-299, 10.5194/amt-8-281-2015, 2015.

Di Noia, A., Hasekamp, O. P., Wu, L., van Diedenhoven, B., Cairns, B., and Yorks, J. E.: Combined neural network/Phillips-Tikhonov approach to aerosol retrievals over land from the NASA Research Scanning Polarimeter, Atmos. Meas. Tech., 2017, 4235–4252, 10.5194/amt-10-4235-2017, 2017.

Duda, R., Hart, P., and Stork, D.: Pattern classification, 2nd edition, John Wiley & Sons, Inc., 2012.

5 Farley, B., and Clark, W.: Simulation of self-organizing systems by digital computer, Transactions of the IRE Professional Group on Information Theory, 4, 76-84, 10.1109/TIT.1954.1057468, 1954.

Hamed, A., Joutsensaari, J., Mikkonen, S., Sogacheva, L., Dal Maso, M., Kulmala, M., Cavalli, F., Fuzzi, S., Facchini, M. C., Decesari, S., Mircea, M., Lehtinen, K. E. J., and Laaksonen, A.: Nucleation and growth of new particles in Po Valley, Italy, Atmos. Chem. Phys., 7, 355-376, 2007.

10 Han, X., Zhong, Y., Cao, L., and Zhang, L.: Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification, Remote Sens., 9, 848, 10.3390/rs9080848, 2017.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778, 2016.

Hebb, D. O.: The organization of behavior: A neuropsychological theory, Psychology Press, 2005.

15 Herrmann, E., Weingartner, E., Henne, S., Vuilleumier, L., Bukowiecki, N., Steinbacher, M., Conen, F., Collaud Coen, M., Hammer, E., Jurányi, Z., Baltensperger, U., and Gysel, M.: Analysis of long-term aerosol size distribution data from Jungfraujoch with emphasis on free tropospheric conditions, cloud influence, and air mass transport, J. Geophys. Res. Atmos., 120, 9459-9480, 10.1002/2015JD023660, 2015.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R.: Improving neural networks by 20 preventing co-adaptation of feature detectors, arXiv preprint, arXiv:1207.0580, 2012.

Hirsikko, A., Bergman, T., Laakso, L., Dal Maso, M., Riipinen, I., Hõrrak, U., and Kulmala, M.: Identification and classification of the formation of intermediate ions measured in boreal forest, Atmos. Chem. Phys., 7, 201-210, 10.5194/acp-7-201-2007, 2007.

Hu, F., Xia, G.-S., Hu, J., and Zhang, L.: Transferring Deep Convolutional Neural Networks for the Scene Classification of 25 High-Resolution Remote Sensing Imagery, Remote Sens., 7, 14680, 2015.

Huber, P. J.: Robust Statistics, John Wiley, New York, 1981.

Huttunen, J., Kokkola, H., Mielonen, T., Mononen, M. E. J., Lipponen, A., Reunanen, J., Lindfors, A. V., Mikkonen, S., Lehtinen, K. E. J., Kouremeti, N., Bais, A., Niska, H., and Arola, A.: Retrieval of aerosol optical depth from surface solar radiation measurements using machine learning algorithms, non-linear regression and a radiative transfer-based look-up 30 table, Atmos. Chem. Phys., 16, 8181-8191, 10.5194/acp-16-8181-2016, 2016.

Hyvönen, S., Junninen, H., Laakso, L., Dal Maso, M., Grönholm, T., Bonn, B., Keronen, P., Aalto, P., Hiltunen, V., Pohja, T., Launiainen, S., Hari, P., Mannila, H., and Kulmala, M.: A look at aerosol formation using data mining techniques, Atmos. Chem. Phys., 5, 3345-3356, 10.5194/acp-5-3345-2005, 2005.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., 2013.

Jeong, C.-H., Hopke, P. K., Chalupa, D., and Utell, M.: Characteristics of Nucleation and Growth Events of Ultrafine Particles Measured in Rochester, NY, Environ. Sci. Technol., 38, 1933-1940, 10.1021/es034811p, 2004.

Junninen, H., Riipinen, I., Dal Maso, M., and Kulmala, M.: An algorithm for automatic classification of two-dimensional aerosol data, in: Nucleation and Atmospheric Aerosols, Springer, Dordrecht, 957-961, 2007.

Kaelbling, L. P., Littman, M. L., and Moore, A. W.: Reinforcement learning: a survey, J. Artif. Int. Res., 4, 237-285, 10.1613/jair.301, 1996.

Kerminen, V. M., Paramonov, M., Anttila, T., Riipinen, I., Fountoukis, C., Korhonen, H., Asmi, E., Laakso, L., Lihavainen, H., Swietlicki, E., Svenningsson, B., Asmi, A., Pandis, S. N., Kulmala, M., and Petäjä, T.: Cloud condensation nuclei production associated with atmospheric nucleation: a synthesis based on existing literature and new results, Atmos. Chem. Phys., 12, 12037-12059, 10.5194/acp-12-12037-2012, 2012.

Kim, Y., Yoon, S.-C., Kim, S.-W., Kim, K.-Y., Lim, H.-C., and Ryu, J.: Observation of new particle formation and growth events in Asian continental outflow, Atmos. Environ., 64, 160-168, 10.1016/j.atmosenv.2012.09.057, 2013.

Kim, Y., Kim, S.-W., and Yoon, S.-C.: Observation of new particle formation and growth under cloudy conditions at Gosan Climate Observatory, Korea, Meteorol. Atmos. Phys., 126, 81-90, 10.1007/s00703-014-0336-2, 2014.

Kivekäs, N., Sun, J., Zhan, M., Kerminen, V. M., Hyvärinen, A., Komppula, M., Viisanen, Y., Hong, N., Zhang, Y., Kulmala, M., Zhang, X. C., Deli, G., and Lihavainen, H.: Long term particle size distribution measurements at Mount Waliguan, a high-altitude site in inland China, Atmos. Chem. Phys., 9, 5461-5474, 10.5194/acp-9-5461-2009, 2009.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems 25, edited by: Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., Curran Associates, Inc., 1097-1105, 2012.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, Commun. ACM, 60, 84-90, 10.1145/3065386, 2017.

Kulmala, M., Vehkamäki, H., Petäjä, T., Dal Maso, M., Lauri, A., Kerminen, V. M., Birmili, W., and McMurry, P. H.: Formation and growth rates of ultrafine atmospheric particles: a review of observations, J. Aerosol Sci., 35, 143-176, 2004.

