Real-time pollen monitoring using digital holography

Eric Sauvageat[1,2*], Yanick Zeder[3,4*], Kevin Auderset[5], Bertrand Calpini[1], Bernard Clot[1], Benoît Crouzy[1], Thomas Konzelmann[1], Gian Lieberherr[1], Fiona Tummon[1], and Konstantina Vasilatou[5]

[*]Yanick Zeder and Eric Sauvageat contributed equally to this work.
[1]Federal Office of Meteorology and Climatology MeteoSwiss, Payerne, Switzerland
[2]Institute of Applied Physics and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
[3]Lucerne University of Applied Sciences and Arts, Lucerne, Switzerland
[4]Swisens AG, Horw, Switzerland
[5]Swiss Federal Institute of Metrology METAS, Bern-Wabern, Switzerland

Correspondence: Eric Sauvageat (eric.sauvageat@iap.unibe.ch)

Abstract. We present the first validation of the Swisens Poleno, currently the only operational automatic pollen monitoring system based on digital holography. The device provides in-flight images of all coarse aerosols and here we develop a two-step classification algorithm that uses these images to identify a range of pollen taxa. Deterministic criteria based on the shape of the particle are applied to initially distinguish between intact pollen grains and other coarse particulate matter. This first level of discrimination identifies pollen with an accuracy of 96 %. Thereafter, individual pollen taxa are recognised using supervised learning techniques. The algorithm is trained using data obtained by inserting known pollen types into the device and out of eight pollen taxa six can be identified with an accuracy of above 90 %. In addition to the ability to correctly identify aerosols, an automatic pollen monitoring system needs to be able to correctly determine particle concentrations. To further verify the device, controlled chamber experiments using polystyrene latex beads were performed. This provided reference aerosols with traceable particle size and number concentrations to ensure particle size and sampling volume were correctly characterised.

1 Introduction

The incidence of pollinosis and related diseases has increased considerably over the past decades, sparking growing research interest into aeroallergens and pollen monitoring. Among aeroallergens, pollen is the most important impacting approximately 20 % of the population in Switzerland and other high income countries (D'Amato et al., 2007; Wüthrich et al., 1995; Ring et al., 2001). Most often, sensitized patients exposed to allergenic pollen experience symptoms of allergic rhinitis or hay fever, but exposure to pollen has also been shown to exacerbate the development of more severe diseases like asthma, all of which have significant effects on public health and the economy (Greiner et al., 2012; Gamble et al., 2008).

Beyond the issue of public health, the airborne transport of pollen plays a key role in ecosystem dynamics, with important implications for agriculture, forestry, and the geographic dispersion of plants (Garzia-Mozo, 2011; Oteros et al., 2014). The

relevance of pollen and other bioaerosols for atmospheric chemistry and physics has also been increasingly acknowledged since they represent a significant fraction of atmospheric particulate matter and have been shown to influence cloud formation and precipitation (Jaenicke, 2005; Pöschl, 2005; Möhler et al., 2007; Deguillaume et al., 2008; Pope, 2010). Furthermore, in the context of climate change, pollen concentrations undergo fluctuations in terms of taxa, abundance, and seasonal trends. Pollen monitoring thus provides valuable information about the evolution of the local biosphere and its response to anthropogenic forcings such as pollutant emission or intensive urbanization. While still uncertain, some evidence shows that the combination of a globally warming climate and the perpetuation of contemporary human lifestyle is very likely to increase the prevalence, intensity, and related costs of pollen-related allergenic diseases in the coming decades (Ring et al., 2001; D'Amato et al., 2001, 2016; Beggs, 2016).

Airborne pollen has been monitored since the mid-twentieth century in Switzerland and elsewhere in Europe (Clot, 2003; Spieksma, 1990), most commonly with Hirst-type samplers (Hirst, 1952). These instruments continuously collect airborne particles on a rotating cylinder tape which is then collected and pollen particles manually identified and counted using optical microscopy, typically on a weekly basis. Because this is such a time- and labour-intensive method, the spatial and temporal resolution of the measurements is severely limited. Another drawback of this type of sampler is the inevitable delay between the observations and their analysis (up to 9 days), which has important implications in terms of pollen forecasts. In particular, the availability of real-time data with high temporal resolution is a key step (Sofiev, 2019) in the development of accurate forecasting models for atmospheric pollen transport (Pasken and Pietrowicz, 2005; Schueler et al., 2006; Sofiev et al., 2006; Vogel et al., 2008; Zink et al., 2013; Sofiev, 2017). More accurate predictions would represent a tremendous asset for both the scheduling of patients' activities and the planning of their medical treatment.

To respond to the need for real-time pollen information, numerous partly- or fully-automated monitoring systems have been developed and investigated over the past decade, with some recently having reached an operational level. Among the existing devices on the market, two main categories of instruments can be identified in terms of the different techniques utilised, either microscope-based or in-situ measurements (Kawashima et al., 2007; Perring et al., 2015; Oteros et al., 2015; Crouzy et al., 2016; Šauliene et al., 2019). The former aim to automatise the microscopic analysis process, while the latter make use of air-flow cytometry measurements, avoiding the collection step and performing real-time particle-by-particle identification. In the category of air-flow cytometers, most existing devices rely on fluorescence and elastic light-scattering measurements combined with machine-learning algorithms to identify and quantify airborne pollen concentrations. Some of these systems have already shown promising results and are currently tested in different European countries (Crouzy et al., 2016; Šauliene et al., 2019; Chappuis et al., 2019). Automatic pollen monitoring is part of a broader field of research on automatic bioaerosol monitoring (Kawashima et al., 2017; Calvo et al., 2018; O'Connor et al., 2015; Savage et al., 2017), that was the object of a recent review article (Huffman et al., 2019).

In this paper we evaluated a new automated pollen monitoring system based on air-flow cytometry, the Swisens Poleno. This device captures holographic images of each airborne particle in addition to measurements of optical properties such as fluorescence intensity, lifetime, and elastic light scattering. Here, we focus on the use of digital holography for online pollen monitoring since this technique allows a certain degree of visual identification of pollen taxa. We use a combination of

classical image analysis and a neural network algorithm to assess the instrument's performance in terms of pollen identification compared to manually-classified calibration sets. Aerosol sampling, particle sizing and counting performance are evaluated using a reference particle counter at the Swiss Federal Institute of Metrology, METAS (Horender et al., 2019).

In the following section, the Swisens Poleno air flow cytometer is presented as well as the methodology used for the data analysis. Thereafter, the performance achieved in pollen identification and counting using holographic images from the device is shown. Although the focus of the present paper lies on the use of digital holography to identify pollen grains, a validation of the output of the fluorescence using standard particles is also performed. Finally, the significance of the results for pollen monitoring are discussed and an overview of the future perspectives for this new technology is provided.

2 Materials and Methods

2.1 Swisens Poleno

We used the first unit of the commercially-available Poleno device developed by Swisens AG (Switzerland). The device provides in-flight measurement of particle shape, size, and fluorescence using various light sources and detectors. The schematic structure of the device is presented in Figure 1. Laser light scattering triggers the measurement, together with providing a first estimation of particle size, velocity, and alignment by combining the information of two trigger lasers. Following the trigger, two focused images at $90°$ from each other are reconstructed using digital holography as in Berg and Videen (2011), and UV-induced fluorescence produces information regarding the particle composition. UV-induced fluorescence lifetime and spectra are measured at three different excitation wavelengths (280, 365, and 405 nm) using five measurement emission windows between 320 and 720 nm. Finally, a measurement of the time-resolved optical polarization characteristics of the particle is acquired before it exits the device.

