Atmospheric
Chemistry
and Physics

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, 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: 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 by researchers manually 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 that a deep learning method, i.e., transfer learning of a convolutional neural network (CNN), has successfully been used to automatically classify NPF events into different classes directly from particle size distribution images, similarly to how the researchers carry out the manual classification. The developed method is based on image analysis of particle size distributions using a pretrained 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 the 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 of 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 nonevent 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, and 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 consistent for 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 "nonquantifiable" (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 nonquantifiable days. Furthermore, we investigated more deeply the days that are classified as clear events by experts and recognized as nonevents 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 pretrained 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 representative of 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 (SPC) 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 sometimes in the free troposphere, there are less and shorter continuous measurements of NPF available. However, NPF has also been observed to occur regularly in these high-altitude sites (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, 2013).

Currently, all the NPF studies published in the literature have utilized visually 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 visually 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. 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 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 (ROIs) from the time series of measured aerosol size-distribution data. These ROIs 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 also take into account 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 discriminant analysis for multiyear datasets of aerosol size distributions and several gas and meteorological parameters measured at Hyytiälä, Finland and San Pietro Capofiume, Italy, respectively. Both of these stud-

ies 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 (SOMs) and a decision tree to classify aerosol size distributions. More recently, a 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 machines (SVMs), Boltzmann machines (BMs), decision trees, etc. Our method choice is the deep feed-forward neural network (NN). The idea of a NN is not 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. 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 a pretrained 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 significantly reduces 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 Appendices A and B and Fig. 2.

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 pretrained 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 SVM approach for classification of dust and clouds from satellite data. Other applications in atmospheric science include, for example, air quality predictions (Li et al., 2016; Ong et al., 2016), characterization of aerosol particles with an ultraviolet laser-induced fluorescence (UV-LIF) spectrometer (Ruske et al., 2017) and aerosol retrievals from ground-based observations (Di Noia et al., 2015, 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-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 pretrained 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 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

In this study, we analyzed a long-term particle size distribution (PSD) dataset measured at the San Pietro Capofiume 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 the 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 and has a high population density, and hence it is one of the most polluted areas in Europe. On the average at SPC, NPF events occur on 36 % of the days whilst 33 % are clearly nonevent (NE) 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 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 were 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 2002 until 16 May 2017 (5534 days in total). 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 series of aerosol size distribution data and time evolution of nucleation-mode particles (particle diameter 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 are classified into days with NPF events and days without particle formation (nonevent days). A day is considered as a NPF event day if the 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 preexisting particles in the smallest size ranges. Class 1 events show an intensive and clear formation of small particles with continuous growth of particles for 7–10 h.

– *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, for example, fluctuations in the size distribution are larger. Furthermore, the growth lasts for a shorter time than for Class 1, being about 5 h 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 the 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 are 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 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 quantifiable and nonquantifiable 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 nonquantifiable days.

Other classes of days are as follows:

– *Nonevent days (NE)* (Class_NE). Days with no NPF in the nucleation mode particle size range are classified into nonevent days. These days are also interesting because, for example, differences in conditions (meteorology, gas concentration, precipitation) during event and nonevent periods are often studied in order to improve understanding of processes behind NPF.

– *Class 0* (Class_00). Some of days do not fulfil the criteria to be classified either event or nonevent 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.

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 Appendices A and B. We used a large, deep convolutional neural network named AlexNet (Krizhevsky et al., 2012, 2017), which was originally 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). Since the model was originally

Table 1. Summary of the 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 %).