Kulmala, M., Petäjä, T., Nieminen, T., Sipilä, M., Manninen, H. E., Lehtipalo, K., Dal Maso, M., Aalto, P. P., Junninen, H., Paasonen, P., Riipinen, I., Lehtinen, K. E. J., Laaksonen, A., and Kerminen, V.-M.: Measurement of the nucleation of atmospheric aerosol particles, Nat. Protoc., 7, 1651-1667, 10.1038/nprot.2012.091, 2012.

Kulmala, M., Petäjä, T., Kerminen, V.-M., Kujansuu, J., Ruuskanen, T., Ding, A., Nie, W., Hu, M., Wang, Z., Wu, Z., Wang, L., and Worsnop, D. R.: On secondary new particle formation in China, Front. Env. Sci. Eng., 10, 8, 10.1007/s11783-016-0850-1, 2016.

Kyrö, E. M., Väänänen, R., Kerminen, V. M., Virkkula, A., Petäjä, T., Asmi, A., Dal Maso, M., Nieminen, T., Juhola, S., Shcherbinin, A., Riipinen, I., Lehtipalo, K., Keronen, P., Aalto, P. P., Hari, P., and Kulmala, M.: Trends in new particle formation in eastern Lapland, Finland: effect of decreasing sulfur emissions from Kola Peninsula, Atmos. Chem. Phys., 14, 4383-4396, 10.5194/acp-14-4383-2014, 2014.

5   Laaksonen, A., Hamed, A., Joutsensaari, J., Hiltunen, L., Cavalli, F., Junkermann, W., Asmi, A., Fuzzi, S., and Facchini, M. C.: Cloud condensation nucleus production from nucleation events at a highly polluted region, Geophys. Res. Lett., 32, L06812, 10.1029/2004gl022092, 2005.

Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L.: Machine learning in geosciences and remote sensing, Geosci. Front., 7, 3-10, 10.1016/j.gsf.2015.07.003, 2016.

10   Le, Q. V.: Building high-level features using large scale unsupervised learning, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, 8595-8598,

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86, 2278-2324, 10.1109/5.726791, 1998.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436-444, 10.1038/nature14539, 2015.

15   Li, X., Peng, L., Hu, Y., Shao, J., and Chi, T.: Deep learning architecture for air quality predictions, Environ. Sci. Pollut. Res. Int., 23, 22408-22417, 10.1007/s11356-016-7812-9, 2016.

Lähivaara, T., Kärkkäinen, L., Huttunen, J. M. J., and Hesthaven, J. S.: Deep convolutional neural networks for estimating porous material parameters with ultrasound tomography, ~~arXiv preprint, arXiv:1709.09212, 2017~~The Journal of the Acoustical Society of America, 143, 1148-1158, 10.1121/1.5024341, 2018.

20   Ma, Y., Gong, W., and Mao, F.: Transfer learning used to analyze the dynamic evolution of the dust aerosol, J. Quant. Spectrosc. Radiat. Transfer, 153, 119-130, 10.1016/j.jqsrt.2014.09.025, 2015.

MathWorks: Deep Learning: Transfer Learning in 10 lines of MATLAB Code: https://se.mathworks.com/matlabcentral/fileexchange/61639-deep-learning--transfer-learning-in-10-lines-of-matlab-code, last access: 29.11.2017, 2017.

25   MathWorks: MathWorks Documentation: imresize: https://se.mathworks.com/help/images/ref/imresize.html, last access: 14.6.2018, 2018.

Merikanto, J., Spracklen, D. V., Mann, G. W., Pickering, S. J., and Carslaw, K. S.: Impact of nucleation on global CCN, Atmos. Chem. Phys., 9, 8601-8616, 10.5194/acp-9-8601-2009, 2009.

Mesnil, G., Dauphin, Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I., Lavoie, E., Muller, X., Desjardins, G., Warde-
30   Farley, D., Vincent, P., Courville, A., and Bergstra, J.: Unsupervised and Transfer Learning Challenge: a Deep Learning Approach, Proceedings of Machine Learning Research, Proceedings of ICML Workshop on Unsupervised and Transfer Learning, 27, 97-110, 2012.

Mikkonen, S., Lehtinen, K. E. J., Hamed, A., Joutsensaari, J., Facchini, M. C., and Laaksonen, A.: Using discriminant analysis as a nucleation event classification method, Atmos. Chem. Phys., 6, 5549-5557, 10.5194/acp-6-5549-2006, 2006.

Mikkonen, S., Korhonen, H., Romakkaniemi, S., Smith, J. N., Joutsensaari, J., Lehtinen, K. E. J., Hamed, A., Breider, T. J., Birmili, W., Spindler, G., Plass-Duelmer, C., Facchini, M. C., and Laaksonen, A.: Meteorological and trace gas factors affecting the number concentration of atmospheric Aitken (D_p = 50 nm) particles in the continental boundary layer: parameterization using a multivariate mixed effects model, Geosci. Model Dev., 4, 1-13, 10.5194/gmd-4-1-2011, 2011.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M.: Playing atari with deep reinforcement learning., arXiv preprint, arXiv:1312.5602, 2013.

Mäkelä, J. M., Dal Maso, M., Pirjola, L., Keronen, P., Laakso, L., Kulmala, M., and Laaksonen, A.: Characteristics of the atmospheric particle formation events observed at a borel forest site in southern Finland, Boreal Environ. Res., 5, 299-313, 2000.

Nel, A.: Air Pollution-Related Illness: Effects of Particles, Science, 308, 804-806, 10.1126/science.1108752, 2005.

Nieminen, T., Asmi, A., Dal Maso, M., Aalto, P. P., Keronen, P., Petaja, T., Kulmala, M., and Kerminen, V. M.: Trends in atmospheric new-particle formation: 16 years of observations in a boreal-forest environment, Boreal Environ. Res., 19, 191-214, 2014.

Nieminen, T., Yli-Juuti, T., Manninen, H. E., Petäjä, T., Kerminen, V. M., and Kulmala, M.: Technical note: New particle formation event forecasts during PEGASOS–Zeppelin Northern mission 2013 in Hyytiälä, Finland, Atmos. Chem. Phys., 15, 12385-12396, 10.5194/acp-15-12385-2015, 2015.