Air inlet	
Beam dump	
Particle concentrator	
Measurement trigger & holography	
Two images taken at 90° angle 200x200 resolution b/w	
Fluorescence lifetime & spectrum measurement	
Excitation wavelength [nm]: 405, 365, 280 Measurement windows [nm]: 320-380, 400-440, 450-520, 530-600, 650-750	
Polarized time-resolved scattered light measurement	
Sample collector (opt.)	
Air outlet	
Modulated laser source for fluorescence & scattered light measurement	

Figure 1. Measurement principle of the Swisens Poleno (courtesy Swisens AG).

75   Since atmospheric pollen concentrations are typically low compared to other aerosols (on the order of 10-10 000 pollen grains per cubic metre), a high sampling rate is necessary to sample pollen effectively. This is all the more relevant since the threshold for allergic response is typically even lower, varying depending on the taxa, from just a few grains to a few tens of grains per cubic metre. In the Swisens Poleno, this level of sampling is achieved using a concentrator based on the principle of a virtual impactor that enables an effective flow rate of 40 litres per minute. An hourly time resolution providing

concentrations relevant for pollen exposure thresholds can thus easily be achieved with this sampling rate. The drawback, however, is that the saturation level occurs at coarse particle (>10 μm) concentrations above 30 000 particles per cubic metre. Note that size-dependent particle loss occurs in the concentrator: Corrected concentration factors were determined from the controlled chamber experiments presented at the end of the manuscript. A Sigma2 inlet was chosen to protect the device from precipitation as its sampling, in particular the role of wind speed and particle size, has been documented in detail (Verein Deutscher Ingenieure, 2013).

2.2 Calibration dataset

A large (> 750 particles per pollen taxon) calibration dataset was collected for eight different pollen taxa using online measurements from the Swisens Poleno device. The taxa were chosen to present a good range of particle size - from small nettle pollen grains through to large pine pollen grains - and morphology. Note that the list includes taxa relevant for pollen allergies such as two different Betulaceae, *Dactylis glomerata* as a proxy for grasses, and ash. These samples were used to train a machine learning algorithm applied to identify the different pollen taxa. Only one particle type was calibrated at a time, allowing the data points to be labelled directly, although dirt, debris and agglomerates needed to be eliminated from the dataset manually through visual inspection of the holographic images. To generate a large number of events without saturating the detector, pollen samples were continuously aerosolised using sound waves in a closed chamber around the detector inlet. Figure 2 shows examples of the reconstructed images generated for the calibration dataset. The pollen identification presented here is based just on these reconstructed images since they are expected to contain enough relevant morphological information to permit sufficient identification of the taxa of interest. Fluorescence and lifetime measurements are expected to be pertinent for extending the scope of the device to characterise other bioaerosols (e.g. spores) and pollutants. The dataset obtained includes 12234 pollen grains (two images per grain) and is summarised in Table 1; 80 % of this dataset was used for algorithm calibration and 20 % for validation purposes. The images are grey scale and have a resolution of 200x200 pixels. Each pixel represents a 0.56 μm by 0.56 μm physical domain.
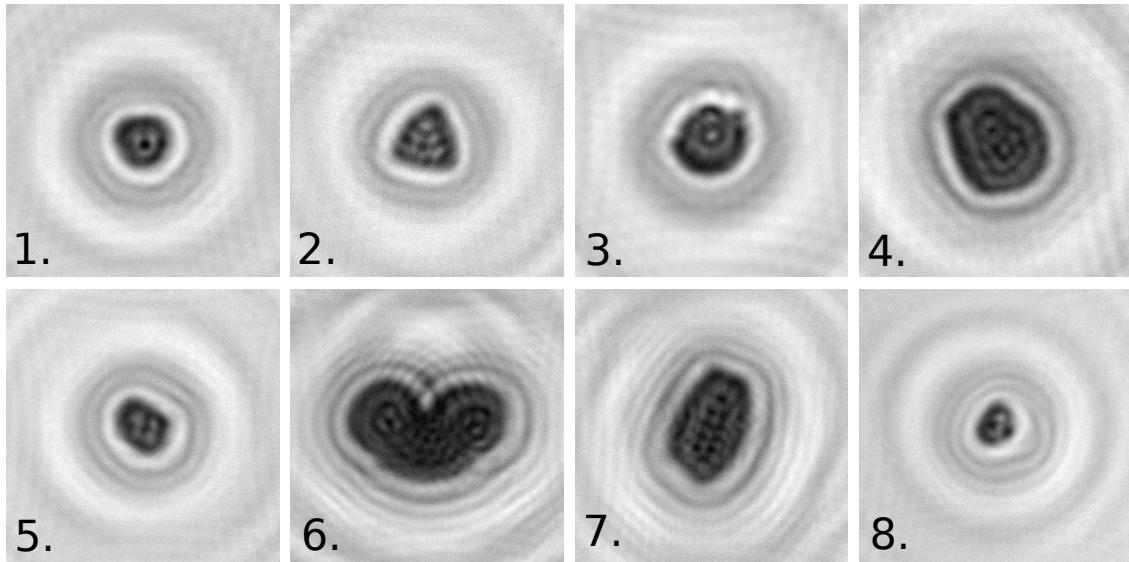
Figure 2. Reconstructed holographic images from the Swisens Poleno for different pollen taxa: 1. *Ambrosia artemisiifolia*, 2. *Corylus avellana*, 3. *Dactylis glomerata*, 4. *Fagus sylvatica*, 5. *Fraxinus excelsior*, 6. *Pinus sylvestris*, 7. *Quercus robur*, 8. *Urtica dioica*.

Table 1. List of pollen taxa used to train the classification algorithm, including the number of events used for training and validation of the machine learning algorithm for each taxa. Note that all pollen were in a dry state.

Common Name	Taxa (Latin)	Supplier	# Training events	# Validation events
Ragweed	*Ambrosia artemisiifolia*	Bonapol	1063	266
Hazel	*Corylus avellana*	Bonapol	1156	289
Grasses	*Dactylis glomerata*	Bonapol	602	151
Beech	*Fagus sylvatica*	Allergon	859	215
Ash	*Fraxinus excelsior*	Allergon	826	207
Pine	*Pinus sylvestris*	Bonapol	3601	901
Oak	*Quercus robur*	Bonapol	775	194
Nettle	*Urtica dioica*	Bonapol	903	226

2.3 Shape analysis for pollen detection

A large range of coarse aerosol particles were seen in the events recorded by the Swisens Poleno. To provide a clean dataset to the pollen classification algorithm, the pollen particles needed to be discriminated from all others. In principle this can be

done by applying thresholds to the confidence estimates provided by deep-learning pollen-classification algorithms, however, this simple method did not yield the required level of accuracy. An additional step was therefore implemented in the algorithm (thus becoming a two-step classifier), using shape analysis to discriminate between pollen and non-pollen particles prior to applying the full pollen classification.

In general, unbroken biological particles tend to have a smooth, convex shape while dust, debris, or other non-biological particles have rougher, more chaotic shapes (see, for example, Figure 3). Two deterministic image analysis routines were developed and evaluated to select the best available method for distinguishing pollen from other detected particles. Both use the contour of each particle, which is extracted from the reconstructed holographic images in three steps: (1) pixels are separated into two classes using the Otsu binarisation algorithm (Otsu, 1979), which is based on a dynamic intensity threshold; (2) the largest cluster corresponding to the particle of interest is then identified; and (3), a convolution operation extracts the contour of the particle.

The first routine for biological particle identification uses the OpenCV2 library (Bradski, 2000) to fit (in a least-squares sense) an ellipse to each contour (Fitzgibbon et al., 1999). As a feature for biological particle identification the fraction f_c of the contour located further than a certain distance from the fitted ellipse is considered (red pixels in Figure 3). For pollen grains, this value is typically low, while for more fragmented particles this fraction can reach up to 60 % of the contour (0 % and 46 % respectively for the examples shown in Figure 3).