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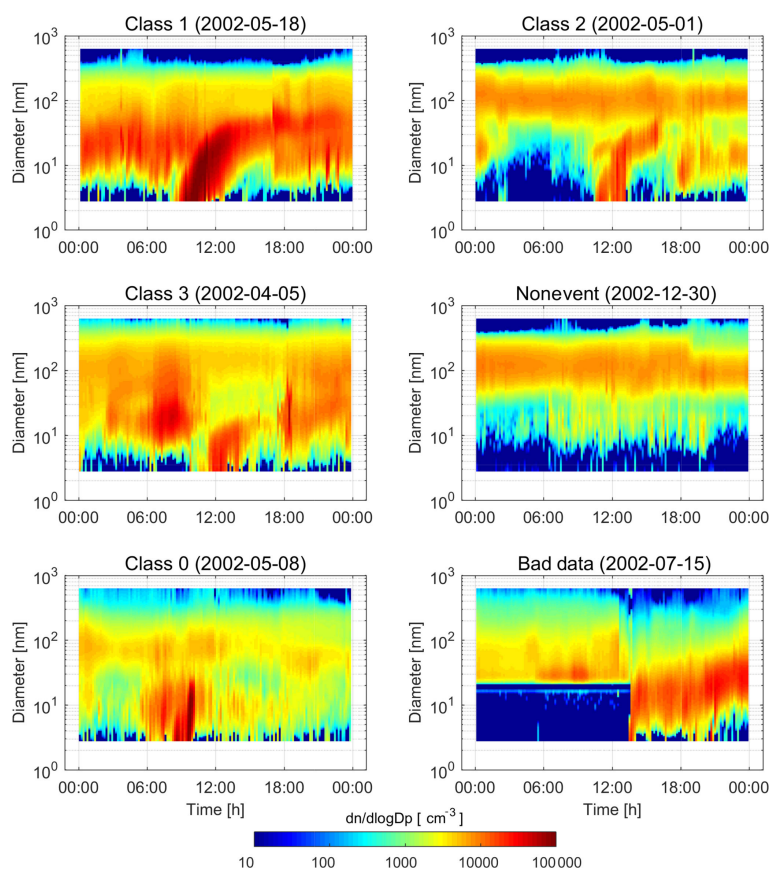
Figure 1. Examples of particle size distributions (time in x axis, particle diameter in y axis and particle concentration presented by different colors) for different NPF event classes. Name of event class is indicated in plot title and date of measurement in brackets.

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 im-ages 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 last 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.
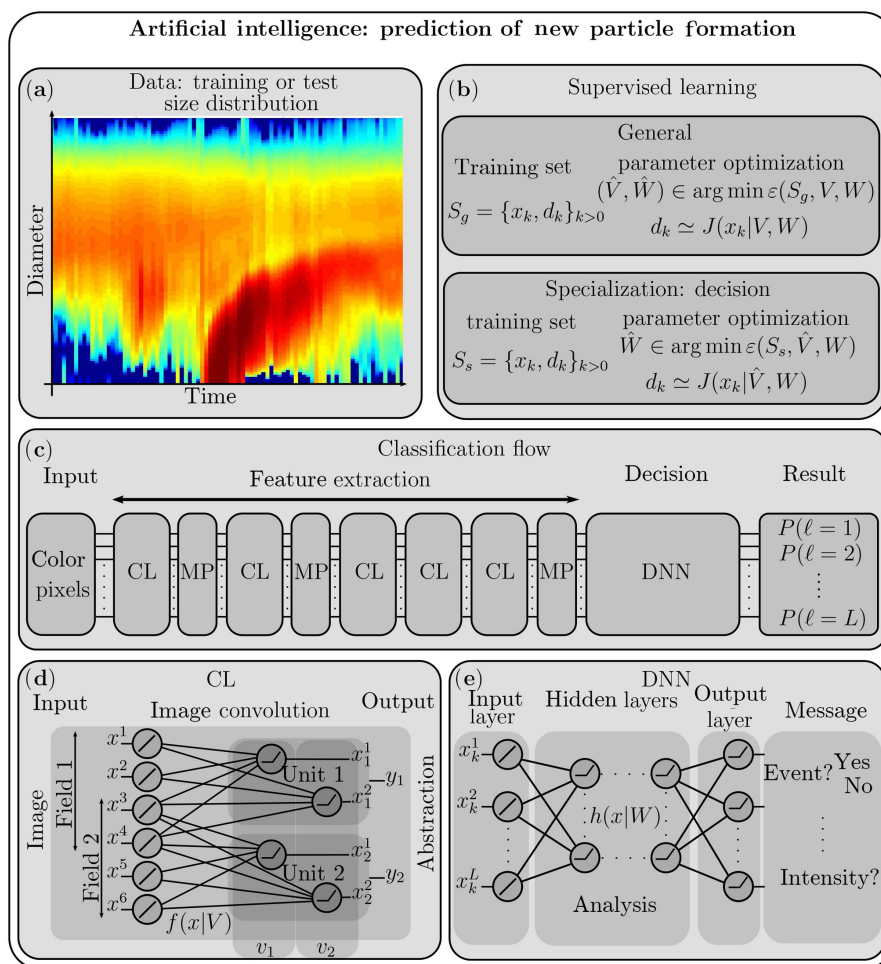
Figure 2. Visual summary of the classification method: **(a)** a typical dataset with a NPF event, **(b)** the learning process of the optimization problem, **(c)** the classification flow of the convolutional neural network (CNN) and the two types of layers of the CNN, **(d)** the convolutional layer (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 three layers. The method and variables in the figure are described in detail in Sect. 2.3 and Appendices A and B.