Nieminen, T., Kerminen, V. M., Petäjä, T., Aalto, P. P., Arshinov, M., Asmi, E., Baltensperger, U., Beddows, D. C. S., Beukes, J. P., Collins, D., Ding, A., Harrison, R. M., Henzing, B., Hooda, R., Hu, M., Hõrrak, U., Kivekäs, N., Komsaare, K., Krejci, R., Kristensson, A., Laakso, L., Laaksonen, A., Leaitch, W. R., Lihavainen, H., Mihalopoulos, N., Németh, Z., Nie, W., O'Dowd, C., Salma, I., Sellegri, K., Svenningsson, B., Swietlicki, E., Tunved, P., Ulevicius, V., Vakkari, V., Vana, M., Wiedensohler, A., Wu, Z., Virtanen, A., and Kulmala, M.: Global analysis of continental boundary layer new particle formation based on long-term measurements, Atmos. Chem. Phys. Discuss., 2018, 1-34, 10.5194/acp-2018-304, 2018.

O'Dowd, C., Monahan, C., and Dall'Osto, M.: On the occurrence of open ocean particle production and growth events, Geophys. Res. Lett., 37, doi:10.1029/2010GL044679, 2010.

Ong, B. T., Sugiura, K., and Zettsu, K.: Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM2.5, Neural Comput. Appl., 27, 1553-1566, 10.1007/s00521-015-1955-3, 2016.

Pan, S. J., and Yang, Q.: A Survey on Transfer Learning, IEEE T. Knowl. Data Eng., 22, 1345-1359, 10.1109/TKDE.2009.191, 2010.

Pascanu, R., Mikolov, T., and Bengio, Y.: On the difficulty of training recurrent neural networks, Proceedings of the 30th International Conference on Machine Learning, Proceedings of Machine Learning Research, 28, 1310-1318, 2013.
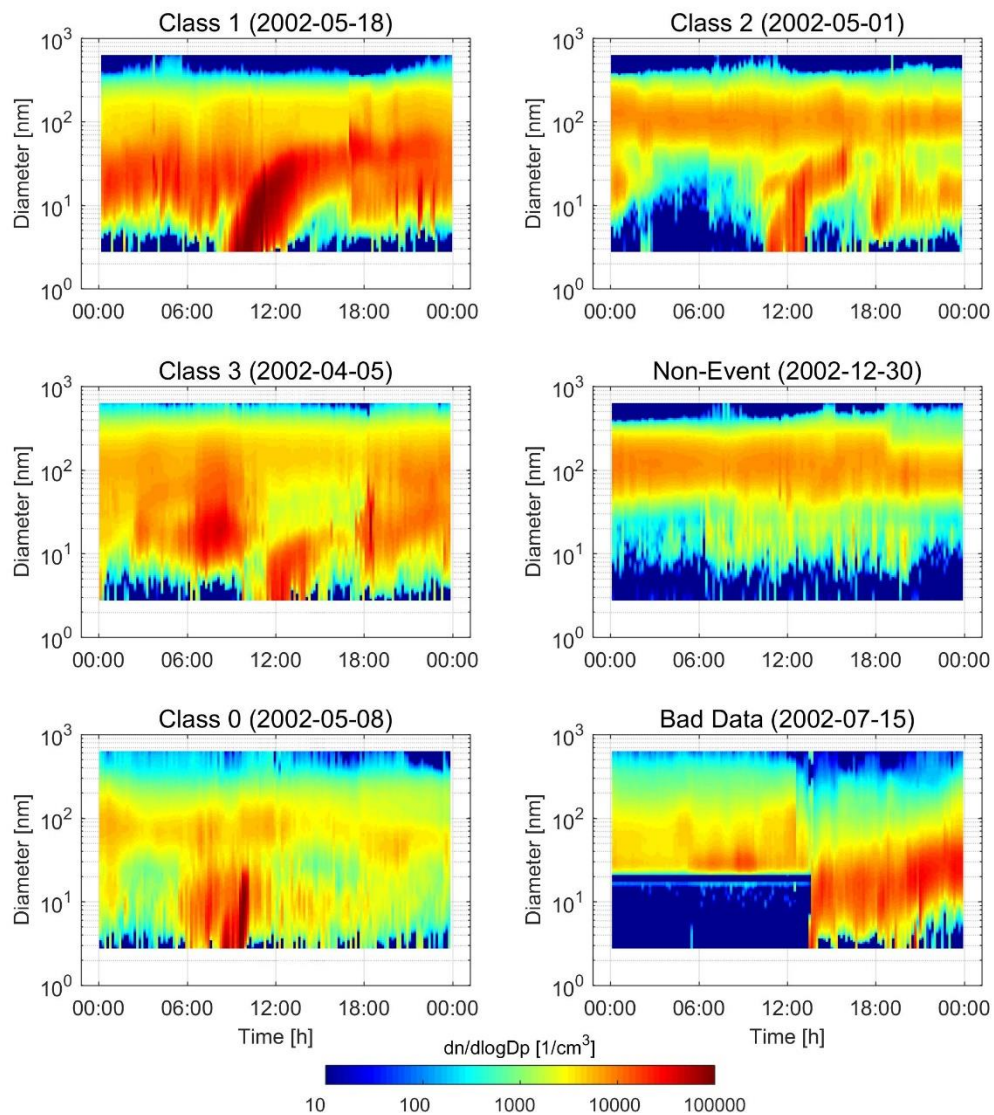
Qi, X. M., Ding, A. J., Nie, W., Petäjä, T., Kerminen, V. M., Herrmann, E., Xie, Y. N., Zheng, L. F., Manninen, H., Aalto, P., Sun, J. N., Xu, Z. N., Chi, X. G., Huang, X., Boy, M., Virkkula, A., Yang, X. Q., Fu, C. B., and Kulmala, M.: Aerosol size distribution and new particle formation in the western Yangtze River Delta of China: 2 years of measurements at the SORPES station, Atmos. Chem. Phys., 15, 12445-12464, 10.5194/acp-15-12445-2015, 2015.

5   R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria: https://www.R-project.org/, 2017.

Radford, A., Metz, L., and Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint, arXiv:1511.06434v2, 2016.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning internal representations by error propagation, in: Parallel

10   distributed processing: explorations in the microstructure of cognition, vol. 1, edited by: David, E. R., James, L. M., and Group, C. P. R., MIT Press, 318-362, 1986.