The second method is based on the fractal dimension, which characterises the state of self-similarity or roughness, and is also estimated from the particle contour. Such analysis of natural objects was first introduced by Mandelbrot (1983), and is now widely used in a variety of applications, such as plant leaf or sediment identification (Backes et al., 2009; Orford and Whalley, 1983). We make use of the so-called box-counting algorithm to estimate the fractal dimension of the holographic images because of its computational simplicity (Theiler, 1990). This method consists of splitting each image into grid boxes and counting the number of boxes $N(s)$ that contain a fragment of the particle's perimeter, where s is the box size. By repeating the procedure for decreasing values of s, the fractal dimension is then estimated by computing the slope of $log(N(s))$ against $log(\frac{1}{s})$ (see lower panel, Figure 3).

The performance of these two methods (ellipse fitting and fractal dimension) is compared using the reference dataset which contains images from all calibrated pollen taxa (1640 particles) and manually-selected coarse aerosols (1554 non-pollen particles), measured during late spring and summer 2018 in Payerne, Switzerland. Smaller particles were filtered out prior to this comparison to keep only particles roughly in the pollen size range (10 to 200 μm). Note that only a reduced subset of pollen calibrations was used to keep a balanced dataset with respect to the non-pollen category.

Using manually-labelled data to verify the output of the classification algorithm, we performed a grid optimization to find the set of parameters that best discriminates between pollen and other airborne particles, that is, a filter that simultaneously provides sufficient recall (ensuring that most pollen grains are classified as pollen) and precision (ensuring that only pollen grains are classified as pollen).
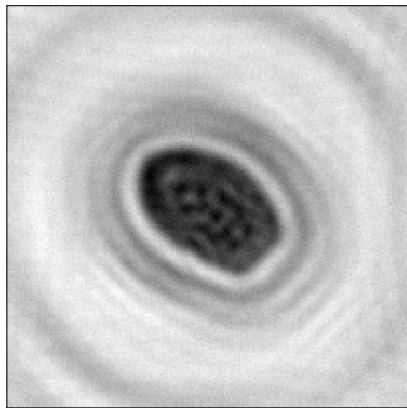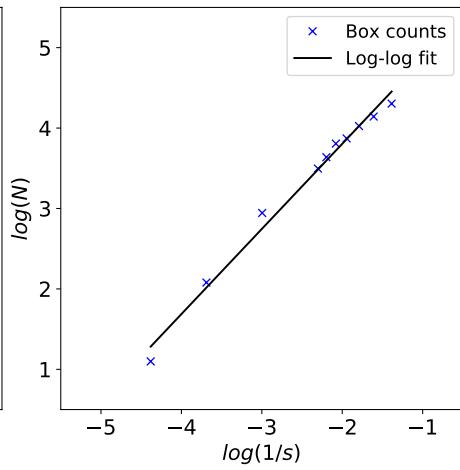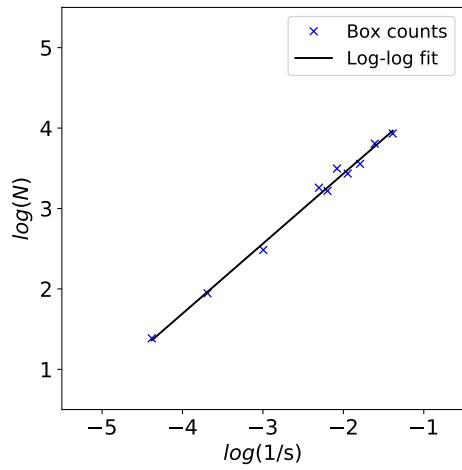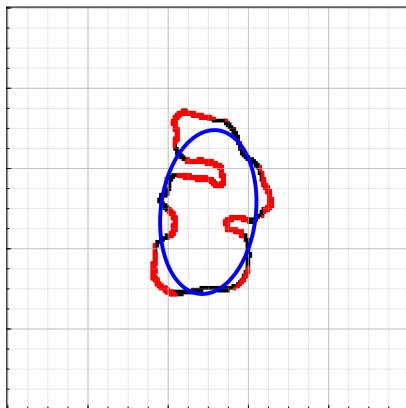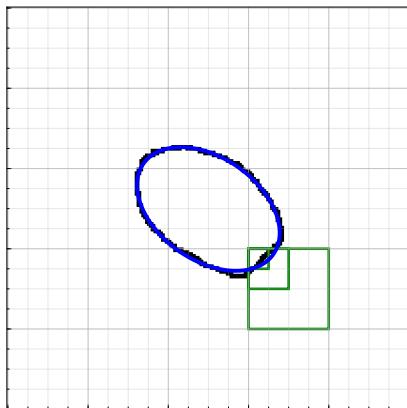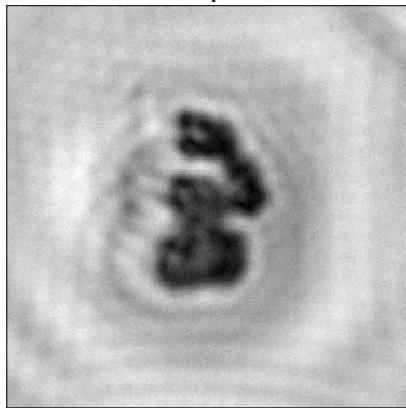
Figure 3. Illustration of the image analysis routines applied to a pollen grain (*Quercus robur*; left) and coarse particulate matter (right).

2.4 Pollen classification using deep learning

Developments in computer hardware have made it possible to perform efficient training of complex artificial, or 'deep', neural
networks; their use in image recognition problems is the iconic application of deep learning. Mimicking the visual cortex, a
series of so-called *convolutional layers* identify relevant patterns and concentrate the information diluted over a large image.
Extracted features are then used as input for fully-connected layers of artificial neurons, which combine the features to deter-
mine associated labels for each image. This technique is part of the family of supervised-learning algorithms; networks need
to be trained using images for which the label is known. We used the open source software library Keras (Chollet, 2015) with
TensorFlow (Abadi et al., 2015) as computational back-end to implement the pollen identification algorithm.
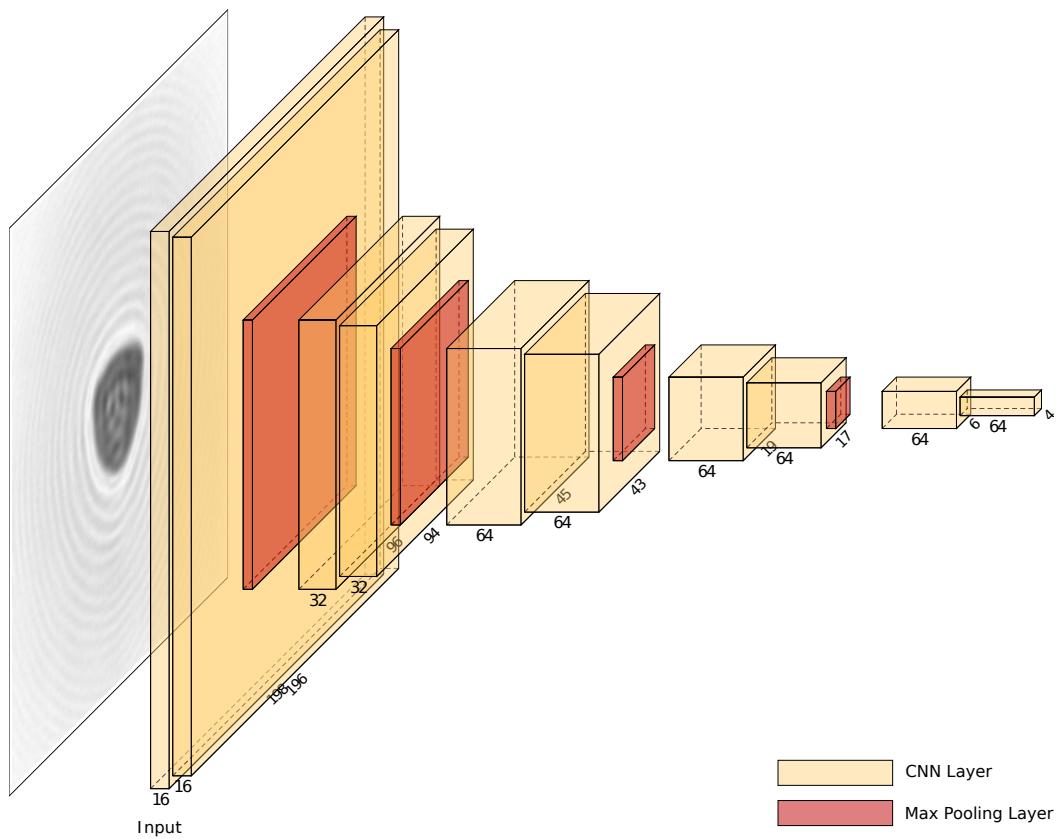
Figure 4. Vision model based on the VGG16 architecture (Simonyan and Zisserman, 2014) as used here for pollen classification.