In a transfer learning process of our dataset, 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 and testing were selected randomly (certain percent of each category) and this was repeated 10 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 Neural Network Toolbox Model for the 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 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 0.5 to 1 h with the used server and programs. The AlexNet CNN has been originally implemented in CUDA, but it is also available in MATLAB. Other programming languages provide also good libraries for implementing the AlexNet such as Python or C++, and dedicated frameworks such as Ten-

soFlow, Theano or Caffe are good candidates for an easy implementation of the NN and the learning stage.

To find out the best performance for the NPF event classification, we tested three different sets of particle size distribution images (Image Set, JPEG 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 1801×951 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 1646×751 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 831×746 pixels (file size ca. 60 kB).

Image sets with the different sizes were tested because all images have to resize to 227×227 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 cause the loss of some information on 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

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 the 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 a 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 dataset 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 and 100 % (see Fig. 3a). For all cases (all image sets and training ratios), mean accuracies of 10 different runs varied from 93 to 98 %. The results show that the method can easily classify training datasets 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 datasets (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 the complete absence of particles, and unusually high particle concentration in a part of the 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 nonevent 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 nonquantifiable 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 and nonquantifiable events, in addition to nonevent, undefined (Class 0) and bad-data 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
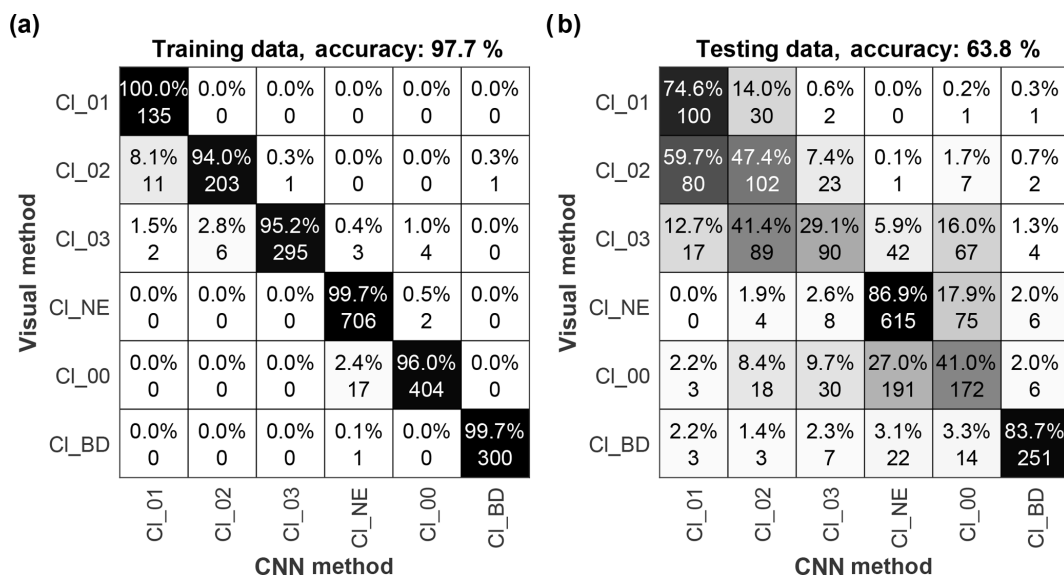
Figure 3. Confusion matrices of **(a)** training and **(b)** testing datasets from one run of Image Set 1 with a training–testing ratio of 50 % / 50 %. Numbers (percent and absolute number of days) in rows indicate how a certain class classified by a human (visually) were categorized in the CNN classification.

Table 2. Summary of classification accuracy (%) of the transfer-learned CNN when applied for test datasets (mean value ± SD from 10 model simulations) for three image sets with three different percentages 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			Image Set 2			Image Set 3			Overall
	80 %	50 %	20 %	80 %	50 %	20 %	80 %	50 %	20 %	
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

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 com-

promise between accuracy, reliability and number of training days for the used dataset.