Ruske, S., Topping, D. O., Foot, V. E., Kaye, P. H., Stanley, W. R., Crawford, I., Morse, A. P., and Gallagher, M. W.: Evaluation of machine learning algorithms for classification of primary biological aerosol using a new UV-LIF spectrometer, Atmos. Meas. Tech., 10, 695-708, 10.5194/amt-10-695-2017, 2017.

15   Samet , J. M., Dominici , F., Curriero , F. C., Coursac , I., and Zeger , S. L.: Fine Particulate Air Pollution and Mortality in 20 U.S. Cities, 1987–1994, New Engl. J. Med., 343, 1742-1749, 10.1056/nejm200012143432401, 2000.

Schmeissner, T., Krejci, R., Ström, J., Birmili, W., Wiedensohler, A., Hochschild, G., Gross, J., Hoffmann, P., and Calderon, S.: Analysis of number size distributions of tropical free tropospheric aerosol particles observed at Pico Espejo (4765 m a.s.l.), Venezuela, Atmos. Chem. Phys., 11, 3319-3332, 10.5194/acp-11-3319-2011, 2011.

20   Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Networks, 61, 85-117, 10.1016/j.neunet.2014.09.003, 2015.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M.: Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, IEEE T. Med. Imaging, 35, 1285-1298, 10.1109/TMI.2016.2528162, 2016.

25   Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., and Hassabis, D.: Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm, arXiv preprint, arXiv:1712.01815, 2017a.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., and Hassabis, D.: Mastering the game of Go

30   without human knowledge, Nature, 550, 354-359, 10.1038/nature24270, 2017b.

Spracklen, D. V., Carslaw, K. S., Kulmala, M., Kerminen, V. M., Sihto, S. L., Riipinen, I., Merikanto, J., Mann, G. W., Chipperfield, M. P., Wiedensohler, A., Birmili, W., and Lihavainen, H.: Contribution of particle formation to global cloud condensation nuclei concentrations, Geophys. Res. Lett., 35, L06808, 10.1029/2007gl033038, 2008.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res., 15, 1929-1958, 2014.

Sutton, R. S.: Learning to Predict by the Methods of Temporal Differences, Mach. Learn., 3, 9-44, 10.1023/a:1022663531479, 1988.

5   Szegedy, C., ~~Wei, L., Yangqing, J.,~~ Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7-12 June 2015, 1-9, 10.1109/CVPR.2015.7298594, 2015.

Twomey, S.: Pollution and the planetary albedo, Atmos. Environ., 8, 1251-1256, doi.org/10.1016/0004-6981(74)90004-3, 1974.

10   Vaattovaara, P., Huttunen, P. E., Yoon, Y. J., Joutsensaari, J., Lehtinen, K. E. J., O'Dowd, C. D., and Laaksonen, A.: The composition of nucleation and Aitken modes particles during coastal nucleation events: evidence for marine secondary organic contribution, Atmos. Chem. Phys., 6, 4601-4616, 2006.

Vana, M., Ehn, M., Petäjä, T., Vuollekoski, H., Aalto, P., de Leeuw, G., Ceburnis, D., O'Dowd, C. D., and Kulmala, M.: Characteristic features of air ions at Mace Head on the west coast of Ireland, Atmos. Res., 90, 278-286,
15   10.1016/j.atmosres.2008.04.007, 2008.

Wang, J.-F., Hu, M.-G., Xu, C.-D., Christakos, G., and Zhao, Y.: Estimation of Citywide Air Pollution in Beijing, PLOS ONE, 8, e53400, 10.1371/journal.pone.0053400, 2013.

Wang, Z., Birmili, W., Hamed, A., Wehner, B., Spindler, G., Pei, X., Wu, Z., Cheng, Y., Su, H., and Wiedensohler, A.: Contributions of volatile and nonvolatile compounds (at 300°C) to condensational growth of atmospheric nanoparticles:
20   An assessment based on 8.5 years of observations at the Central Europe background site Melpitz, J. Geophys. Res. Atmos., 122, 485-497, 10.1002/2016JD025581, 2017a.

Wang, Z., Wu, Z., Yue, D., Shang, D., Guo, S., Sun, J., Ding, A., Wang, L., Jiang, J., Guo, H., Gao, J., Cheung, H. C., Morawska, L., Keywood, M., and Hu, M.: New particle formation in China: Current knowledge and further directions, Sci. Total Environ., 577, 258-266, 10.1016/j.scitotenv.2016.10.177, 2017b.

25   Weiss, K., Khoshgoftaar, T. M., and Wang, D.: A survey of transfer learning, J. Big Data, 3, 9, 10.1186/s40537-016-0043-6, 2016.

Venables, W. N., and Ripley, B. D.: Modern Applied Statistics with S, Fourth Edition, Springer, New York, 2002.

WHO: Review of evidence on health aspects of air pollution - REVIHAAP Project: Technical Report, World Health Organization, WHO Regional Office for Europe, Copenhagen, 2013.

30   Widrow, B., and Hoff, M. E.: Adaptive Switching Circuits, Technical Report No. 1553-1, Stanford Electronics Laboratories, Stanford University, CA, 1960.

Winklmayr, W., Reischl, G. P., Lindner, A. O., and Berner, A.: A new electromobility spectrometer for the measurement of aerosol size distributions in the size range from 1 to 1000 nm, J. Aerosol Sci., 22, 289-296, 1991.

Voutilainen, A., Kolehmainen, V., and Kaipio, J. P.: Statistical inversion of aerosol size measurement data, Inverse Probl. Eng., 9, 67-94, 2001.

Vuollekoski, H., Junninen, H., Dal Maso, M., Kerminen, V.-M., and Kulmala, M.: An eigenvector-based approach for automatic classification of new particle formation events, in: Proceedings of Finnish Center of Excellence in 'Physics, Chemistry, Biology and Meteorology of Atmospheric Composition and Climate Change', and Nordic Center of Excellence in 'Cryosphere-Atmosphere Interactions in a Changing Arctic Climate', Annual Meetings 2012, edited by: Kulmala, M., Lappalainen, H. K., Boy, M., Brus, M., and Nieminen, T., Report Series in Aerosol Science, N:o 134, Helsinki, Finland, 653-654, 2012. http://www.atm.helsinki.fi/FAAR/reportseries/rs-134.pdf.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.: How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems 27, edited by: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., 3320-3328, 2014.