The model used to classify pollen grains is based on the VGG16 architecture (Simonyan and Zisserman, 2014), which has
successfully been applied to a wide range of different image classification problems (Russakovsky et al., 2015). The basic
model is shown in Figure 4, with the vision model being applied separately to the two orthogonal images and the output
then being processed with two fully-connected layers. This ensures that the model is able to use the information from both

150   images. For the final layer, softmax activation is used to map the network output to a probability distribution (Chollet, 2015). The predicted pollen label is determined by taking the most probable class. Note that probability information is also useful since the plausibility of the final classifications can be easily verified (Crouzy et al., 2016; Šauliene et al., 2019). Furthermore, although not carried out here, a threshold could be applied to the classification when performing operational measurements to retain only the pollen grains classified above a sufficient confidence level.

155 ## 3   Results and Discussion

3.1   Pollen identification

The discrimination between pollen and other coarse aerosols is evaluated in Figure 5 in the form of a normalised confusion matrix for each of the two image analysis algorithms. Each line in Figure 5 is normalised to one and the values along the diagonal provide the recall for each category.

160   The convexity hypothesis for unbroken biological particles seems to hold particularly well for pollen grains. Indeed, in nearly all cases one of the two images has an almost perfect elliptical fit for pollen particles, which translates into only a very small fraction of contour pixels that strongly deviates from the optimal ellipse, i.e. $f_c \approx 0$. Other particles most often present values of $f_c \geq 20$ % for both images. Good precision can be obtained by using a low f_c threshold, albeit at the expense of recall. The best results (achieving both good precision and recall) were obtained by imposing different thresholds on the two

165   images: at least one of the two images needs to satisfy a hard condition on the value of f_c, while the other should not present excessive deviation from its fitted ellipse. When using optimal values of the two f_c thresholds, an overall accuracy of 96 % was achieved for the discrimination of pollen from other particles.

Visually, the contours of pollen grains clearly exhibit smoother shapes than non-pollen particles. However, it can be noted from Figure 5, that the fractal dimension method did not function as well as the ellipse fitting one. While an overall accuracy of

170   77 % is reached, the number of non-pollen particles classified as pollen (false positives) is still too high to ensure a satisfactory classification in the second step of the deep-learning algorithm. Note that the ellipse fitting algorithm alone performs better than any combination of the two methods (not shown here).
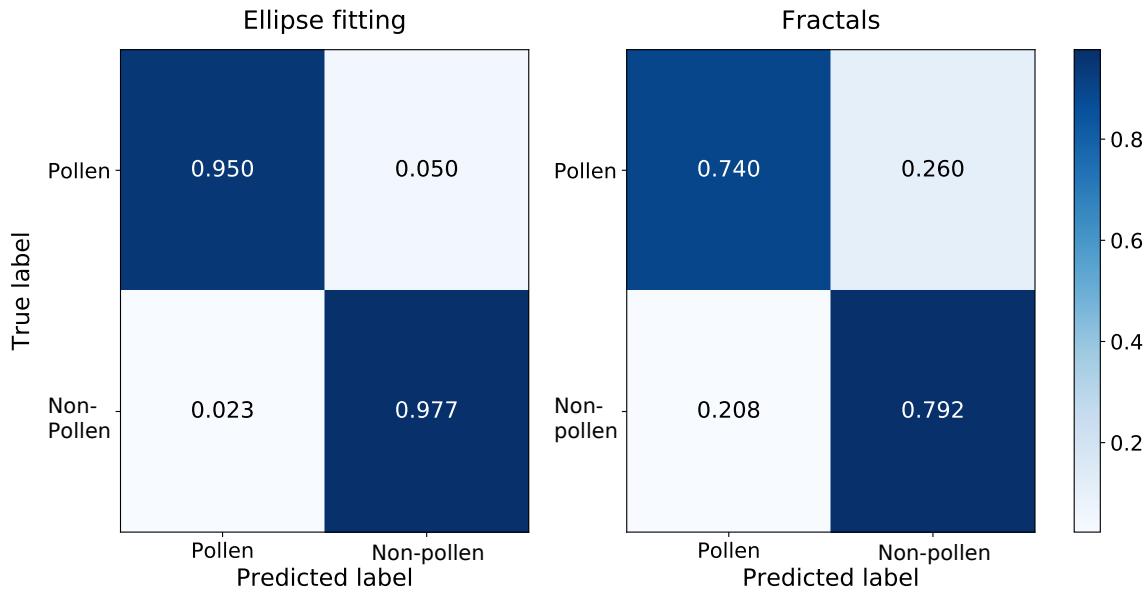
Figure 5. Normalised confusion matrix summarizing the performance of the ellipse-based classifier (left) and fractal-based classifier (right). 'Pollen' refers to a mix of different pollen taxa while 'Non-pollen' encompasses all other coarse aerosols.

Estimating the fractal dimension of an object from a holographic image is sensitive to the image resolution, which is thought to have a significant influence on the determination of the fractal dimension (see Baveye et al. (1998)). Indeed, more detailed images tend to improve the estimation of an object's fractal dimension as more details of the contour become apparent. Furthermore, the binarisation process (i.e. reducing the grey scale holographic image to black and white) may also affect the box-counting calculation. Should higher resolution images be available in future versions of the Poleno, the fractal dimension method may be worth implementing. At this point, given the better accuracy of the ellipse-fitting technique, we utilise this method in the final algorithm.

3.2 Pollen classification

Once a particle has been identified as a pollen grain it needs to be classified into the right taxa. Using the convolutional neural network (CNN) described in Section 2.4, each airborne particle is assigned a taxa with a corresponding confidence level of prediction. Results from the classification model are presented as a normalised confusion matrix in Figure 6. The sum of each line is normalised to one, and the diagonal values indicate the recall for the different pollen taxa.