We also studied the effect of image size (Image Sets 1–3) for classification accuracy. Only when all event classes were inspected together was there 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 subgroups 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 (06:00–18:00 UTC+1). Although images have to resize to the fixed input size (227 × 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 the resizing of images might cause the loss of some 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 a 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 a 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 nonevent 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 a human and right-hand plots by CNN; the first row shows 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 a human) was categorized 2 times into nonevent 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, Figure 5a and b show examples of misclassifications, which are due to clear human errors – probably that 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 preprocessing (fitting of lognormal distributions to the data), feature calculations and extractions (e.g., mean size, standard deviation) 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 nonevent days were considered. When compared to our method, they used several preprocessing steps before classification by the neural network and they did not use pretrained 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) and determines, for example, 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 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, Nonevent and undefined). The overall accuracy of the method was ca. 80 %. When comparing to our method, they preprocessed the data by decreasing the size resolution to 15 bins with variable width size bins in logscale (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. (2006, 2011) 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.
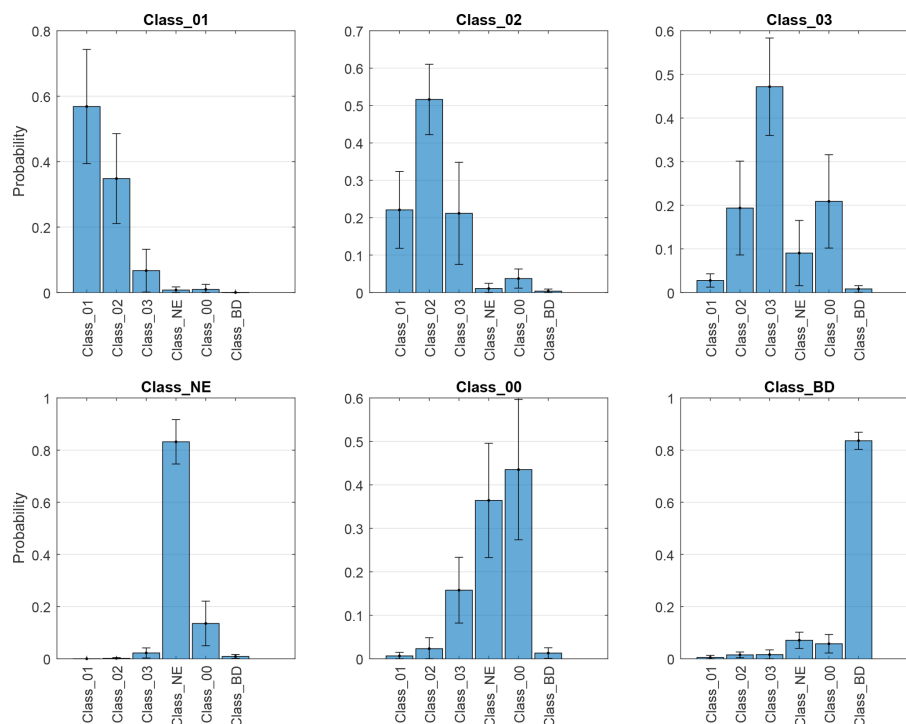
Figure 4. Probability plots of distribution of CNN classification (*x* axis labels) for different human-made classifications (indicated in a title of the plot), for Image Set 1 with a training–testing ratio of 80 % / 20 %. Histograms are mean values of 10 different training–testing runs and error bars indicate standard deviations of the results.