Yu, F., and Luo, G.: Simulation of particle size distribution with a global aerosol model: contribution of nucleation to aerosol and CCN number concentrations, Atmos. Chem. Phys., 9, 7691-7710, 10.5194/acp-9-7691-2009, 2009.

Zaidan, M. A., Haapasilta, V., Relan, R., Junninen, H., Aalto, P. P., Canova, F. F., Laurson, L., and Foster, A. S.: Neural network classifier on time series features for predicting atmospheric particle formation days, in: The 20th International Conference on Nucleation and Atmospheric Aerosols, edited by: Halonen, R., Nikandrova, A., Kontkanen, J., Enroth, J. A., and Vehkamäki, H., Report Series in Aerosol Science, N:o 200, Helsinki, Finland, 687-690, 2017. http://www.atm.helsinki.fi/FAAR/reportseries/rs-200.pdf.

Zeiler, M. D., and Fergus, R.: Visualizing and Understanding Convolutional Networks, in: Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I, edited by: Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., Springer International Publishing, Cham, 818-833, 2014. https://doi.org/10.1007/978-3-319-10590-1_53.

Zhang, L. P., Zhang, L. F., and Du, B.: Deep learning for remote sensing data: A technical tutorial on the state of the art, IEEE Geoscience and Remote Sensing Magazine, 4, 22-40, 10.1109/mgrs.2016.2540798, 2016.
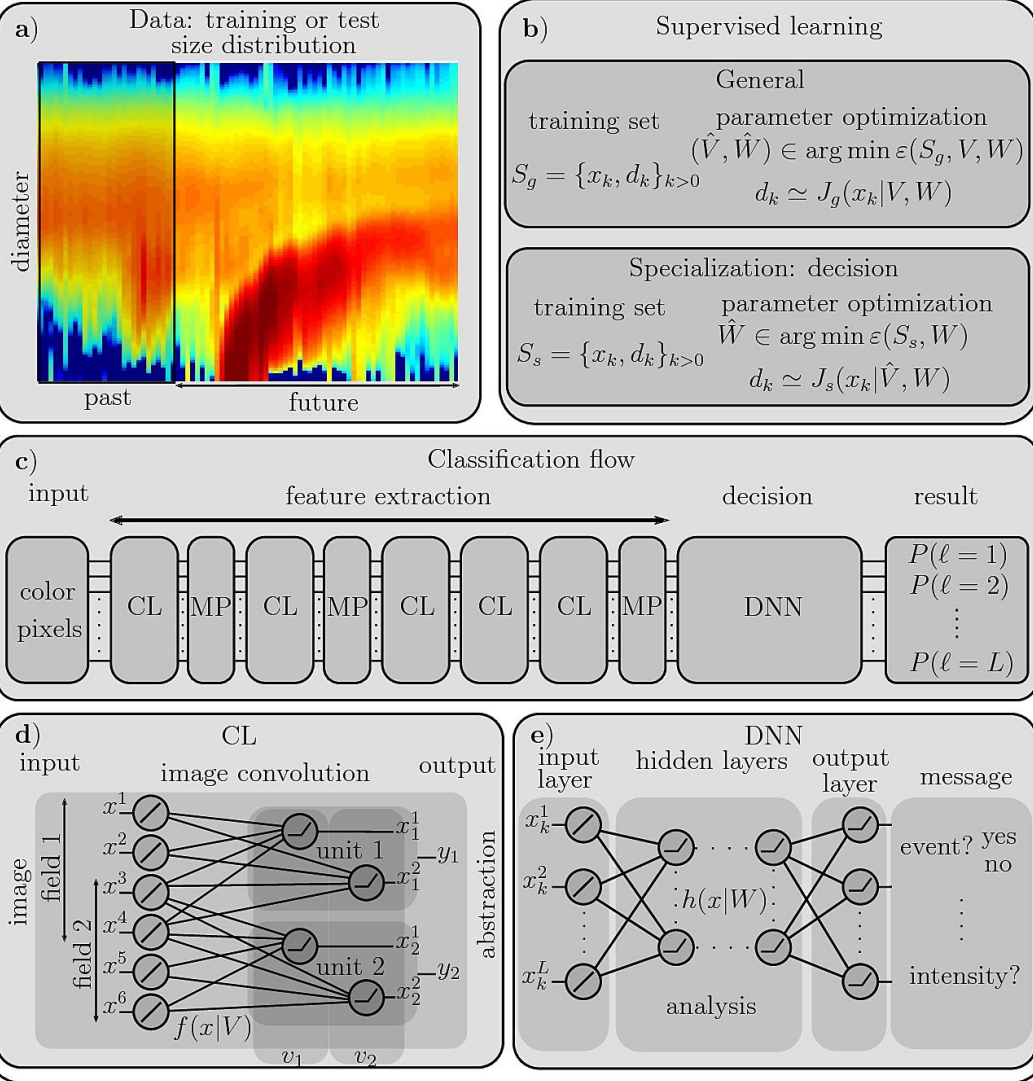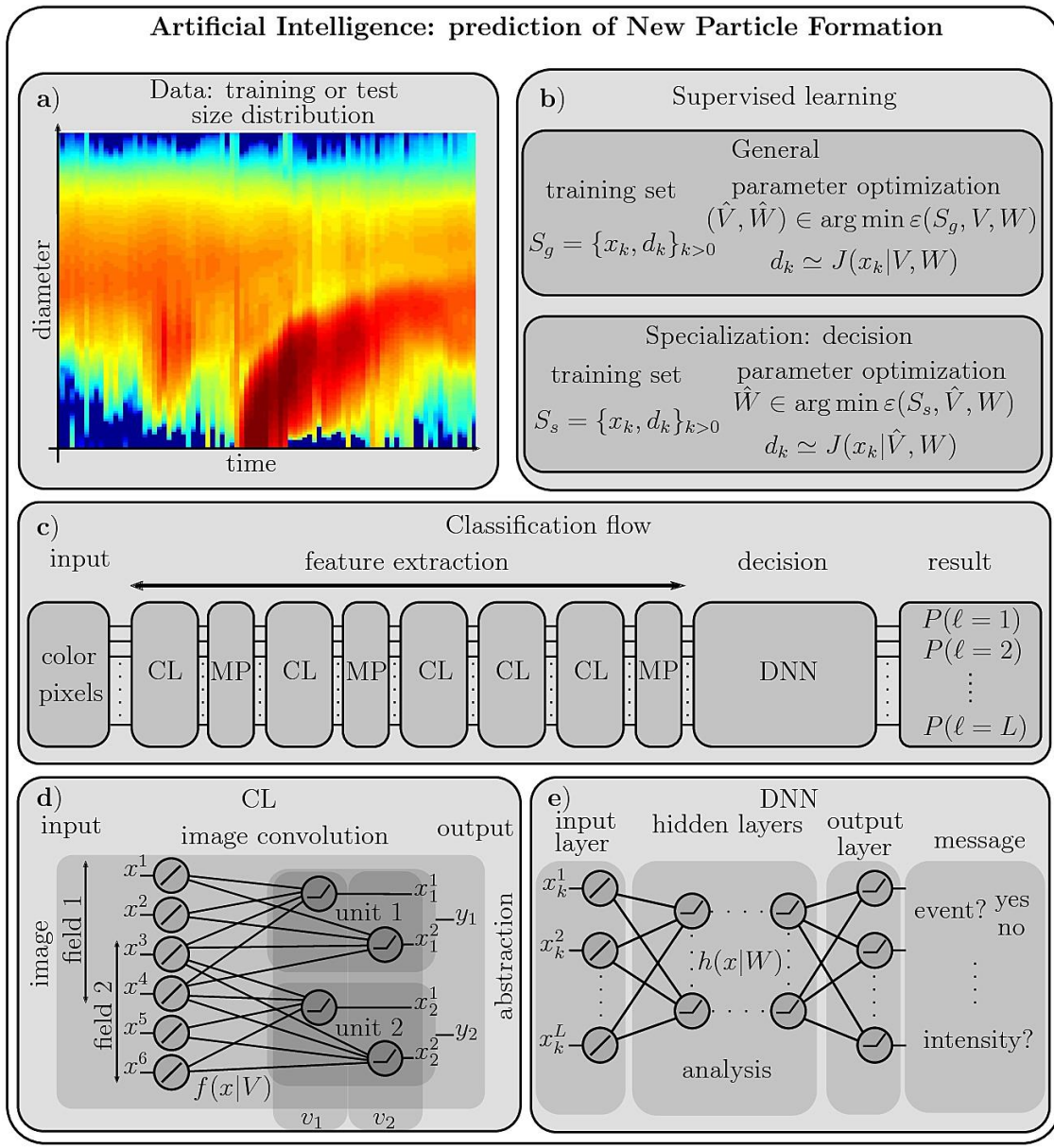
Figures:

Figure 1: Examples of particle size distributions (time in x-axis, particle diameter in y-axis and particle concentration presented by different colors) for different NPF event classes. Name of event class is indicated in plot title and date of measurement in brackets.

5

Artificial Intelligence: prediction of New Particle Formation

a) Data: training or test size distribution

diameter

past | future

b) Supervised learning

General

training set $\qquad$ parameter optimization
$$(\hat{V}, \hat{W}) \in \arg\min \varepsilon(S_g, V, W)$$
$$S_g = \{x_k, d_k\}_{k>0}$$
$$d_k \simeq J_g(x_k | V, W)$$

Specialization: decision

training set $\qquad$ parameter optimization
$$\hat{W} \in \arg\min \varepsilon(S_s, W)$$
$$S_s = \{x_k, d_k\}_{k>0}$$
$$d_k \simeq J_s(x_k | \hat{V}, W)$$

c) Classification flow

input | feature extraction | decision | result

color pixels | CL | MP | CL | MP | CL | CL | CL | MP | DNN | $P(\ell = 1)$ $P(\ell = 2)$ $\vdots$ $P(\ell = L)$

d) CL

input | image convolution | output

image convolution

field 1

field 2

image

x^1
x^2
x^3
x^4
x^5
x^6

unit 1 $\quad x_1^1$
$\quad x_1^2 - y_1$

unit 2 $\quad x_2^1$
$\quad x_2^2 - y_2$

$f(x|V)$

$v_1 \quad v_2$

abstraction

e) DNN

input layer | hidden layers | output layer | message

x_k^1
x_k^2
$\vdots$
x_k^L

$h(x|W)$

$\cdots$

analysis

event? yes no
$\vdots$
intensity?

29

Artificial Intelligence: prediction of New Particle Formation

a) Data: training or test size distribution

diameter / time

b) Supervised learning

General

training set
$$S_g = \{x_k, d_k\}_{k>0}$$

parameter optimization
$$(\hat{V}, \hat{W}) \in \arg\min \varepsilon(S_g, V, W)$$
$$d_k \simeq J(x_k | V, W)$$

Specialization: decision

training set
$$S_s = \{x_k, d_k\}_{k>0}$$

parameter optimization
$$\hat{W} \in \arg\min \varepsilon(S_s, \hat{V}, W)$$
$$d_k \simeq J(x_k | \hat{V}, W)$$

c) Classification flow

input — feature extraction — decision — result

color pixels | CL | MP | CL | MP | CL | CL | CL | MP | DNN
$$P(\ell = 1)$$
$$P(\ell = 2)$$
$$\vdots$$
$$P(\ell = L)$$

d) CL

input — image convolution — output

image field 1, field 2
$x^1, x^2, x^3, x^4, x^5, x^6$
unit 1: x_1^1, x_1^2 — y_1
unit 2: x_2^1, x_2^2 — y_2
$f(x|V)$
v_1 v_2
abstraction

e) DNN

input layer — hidden layers — output layer — message

x_k^1
x_k^2
x_k^L
$h(x|W)$
analysis
event? yes / no
$\vdots$
intensity?

Figure 2: Visual summary of the classification method: a) a typical dataset with a NPF event, b) the learning process of the optimization problem, c) the classification flow of the Convolutional Neural Network (CNN) and the two types of layers of the CNN, d) the Convolutional Layer (CL) and e) the Deep Neural Network (DNN), involved in the total CNN. The classification flow (c) is from the AlexNet CNN (Krizhevsky et al., 2012) composed of 5 CL intertwined with Max Pooling (MP) layers followed by a fully connected DNN of 3 layers. The method and variables in the figure are described in detail in Chapter 2.3 and Appendix A.