Overall the classification algorithm performs very well, with 6 of the 8 pollen taxa being classified with an accuracy of over 90 %. The exceptions are *Corylus*, which is confused in 10 % of cases with *Fraxinus*, and *Dactylis*, which is confused 22 % of the time with *Corylus*. Note that in this regard the problem presented to the algorithm is somewhat artificial: *Corylus* and grass pollen are not likely to be simultaneously present in the atmosphere in concentrations relevant for pollen allergies. Nevertheless, a larger mix of pollen taxa is likely to be observed in reality, highlighting the need for further developments to the

190 classification algorithm using a larger number of species and including fresh pollen. In this line, it will be essential to include birch in the identification algorithm. This may, however, prove to be challenging given the morphological similarities of the members of the Betulaceae family.
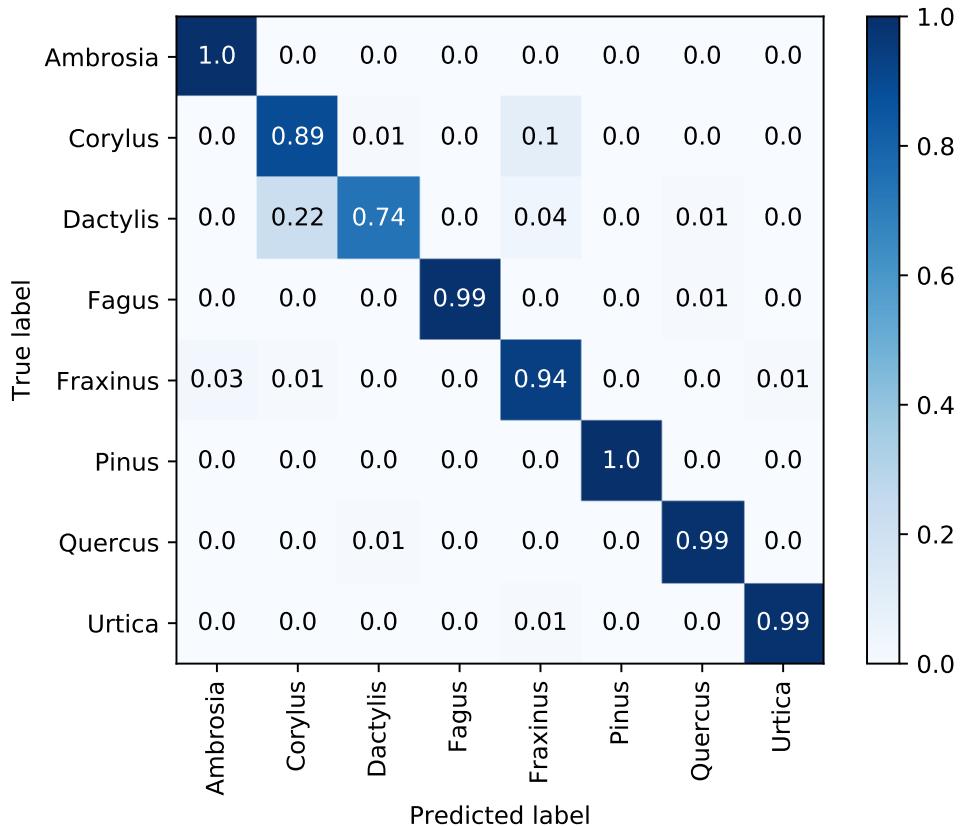
Figure 6. Normalised confusion matrix for the pollen taxa identification, the second step of the classification algorithm.

To better understand the functioning of the neural network, Figure 7 presents activation heatmaps (Kotikalapudi and contributors, 2017) of pollen particles. These show which parts of the image the network focuses on to make the taxa prediction; in
195 our case, strongly on the particle shape. This is apparent in the heatmaps (Figure 7), as the highest activation regions follow the outline of the pollen grains. This may appear as an obvious result but confirms the validity of the CNN step of the classification algorithm and indicates that the predictions are based on a physical feature of the particle and not on some other information embedded in the images. This verification is essential, as differences in light intensity, or the presence of dust on lenses, could lead to discrimination between calibrations not based on pollen morphology but on artifacts.
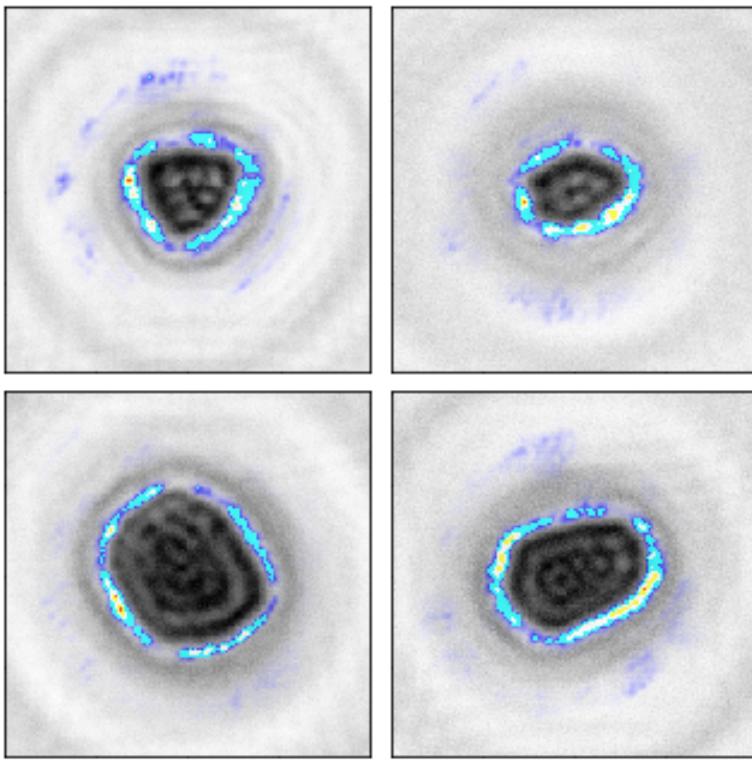
Figure 7. Visualisation of the areas on which the convolutional neural network for pollen classification focuses.

Although there are some limitations to the use of dry pollen for model training purposes, the performance obtained suggests that holography alone is sufficient to distinguish between different pollen taxon. Combined with the results of the previous section on pollen identification, we propose a two-step approach for operational pollen monitoring using digital holography, first applying classical image analysis to identify pollen, and subsequently using deep learning to classify these particles into individual pollen taxa. As mentioned in Section 2.4, the identification algorithm provides a measure of confidence in addition to the predicted label. Note that raw results are presented in the confusion matrix (Figure 6); in an operational setup confidence thresholds could be used to increase precision further. Due to the large sampling of such an automatic system, a certain loss of particles from introducing confidence thresholds can be accepted without losing statistical significance of the sampling.

3.3 Reference particle counts and fluorescence observations

The focus of this study was to assess the performance of the Swisens Poleno in terms of pollen identification. While this is key, it is equally as important to accurately quantify airborne pollen concentrations. At present, this remains a difficult task since no method, standardised or other, exists to aerosolise a known quantity of a known pollen taxa. Pollen grains are both considerably larger than other, non-biological aerosol particles and relatively fragile, so producing homogenised airborne concentrations is currently not possible with conventional techniques.