We have not yet tested the method at other sites. Basically, "banana-type" events, nonevent days and bad data should be recognized from other site data if figures are plotted roughly in a similar way (1-day plot, same size ranges and axes, and color map). The method analyses features from size distribution plots, which are quite similar at different sites in many cases. 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 pretrained convolutional neural network (CNN) called 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 CNNs, has successfully been used to classify automatically NPF events into different classes, including several event and nonevent 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 vary, especially for long-term datasets, which have been analyzed by several researchers at 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 pretrained CNN to recognize images of particle size distributions is a very powerful tool to analyze NPF events. The event classification can be done directly from existing data (figures) without any preprocessing of the data. Although an average classification accuracy of certain classes is ca. 65 %, the overall accuracy is ca. 80 % for nonevent (NE) and clear event classes (classes 1 and 2 combined), which is a good first
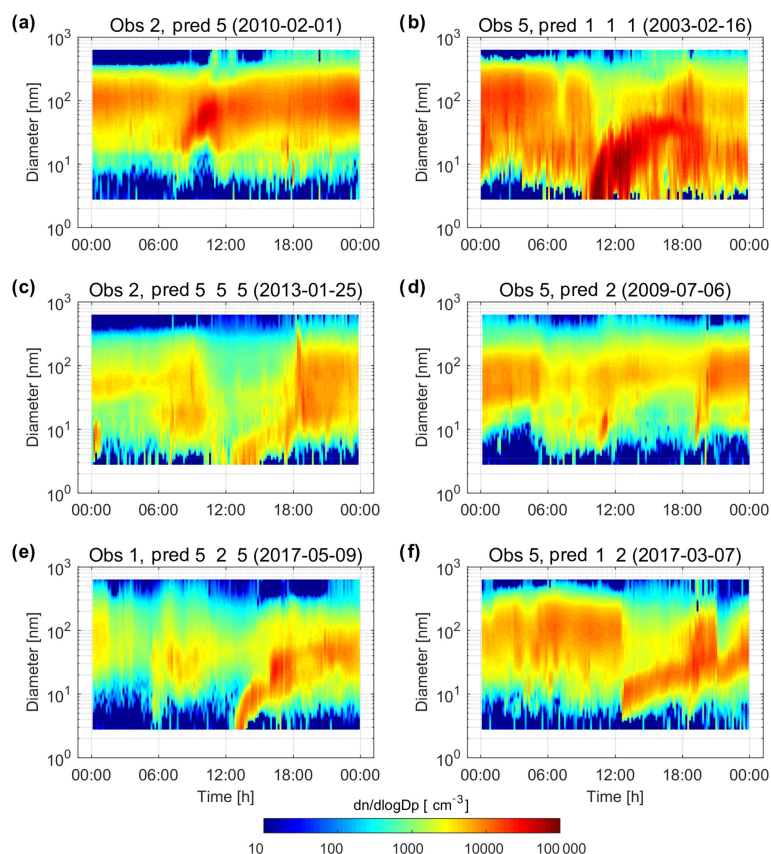
Figure 5. Examples of particle size distributions for different days when visually categorized clear event days (classes 1 and 2) are classified nonevent days by CNN classification or vice versa. In title, numbers after Obs and Pred show NPF event classes determined with visual and CNN-based classification, respectively (date of measurement in brackets). Several event class numbers in Pred cases are results from different calculation runs. The first row **(a, b)** contains examples of human-made misclassifications, the second row **(c, d)** CNN misclassifications and the third row **(e, f)** ambiguous situations (low particle concentrations and change in air masses during the day, respectively). CNN classification is from Image Set 1 with a training–testing ratio of 80 % / 20 %.

result for automatic NPF event analysis. Most of misclassified days have been categorized into the adjacent classes, which can be ambiguous to distinguish from each other. A comparison between CNN-based and human-made classification also showed that often the difference in categorization is due to a wrong or an incorrectly listed classification by a researcher. Human-made classification can easily vary from person to person and can change over time, whereas CNN-based classification is consistent at all times. The CNN-based categorization seems, at least, to be as reliable as human-made categorization and it could be even more reliable if training image sets are selected carefully. Typically, an analysis of a large size distribution dataset requires manual labor and training for several researchers, which is very time consuming, and the 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 more efficiently, accurately and consistently, which helps us to understand aerosol–climate interactions in detail.

The transfer learning of pretrained 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 pretrained CNN, a hundred images can be enough for a precise classification. The pretrained CNNs, as well as other novel machine learning and artificial intelligence methods, and the increased computing capacity due to graphics processing units (GPUs) enable us to analyze very complex and large datasets that are typical in atmospheric science.

Instead of the transfer learning of a pretrained 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, the data are 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 into 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 an unsupervised learning method can help us describe new NPF event classes or merge classes 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 people 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 algorithms are deep CNN-based *AlphaGo Zero* and *AlphaZero* developed by DeepMind (Silver et al., 2017b, a). Those algorithms achieved superhuman performance by tabula rasa reinforcement learning without human data or guidance and defeated a 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 for 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 are 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 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 a 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 the possibility of 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 stations (e.g., Puijo in Kuopio, Finland). We are including more parameters in 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).

Appendix A: General description of neural networks and training process

We used a deep feed-forward 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 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, such 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 (RNNs) are well designed for speech, text and time series classification or prediction (Pascanu et al., 2013) and convolutional NNs (CNNs) 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 providing solutions for the issue – e.g., the back-propagation 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 NNs with at least two layers and very deep NNs as those with more than 10 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 of using part of the existing learned parameters and learning only a subset of the structure. The pretraining method consists of first training the network with an UL algorithm and then continuing 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 as to 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 partially using the weights – e.g., for the first layers of the NN – and training 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 – about 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).