5

Figure 3: Confusion matrices of a) training and b) testing data sets from one run of Image Set 1 with training-testing ratio of 50 %/50 %. Numbers (percent and absolute number of days) in rows indicate how a certain class classified by human (visually) were categorized in the CNN classification.
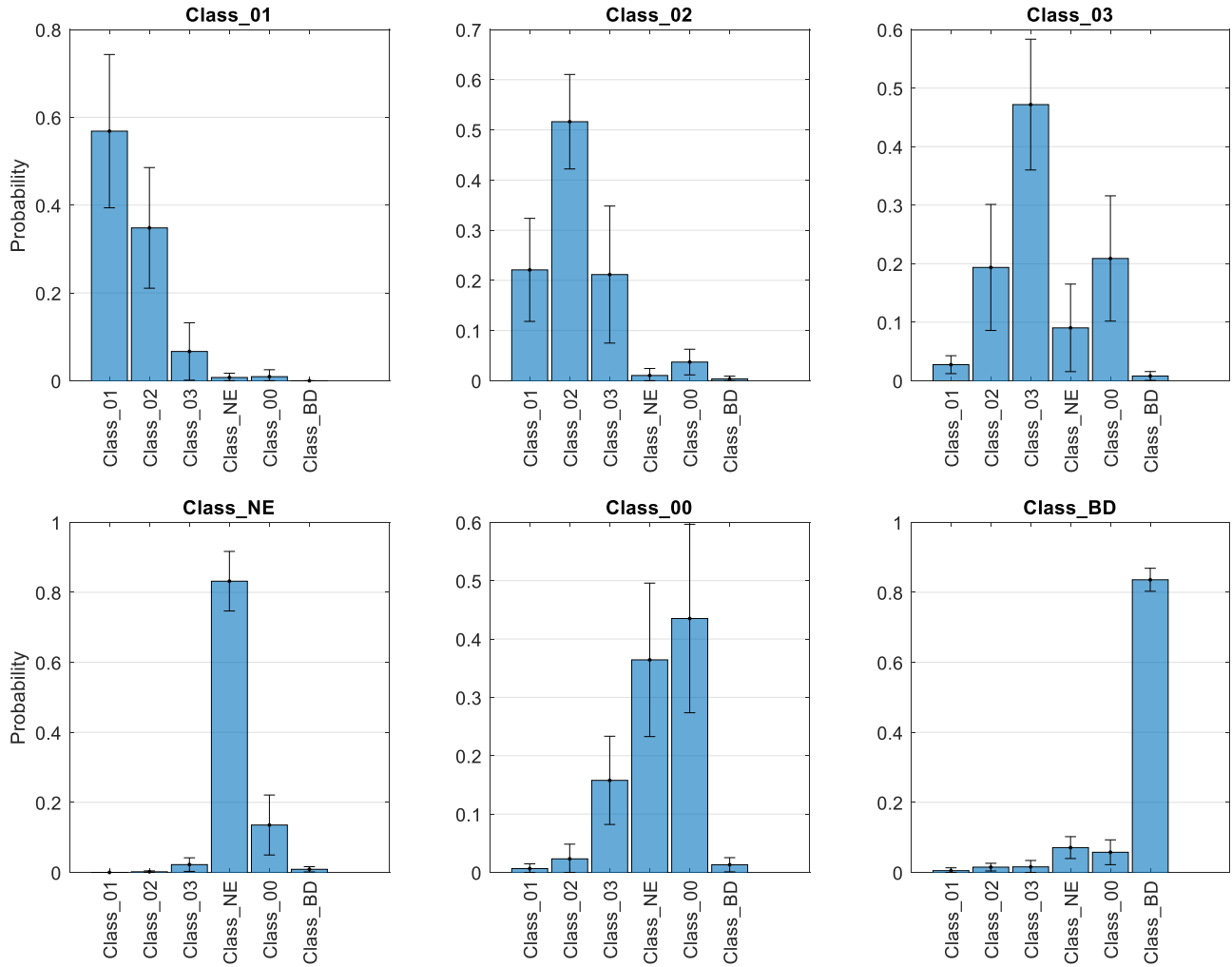
Figure 4: Probability plots of distribution of CNN classification (x-axis labels) for different human-made classifications (indicated in a title of the plot), for Image Set 1 with training-testing ratio of 80 %/20 %. Histograms are mean values of ten different training-testing runs and error bars indicate standard deviations of the results.
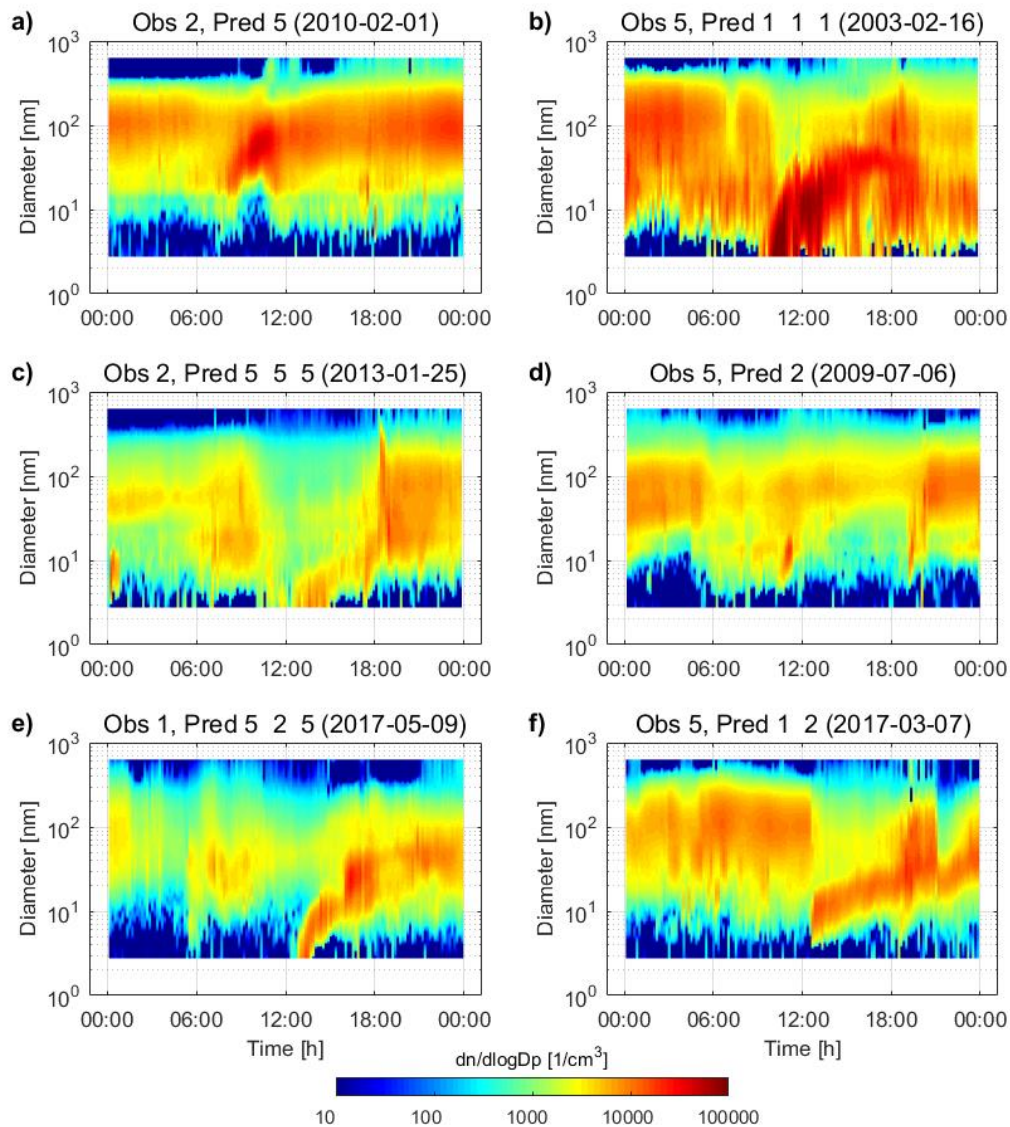
Figure 5: Examples of particle size distributions for different days when visually categorized clear event days (Classes 1 and 2) are classified Non-Event days by CNN classification or vice versa. In title, numbers after *Obs* and *Pred* show NPF event classes determined with visual and CNN-based classification, respectively (date of measurement in brackets). Several event class numbers in Pred-cases are results from different calculation runs. The first row (a and b) contains examples of human made misclassifications, the second row (c and d) CNN misclassifications and the third row (e and f) ambiguous situations (low particle concentrations and change in air masses during day, respectively). CNN classification is from Image Set 1 with training-testing ratio of 80 %/20 %.