To assess the accuracy of the particle concentrations obtained with the Swisens Poleno, a measurement campaign was carried out at the Swiss Federal Institute of Metrology (METAS). The custom-made facility at METAS has been described in detail in Horender et al. (2019). The aim was to compare the Poleno device with reference particle concentrations and fluorescence observations in a controlled calibration chamber using Polystyrene Latex (PSL) spheres. Different sizes, ranging from 0.5-20 µm, were tested along with 3 types of fluorescent PSL (Blue 2.07 µm, Plum Purple 2.07 µm, and Red 2.07 µm) to provide a first insight into the quality of the fluorescence measurements. For each size, the concentrations measured by the Poleno were compared to the reference concentrations for approximately 20 minutes. The fluorescent PSL used here have been fully characterised by the Max Planck Institute for Chemistry (MPIC) (Könemann et al., 2018) for a large range of excitation wavelengths. Those corresponding to the Poleno excitation wavelengths have been reproduced in Figure 8 and serve as a reference for the fluorescence measurements. Since fluorescence intensity is measured in arbitrary units a.u., the fluorescence measured by the Poleno (filled dots) is scaled to the MPIC reference values (solid lines) using the maximum for each of the 5 emission windows located between 335 and 700 nm.
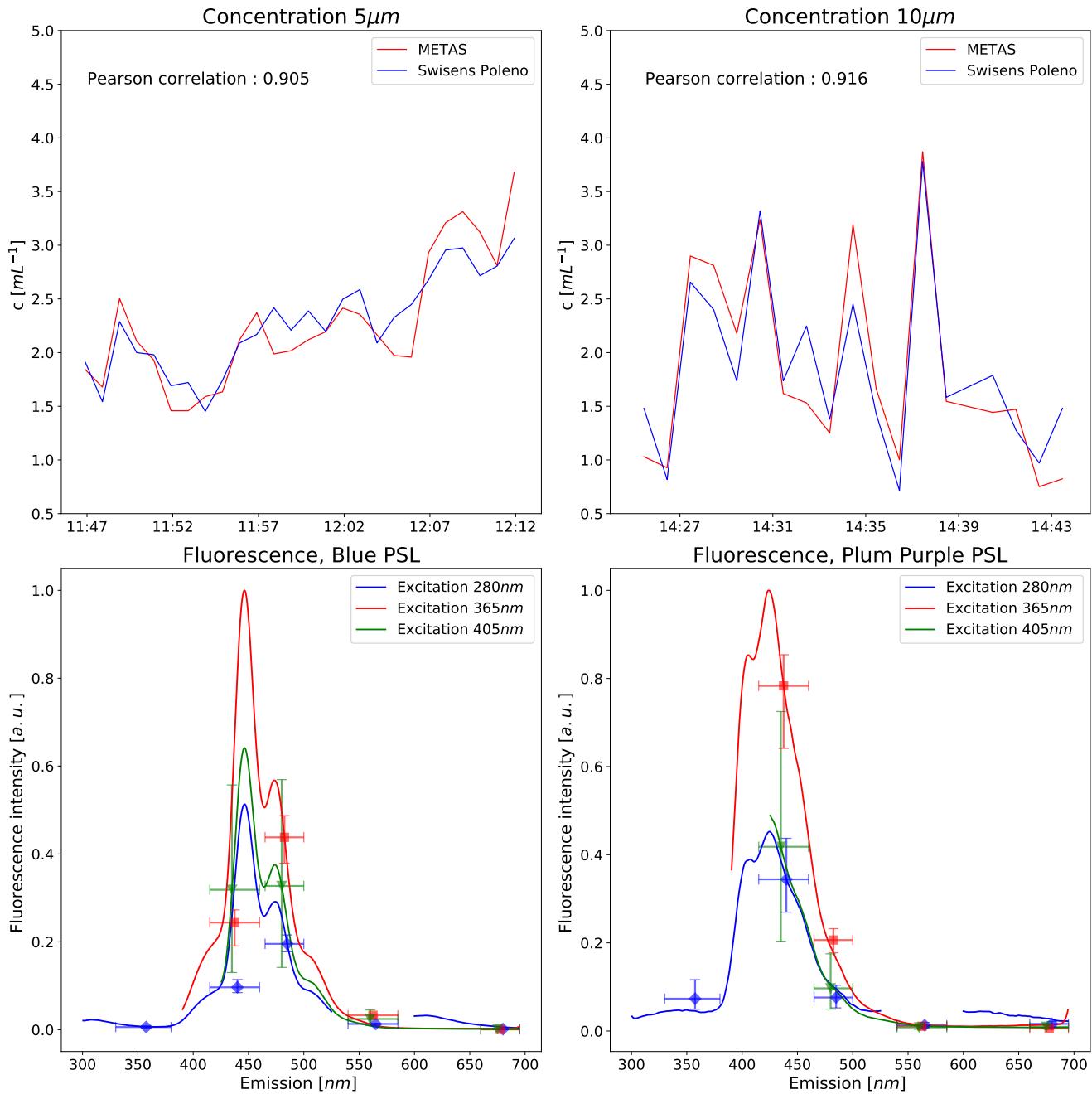
Figure 8. Top: Concentrations (5 and 10 µm PSL) scaled to the METAS reference measurements. Bottom: Comparison of fluorescence measurements. Solid lines are the reference fluorescence intensities measured by the Max Planck Institute of Chemistry presented in Könemann et al. (2018). Median measurements from the Poleno are shown with error bars (interquartile range). Each excitation wavelength is scaled individually (see text for details).

The results presented here are encouraging, both in terms of particle concentration and fluorescence measurements. The Poleno seems to follow the fluctuations in terms of particle concentration very well, with Pearson correlation values of 0.905 and 0.916 for the 5 μm and 10 μm sizes, respectively (see Figure 8). Similar results are observed for the other particle sizes tested (not shown), indicating that the Poleno measures the size of the certified PSL particles correctly. It is important to note, however, that the Poleno values have been scaled to the METAS values since the particle concentrator is size selective, with larger particles being better sampled. Once a size-scaling curve has been established it can be effectively applied to all future measurements, which is a significant improvement compared to the current practice of deriving scaling factors for automatic pollen monitors from Hirst-type measurements (Crouzy et al., 2016; Šauliene et al., 2019). The systematic analysis of the efficiency of the concentrator goes beyond the scope of this paper, but will be described in future work. The reproducibility of the scaling factors obtained was verified by repeating the experiments with the 2 μm particles three times.

Despite the fact that the Poleno does not measure a continuous fluorescence emission spectrum, Figure 8 confirms that it already provides an insight into the shape of the spectra for the different excitation wavelengths. The Poleno fluorescence signals agree well with the off-line reference measurements performed at MPIC (Könemann et al., 2018) for all 5 emission windows and combined with the holographic images, potentially provide the opportunity to extend the number of particle types that can be recognised (e.g. further pollen taxa, spores, or pollutants).

4   Towards operational pollen monitoring

The focus of this study was to assess the performance of the Swisens Poleno, the first operational automatic pollen monitoring system based on digital holography. The potential of using these in-flight images to classify pollen particles in real-time was shown for 8 pollen taxa using a two-step classification algorithm. The first step distinguishes intact pollen grains from other coarse aerosol particles using a deterministic ellipse-fitting method, providing a 96 % discrimination accuracy for pollen. Thereafter, individual pollen taxa are recognized using supervised learning techniques. The algorithm is trained using data obtained by inserting known pollen types into the device and 6 out of 8 pollen taxa can be identified with an accuracy of above 90 %.

The ability of the device to accurately count particles was tested against reference measurements in controlled chamber experiments using polystyrene latex spheres. This is a key aspect for any monitoring device that is to be used operationally and to date has not been accurately assessed. These tests, together with validation of the fluorescence measurements carried out in the same chamber, provide very promising results.

The holographic images open the possibility for a human expert to perform online training and improve the model through a feedback loop. This effectively means that falsely-classified pollen are identified manually and put into the correct class, for the model to use in the next training phase. The same principle could potentially be applied when the device is deployed in a new region with different pollen taxa by creating new pollen classes. Since the Swisens Poleno measures one cubic metre of air every 25 minutes, such new datasets can be created relatively quickly.

Finally, while not included in this study, the use of the fluorescence observations may provide the potential to identify particles other than pollen, for example, spores or other pollutants. Although the use of holography is a clear novelty of the present work, development of the method to additionally include florescence would build upon pioneering work performed using other devices (Toprak and Schnaiter, 2013; Hernandez et al., 2016; Savage et al., 2017). This could lead to synergies with air pollution monitoring networks and be of significant benefit to other sectors, such as agriculture and forestry, where real-time information concerning the distribution of spores could lead to better crop management practices. Future work in this direction is being continued, as is the development of the machine learning algorithm to identify further pollen taxa.

Code and data availability. All data and algorithms presented in the paper are experimental and subject to further development. They are available for research purposes on request to the authors of the paper. Work is in progress to harmonize the algorithms and make them public together with the data via open software and data repositories.

Author contributions. BCl, BCr, ES and YZ designed the study. BCr, KV, KA, and FT carried out the METAS campaign. ES and YZ analysed all available data. ES, YZ, BCr, and FT prepared the paper with contributions from all other authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. METAS has received funding from the EMPIR Projects 16ENV07-Aeromet and 19ENV08-Aeromet II. The EMPIR programme is co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme. This work also contributes to the EUMETNET AutoPollen Programme.