Appendix B: 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 e show 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 with the ZFnet (Zeiler and Fergus, 2014), then followed 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 convolutional layers (CLs, 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 ($P(l)$) to each of the classes' labels (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 convolutions of the image by several kernels (i.e., matrices used in the convolution) – that must be learnt – and generates a multidimensional output of one dimension per kernel. For instance, in Fig. 2d the two kernels $\mathbf{v}_1$ and $\mathbf{v}_2$ are applied to two compact subregions of the data (field 1 and field 2) by two units (unit 1 and unit 2), which generate two 2-D outputs $\left(\mathbf{y}_1 = \left(x_1^1, x_1^2\right)\right.$ and $\left.\mathbf{y}_2 = \left(x_2^1, x_2^2\right)\right)$. 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 256×256 pixels generated by CL will be reduced to smaller image of 64×64 by calculating maximum values of the 4×4 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 feed-forward 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 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 themselves ($\mathbf{x}_k$) 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 $\mathbf{V}$ and $\mathbf{W}$, the first layers and the last layers, respectively. The optimization problem is defined so that the cost function E reaches a minimum for some optimal sets of parameters $\hat{\mathbf{V}}$ and $\hat{\mathbf{W}}$. Once the optimal parameters are known, for the most of elements $\mathbf{x}_k \in S_g$, the output of the NN, i.e., $J\left(\mathbf{x}_k | \hat{\mathbf{V}}, \hat{\mathbf{W}}\right)$, predicts the actual label d_k, which means that the NN is ready to classify data of the same semantic field as those 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 E in Fig. 2b may be the same for the general and specialized learning step, but for the first stage, all the parameters are allowed to vary while only the $\mathbf{W}$ 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, https://doi.org/10.1029/2001JD001417, 2003.

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

Amari, S.-I.: Natural Gradient Works Efficiently in Learning, Neural Comput., 10, 251–276, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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.

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, https://doi.org/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), 1–6, 2012.

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.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 248–255, https://doi.org/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, https://doi.org/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., 10, 4235–4252, https://doi.org/10.5194/amt-10-4235-2017, 2017.

Duda, R. O., Hart, P. E., and Stork, D. G.: Pattern classification, 2nd edition, Wiley-Interscience, John Wiley & Sons, Inc., New York, 2012.

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, https://doi.org/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, https://doi.org/10.5194/acp-7-355-2007, 2007.

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, https://doi.org/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.

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 in-

fluence, and air mass transport, J. Geophys. Res.-Atmos., 120, 9459–9480, https://doi.org/10.1002/2015JD023660, 2015.

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, https://doi.org/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 High-Resolution Remote Sensing Imagery, Remote Sens., 7, 14680, https://doi.org/10.3390/rs71114680, 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, nonlinear regression and a radiative transfer-based look-up table, Atmos. Chem. Phys., 16, 8181–8191, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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-Geer, 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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1029/2004gl022092, 2005.

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, 1148–1158, https://doi.org/10.1121/1.5024341, 2018.

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, https://doi.org/10.1016/j.gsf.2015.07.003, 2016.

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

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

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, https://doi.org/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, https://doi.org/10.1007/s11356-016-7812-9, 2016.

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, https://doi.org/10.1016/j.jqsrt.2014.09.025, 2015.

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.

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 November 2017), 2017.

MathWorks: MathWorks Documentation: imresize: https://se.mathworks.com/help/images/ref/imresize.html, (last access: 14 June 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, https://doi.org/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-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, https://doi.org/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: parameterization using a multivariate mixed effects model, Geosci. Model Dev., 4, 1–13, https://doi.org/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.

Nel, A.: Air Pollution-Related Illness: Effects of Particles, Science, 308, 804–806, https://doi.org/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, https://doi.org/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., https://doi.org/10.5194/acp-2018-304, in review, 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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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.

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.

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, https://doi.org/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, https://doi.org/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, https://doi.org/10.5194/acp-11-3319-2011, 2011.

Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Networks, 61, 85–117, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7–12 June 2015, 1–9, https://doi.org/10.1109/CVPR.2015.7298594, 2015.

Twomey, S.: Pollution and the planetary albedo, Atmos. Environ., 8, 1251–1256, https://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, https://doi.org/10.5194/acp-6-4601-2006, 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, https://doi.org/10.1016/j.atmosres.2008.04.007, 2008.

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

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, No. 134, Helsinki, Finland, 653–654, 2012.

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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1186/s40537-016-0043-6, 2016.

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.

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.

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, https://doi.org/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, No. 200, Helsinki, Finland, 687–690, 2017.

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.

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, https://doi.org/10.1109/mgrs.2016.2540798, 2016.