Tables:

Table 1. Summary of number of days in different NPF event classes based on manual classification and division ~~into~~of days to three different training (used in CNN learning) and testing (used in testing success of training) ratio categories.~~ (80 %/20 %, 50 %/50% and 20 %/80 %).~~

Event class	Days Total	Days %	Training 80 %	Testing 20 %	Training 50 %	Testing 50 %	Training 20 %	Testing 80 %
Class_01	269	6 %	215	54	135	134	54	215
Class_02	431	10 %	345	86	216	215	86	345
Class_03	619	15 %	495	124	310	309	124	495
Class_NE	1416	34 %	1133	283	708	708	283	1133
Class_00	841	20 %	673	168	421	420	168	673
Class_BD	601	14 %	481	120	301	300	120	481
Total	4177	100 %	3342	835	2091	2086	835	3342

Table 2. Summary of classification accuracy (%) of the transfer learned CNN when applied for test datasets (mean value ± standard deviation from ten model simulations) for three image sets with three different percentage of training images. Some classes have been merged together: CL_01-02 is a combination of Class_01 and Class_02, etc.

Event class	Image Set 1 80 %	Image Set 1 50 %	Image Set 1 20 %	Image Set 2 80 %	Image Set 2 50 %	Image Set 2 20 %	Image Set 3 80 %	Image Set 3 50 %	Image Set 3 20 %	Overall
All_Class	65 ± 1	64 ± 1	61 ± 1	65 ± 1	64 ± 1	60 ± 3	62 ± 3	63 ± 1	60 ± 2	63 ± 2
Class_01	57 ± 17	54 ± 20	51 ± 15	63 ± 11	38 ± 10	53 ± 27	54 ± 17	54 ± 10	47 ± 22	52 ± 7
Class_02	52 ± 9	51 ± 15	31 ± 17	40 ± 17	57 ± 14	27 ± 19	47 ± 20	45 ± 14	45 ± 21	44 ± 10
Class_03	47 ± 11	32 ± 11	50 ± 14	46 ± 12	42 ± 14	48 ± 14	40 ± 10	44 ± 10	40 ± 16	43 ± 6
Class_NE	83 ± 9	80 ± 8	81 ± 13	80 ± 10	80 ± 10	81 ± 18	77 ± 12	81 ± 10	79 ± 11	80 ± 2
Class_00	44 ± 16	53 ± 13	38 ± 18	51 ± 13	48 ± 12	36 ± 28	49 ± 12	43 ± 14	41 ± 14	45 ± 6
Class_BD	84 ± 3	85 ± 3	83 ± 3	85 ± 4	86 ± 2	80 ± 2	83 ± 4	84 ± 2	80 ± 2	83 ± 2
Cl_01-02	81 ± 11	85 ± 9	62 ± 19	76 ± 11	78 ± 12	62 ± 19	81 ± 11	77 ± 13	71 ± 24	75 ± 8
Cl_01-02-03	84 ± 6	79 ± 6	81 ± 6	82 ± 3	81 ± 5	80 ± 8	83 ± 6	82 ± 5	82 ± 3	81 ± 2
Cl_03-NE	77 ± 10	69 ± 10	78 ± 12	74 ± 8	73 ± 7	77 ± 18	70 ± 10	75 ± 10	72 ± 10	74 ± 3
Cl_03-00	63 ± 10	60 ± 7	61 ± 13	66 ± 10	63 ± 13	60 ± 18	62 ± 10	60 ± 9	58 ± 14	61 ± 2
Cl_NE-00	90 ± 5	93 ± 3	90 ± 4	92 ± 2	91 ± 3	91 ± 5	91 ± 3	90 ± 4	90 ± 2	91 ± 1
Cl_03-NE-00	93 ± 4	91 ± 4	94 ± 4	93 ± 3	93 ± 3	95 ± 3	92 ± 4	93 ± 3	91 ± 6	93 ± 1