References

275 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Good-fellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, http://tensorflow.org/, software available from tensorflow.org, 2015.

280 Backes, A. R., Casanova, D., and Bruno, O. M.: Plant leaf identification based on volumetric fractal dimension, International Journal of Pattern Recognition and Artificial Intelligence, 23, 1145–1160, 2009.

Baveye, P., Boast, C. W., Ogawa, S., Parlange, J.-Y., and Steenhuis, T.: Influence of image resolution and thresholding on the apparent mass fractal characteristics of preferential flow patterns in field soils, Water resources research, 34, 2783–2796, 1998.

Beggs, P. J.: Impacts of Climate Change on Allergens and Allergic Diseases, Cambridge University Press, 2016.

285 Berg, M. J. and Videen, G.: Digital holographic imaging of aerosol particles in flight, Journal of Quantitative Spectroscopy and Radiative Transfer, 112, 1776 – 1783, https://doi.org/https://doi.org/10.1016/j.jqsrt.2011.01.013, http://www.sciencedirect.com/science/article/pii/S0022407311000380, electromagnetic and Light Scattering by Nonspherical Particles XII, 2011.

Bradski, G.: The OpenCV Library, Dr. Dobb's Journal of Software Tools, 2000.

Calvo, A., Baumgardner, D., Castro, A., Fernández-González, D., Vega-Maray, A., Valencia-Barrera, R., Oduber, F., Blanco-
290 Alegre, C., and Fraile, R.: Daily behavior of urban Fluorescing Aerosol Particles in northwest Spain, Atmospheric Environ-ment, 184, 262 – 277, https://doi.org/https://doi.org/10.1016/j.atmosenv.2018.04.027, http://www.sciencedirect.com/science/article/pii/S1352231018302589, 2018.

Chappuis, C. M., Tummon, F., Clot, B., Konzelmann, T., Calpini, B., and Crouzy, B.: Automatic pollen monitoring: first insights from hourly data, Aerobiologia, pp. 1 – 12, 2019.

295 Chollet, F.: Keras, https://github.com/fchollet/keras, 2015.

Clot, B.: Trends in airborne pollen: an overview of 21 years of data in Neuchâtel (Switzerland), Aerobiologia, 19, 227–234, 2003.

Crouzy, B., M., S., Konzelmann, T., Calpini, B., and Clot, B.: All-optical automatic pollen indentification: Towards an operational system, Atmospheric Environment, 140, 202 – 212, 2016.

D'Amato, G., Liccardi, G., D'Amato, M., and Cazzola, M.: The role of outdoor air pollution and climatic changes on the rising trends in res-
300 piratory allergy, Respiratory Medicine, 95, 606 – 611, https://doi.org/http://dx.doi.org/10.1053/rmed.2001.1112, http://www.sciencedirect.com/science/article/pii/S0954611101911126, 2001.

D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Annesi-Maesano, I., Behrendt, H., Liccardi, G., Popov, T., and Van Cauwenberge, P.: Allergenic pollen and pollen allergy in Europe, Allergy, 62, 976–990, https://doi.org/10.1111/j.1398-9995.2007.01393.x, http://dx.doi.org/10.1111/j.1398-9995.2007.01393.x, 2007.

305 D'Amato, G., Pawankar, R., Vitale, C., Lanza, M., Molino, A., Stanziola, A., Sanduzzi, A., Vatrella, A., and D'Amato, M.: Climate change and air pollution: effects on respiratory allergy, Allergy, asthma & immunology research, 8, 391–395, 2016.

Deguillaume, L., Leriche, M., Amato, P., Ariya, P., Delort, A., Pöschl, U., Chaumerliac, N., Bauer, H., Flossmann, A., and Morris, C.: Microbiology and atmospheric processes: chemical interactions of primary biological aerosols, Biogeosciences Discussions, 5, 841–870, 2008.

310   Fitzgibbon, A., Pilu, M., and Fisher, R. B.: Direct least square fitting of ellipses, IEEE Transactions on pattern analysis and machine intelligence, 21, 476–480, 1999.

Gamble, J. L., Reid, C., Post, E., and Sacks, J.: A review of the impacts of climate variability and change on aeroallergens and their associated effects, Washington: Global Change Research Program, 2008.

Garzia-Mozo, H.: The use of aerobiological data on agronomical studies, Annals of Agricultural and Environmental Medicine, 18, 2011.

315   Greiner, A. N., Hellings, P. W., Rotiroti, G., and Scadding, G. K.: Allergic rhinitis, The Lancet, 378, 2112 – 2122, https://doi.org/http://dx.doi.org/10.1016/S0140-6736(11)60130-X, http://www.sciencedirect.com/science/article/pii/S014067361160130X, 2012.

Hernandez, M., Perring, A. E., McCabe, K., Kok, G., Granger, G., and Baumgardner, D.: Chamber catalogues of optical and fluorescent signatures distinguish bioaerosol classes, Atmospheric Measurement Techniques, 9, 3283–3292, https://doi.org/10.5194/amt-9-3283-2016,
320   https://www.atmos-meas-tech.net/9/3283/2016/, 2016.

Hirst, J. M.: An automatic volumetric spore trap, Annals of Applied Biology, 39, 257–265, https://doi.org/10.1111/j.1744-7348.1952.tb00904.x, http://dx.doi.org/10.1111/j.1744-7348.1952.tb00904.x, 1952.

Horender, S., Auderset, K., and Vasilatou, K.: Facility for calibration of optical and condensation particle counters based on a turbulent aerosol mixing tube and a reference optical particle counter, Review of Scientific Instruments, 90, 075 111, https://doi.org/10.1063/1.5095853,
325   https://doi.org/10.1063/1.5095853, 2019.

Huffman, J. A., Perring, A. E., Savage, N. J., Clot, B., Crouzy, B., Tummon, F., Shoshanim, O., Damit, B., Schneider, J., Sivaprakasam, V., Zawadowicz, M. A., Crawford, I., Gallagher, M., Topping, D., Doughty, D. C., Hill, S. C., and Pan, Y.: Real-time sensing of bioaerosols: Review and current perspectives, Aerosol Science and Technology, 0, 1–31, https://doi.org/10.1080/02786826.2019.1664724, https://doi.org/10.1080/02786826.2019.1664724, 2019.

330   Jaenicke, R.: Abundance of cellular material and proteins in the atmosphere, Science, 308, 73–73, 2005.

Kawashima, S., Clot, B., Fujita, T., Takahashi, Y., and Nakamura, K.: An algorithm and a device for counting airborne pollen automatically using laser optics, Atmospheric Environment, 41, 7987 – 7993, https://doi.org/http://dx.doi.org/10.1016/j.atmosenv.2007.09.019, http://www.sciencedirect.com/science/article/pii/S1352231007008217, 2007.

Kawashima, S., Thibaudon, M., Matsuda, S., Fujita, T., Lemonis, N., Clot, B., and Oliver, G.: Automated pollen monitoring system using
335   laser optics for observing seasonal changes in the concentration of total airborne pollen, Aerobiologia, 33, https://doi.org/10.1007/s10453-017-9474-6, 2017.

Könemann, T., Savage, N. J., Huffman, J. A., and Pöhlker, C.: Characterization of steady-state fluorescence properties of polystyrene latex spheres using off-and online spectroscopic methods., Atmospheric Measurement Techniques, 11, 2018.

Könemann, T., Savage, N. J., Huffman, J. A., and Pöhlker, C.: Characterization of steady-state fluorescence properties of polystyrene latex
340   spheres using off- and online spectroscopic methods, Atmospheric Measurement Techniques, 11, 3987–4003, https://doi.org/10.5194/amt-11-3987-2018, https://www.atmos-meas-tech.net/11/3987/2018/, 2018.

Kotikalapudi, R. and contributors: keras-vis, https://github.com/raghakot/keras-vis, 2017.

Mandelbrot, B. B.: The fractal geometry of nature, vol. 173, WH freeman New York, 1983.

345   Möhler, O., DeMott, P., Vali, G., and Levin, Z.: Microbiology and atmospheric processes: the role of biological particles in cloud physics, Biogeosciences Discussions, 4, 2559–2591, 2007.

Orford, J. and Whalley, W.: The use of the fractal dimension to quantify the morphology of irregular-shaped particles, Sedimentology, 30, 655–668, 1983.

Oteros, J., Orlandi, F., García-Mozo, H., Aguilera, F., Dhiab, A. B., Bonofiglio, T., Abichou, M., Ruiz-Valenzuela, L., del Trigo, M. M.,
350   Díaz de la Guardia, C., Domínguez-Vilches, E., Msallem, M., Fornaciari, M., and Galán, C.: Better prediction of Mediterranean olive production using pollen-based models, Agronomy for Sustainable Development, 34, 685–694, https://doi.org/10.1007/s13593-013-0198-x, https://doi.org/10.1007/s13593-013-0198-x, 2014.

Oteros, J., Pusch, G., Weichenmeier, I., Heimann, U., Möller, R., Röseler, S., Traidl-Hoffmann, C., Schmidt-Weber, C., and Buters, J. T. M.: Automatic and Online Pollen Monitoring, International Archives of Allergy and Immunology, 167, 158–166, http://www.karger.com/DOI/
355   10.1159/000436968, 2015.

Otsu, N.: A threshold selection method from gray-level histograms, IEEE transactions on systems, man, and cybernetics, 9, 62–66, 1979.

O'Connor, D. J., Daly, S. M., and Sodeau, J. R.: On-line monitoring of airborne bioaerosols released from a composting/green waste site, Waste Management, 42, 23 – 30, https://doi.org/https://doi.org/10.1016/j.wasman.2015.04.015, http://www.sciencedirect.com/science/article/pii/S0956053X15002871, 2015.

360 Pasken, R. and Pietrowicz, J. A.: Using dispersion and mesoscale meteorological models to forecast pollen concentrations, Atmospheric Environment, 39, 7689 – 7701, https://doi.org/https://doi.org/10.1016/j.atmosenv.2005.04.043, http://www.sciencedirect.com/science/article/pii/S1352231005007806, 2005.

Perring, A. E., Schwarz, J. P., Baumgardner, D., Hernandez, M. T., Spracklen, D. V., Heald, C. L., Gao, R. S., Kok, G., McMeeking, G. R., McQuaid, J. B., and Fahey, D. W.: Airborne observations of regional variation in fluorescent aerosol across the United States, Journal of
365   Geophysical Research: Atmospheres, 120, 1153–1170, https://doi.org/10.1002/2014JD022495, http://dx.doi.org/10.1002/2014JD022495, 2014JD022495, 2015.

Pope, F.: Pollen grains are efficient cloud condensation nuclei, Environmental Research Letters, 5, 044 015, 2010.

Pöschl, U.: Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects, Angewandte Chemie International Edition, 44, 7520–7540, https://doi.org/10.1002/anie.200501122, http://dx.doi.org/10.1002/anie.200501122, 2005.

370 Ring, J., Krämer, U., Schäfer, T., and Behrent, H.: Why are allergies increasing?, Current Opinion in Immunology, 13, 701 – 708, https://doi.org/http://dx.doi.org/10.1016/S0952-7915(01)00282-5, http://www.sciencedirect.com/science/article/pii/S0952791501002825, 2001.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision (IJCV), 115, 211–252,
375   https://doi.org/10.1007/s11263-015-0816-y, 2015.

Savage, N. J., Krentz, C. E., Könemann, T., Han, T. T., Mainelis, G., Pöhlker, C., and Huffman, J. A.: Systematic characterization and fluorescence threshold strategies for the wideband integrated bioaerosol sensor (WIBS) using size-resolved biological and interfering particles, Atmospheric Measurement Techniques, 10, 4279–4302, https://doi.org/10.5194/amt-10-4279-2017, https://www.atmos-meas-tech.net/10/4279/2017/, 2017.

380 Savage, N. J., Krentz, C. E., Koenemann, T., Han, T. T., Mainelis, G., Poehlker, C., and Huffman, J. A.: Systematic characterization and fluorescence threshold strategies for the wideband integrated bioaerosol sensor (WIBS) using size-resolved biological and interfering particles, ATMOSPHERIC MEASUREMENT TECHNIQUES, 10, 4279–4302, https://doi.org/10.5194/amt-10-4279-2017, 2017.

Schueler, S., Schlünzen, and Heinke, K.: Modeling of oak pollen dispersal on the landscape level with a mesoscale atmospheric model, Environmental Modeling & Assessment, 11, 179, https://doi.org/10.1007/s10666-006-9044-8, https://doi.org/10.1007/s10666-006-9044-8, 2006.

Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv 1409.1556, 2014.

Sofiev, M.: On impact of transport conditions on variability of the seasonal pollen index, Aerobiologia, 33, 167–179, https://doi.org/10.1007/s10453-016-9459-x, https://doi.org/10.1007/s10453-016-9459-x, 2017.

Sofiev, M.: On possibilities of assimilation of near-real-time pollen data by atmospheric composition models, Aerobiologia, 35, 523–531, https://doi.org/10.1007/s10453-019-09583-1, https://doi.org/10.1007/s10453-019-09583-1, 2019.

Sofiev, M., Siljamo, P., Ranta, H., and Rantio-Lehtimäki, A.: Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study, International Journal of Biometeorology, 50, 392, https://doi.org/10.1007/s00484-006-0027-x, https://doi.org/10.1007/s00484-006-0027-x, 2006.

Spieksma, F. T. M.: Pollinosis in Europe: new observations and developments, Review of Palaeobotany and Palynology, 64, 35–40, 1990.

Theiler, J.: Estimating fractal dimension, JOSA A, 7, 1055–1073, 1990.

Toprak, E. and Schnaiter, M.: Fluorescent biological aerosol particles measured with the Waveband Integrated Bioaerosol Sensor WIBS-4: laboratory tests combined with a one year field study, Atmospheric Chemistry and Physics, 13, 225–243, https://doi.org/10.5194/acp-13-225-2013, https://www.atmos-chem-phys.net/13/225/2013/, 2013.

Verein Deutscher Ingenieure: VDI 2119, techreport, 2013.

Vogel, H., Pauling, A., and Vogel, B.: Numerical simulation of birch pollen dispersion with an operational weather forecast system, International Journal of Biometeorology, 52, 805–814, https://doi.org/10.1007/s00484-008-0174-3, https://doi.org/10.1007/s00484-008-0174-3, 2008.

Šauliene, I., Šukiene, L., Daunys, G., Valiulis, G., Vaitkevičius, L., Matavulj, P., Brdar, S., Panic, M., Sikoparija, B., Clot, B., Crouzy, B., and Sofiev, M.: Automatic pollen recognition with the Rapid-E particle counter: the first-level procedure, experience and next steps, Atmospheric Measurement Techniques, 12, 3435–3452, https://doi.org/10.5194/amt-12-3435-2019, https://www.atmos-meas-tech.net/12/3435/2019/, 2019.

Wüthrich, B., Schindler, C., Leuenberger, P., and Ackermann-Liebrich, U.: Prevalence of atopy and pollinosis in the adult population of Switzerland (SAPALDIA study), International archives of allergy and immunology, 106, 149–156, 1995.

Zink, K., Pauling, A., Rotach, M. W., Vogel, H., Kaufmann, P., and Clot, B.: EMPOL 1.0: a new parameterization of pollen emission in numerical weather prediction models, Geoscientific Model Development, 6, 1961–1975, https://doi.org/10.5194/gmd-6-1961-2013, https://www.geosci-model-dev.net/6/1961/2013/, 2013.