Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

Atmospheric Measurement Techniques Discussions

Improved real-time bio-aerosol classification using Artificial Neural Networks

Maciej Leśkiewicz[1], *Miron Kaliszewski[2], Maksymilian Włodarski[2], Jarosław Młyńczak[2], Zygmunt Mierczyk[2], Krzysztof Kopczyński[2].

1. PCO S.A. ul. Jana Nowaka-Jeziorańskiego 28, 03-982 Warsaw, Poland.
2. Institute of Optoelectronics, Military University of Technology, ul. Gen. Witolda Urbanowicza 2, 00-908 Warsaw, Poland

*Corresponding author: miron.kaliszewski@wat.edu.pl

Keywords: Bio-aerosol, Fluorescence, Real-time analysis, Artificial Neural Network, PBAP.

1. Abstract

Air contamination has had stronger and stronger impact on everyday life of humans. An increasing number of people are aware of the health problems that may result from inhaling air containing dust, bacteria, pollens or fungi. Society is awaiting anxiously for a system that could inform them in real-time about a real danger that is suspended in the air. The devices, currently available on the market, are able to detect some particles in the air, but cannot classify them by the health threats. Fortunately, a new type of technology is emerging as a really promising solution. Laser based bio-detectors are opening a new era in aerosol research. They are capable of characterizing a great number of individual particles in seconds by analyzing optical scattering and fluorescence characteristics. In this study we demonstrate application of Artificial Neural Network (ANN) to real-time analysis of single particle fluorescence fingerprints. We gathered a total of 114779 spectra of 48 aerosols. We discuss an entirely new approach to data analysis using decision tree comprising 22 independent neural networks. Applying confusion matrices and ROC analysis the best sets of ANN's for each group of similar aerosols has been determined. As a result we achieved very high performance of aerosol classification in real-time. We found that for some substances that have characteristic spectra almost each particle can be properly classified. The aerosols with similar spectral characteristics can be classified as a specific cloud with high probability.

2. Introduction

The ambient air contains a variety of particles like dust, bacteria, pollens, fungi and other parts of biological and non-biological origin(Pöhlker et al., 2013)(Górny, 2004). The aerosols are involved in various atmospheric processes like ice nuclei formation, precipitation and global climate effects (Deguillaume et al., 2008) (Fröhlich-Nowoisky et al., 2016) (Gabey et al., 2010)-(Pósfai and Buseck, 2010) (Fuzzi et al., 2015). They also strongly influence human health (Davidson et al., 2005) (Pope III and Dockery, 2006) (Michaels, 2017) (Shiraiwa et al., 2012). Therefore, the characterization of ambient air is important for estimating potential health hazards and environmental impact (Mauderly and Chow, 2008) (Lim et al., 2005). Standard methods of aerosol composition assessment usually include microscopic inspection or molecular analysis of filter (Miaskiewicz-Peska and Lebkowska, 2012), tape or liquid trapped particles. Nevertheless, they suffer from low time resolution due to periodical and relatively long analytical procedures. They are also ineffective for the detection of non-culturable microorganisms (Blais-Lecours et al., 2015) (Trafny et al., 2014).

The detection and classification of biological particles is possible using fluorescence techniques

due to the presence of proteins, NADH, and some vitamins that emit light when excited with UV light (Lakowicz, 1999). This feature is utilized in single particle fluorescence detectors. In the flowing air each particle is characterized for size/shape using light scattering as well as fluorescence properties. This approach ensures continuous measurement and immediate response. Thus the analysis process can be facilitated and accelerated compared with other commonly used analytical procedures (Hill et al., 1999) (Choi et al., 2014) (Taketani et al., 2013)(Feugnet et al., 2008).

Several studies using single particle fluorescence detectors demonstrated that fluctuations of aerosol concentration and variations in its fluorescence properties are strongly dependent on the season, day time, location and a place occupancy (Gabey et al., 2011) (Huffman et al., 2010) (Pinnick et al., 2004) (Bhangar et al., 2014) (Fennelly et al., 2018). Each single particle passing the instrument is labelled with the time, scattering properties (size and/or shape) and fluorescence characteristics. It is obvious that continuous single particle measurements bring a new potential and quality to environmental research. However, particles of the same type and batch display slightly different spectral characteristics due to variations in biochemical composition, size, age in a population (Agranovski et al., 2003), degradation or stress level (Lee et al., 2010) and the particle position within instrument's interrogation point (Pan et al., 2011). The simple statistics, like data averaging and graphical spectra representation, are not sufficient. Therefore, the huge amount of data and occurring spectral variations require more advanced algorithms supporting automatic data classification. Various analytical methods of particle discrimination and classification were applied. It has been shown that Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Hierarchical cluster Analysis (HCA) of fluorescence spectra strongly increases discrimination of particles compared with methods based on spectra averaging or fluorescence threshold (Leśkiewicz et al., 2016)(Kaliszewski et al., 2013) (Pan et al., 2012) (Hernandez et al., 2016). Artificial neural network (ANN) is an emerging analytical approach that becomes more widely and successfully applied in various life domains like chemical analysis (Borecki et al., 2008), image recognition (Antowiak and Chałasińska-Macukow, 2003), data mining and weather forecasting (Purnomo et al., 2017). It has been shown that ANN can be applied in bio-aerosol classification (Kohlus and Bottlinger, 1993). However, it usually requires more user input comparing to other analytical procedures (Ruske et al., 2017).

This paper focuses on the application of ANN for real time discrimination of bio-aerosols basing on single particle fluorescence characteristics. We demonstrated a new approach to data analysis using ANN allowing automatization of data preparation procedures and minimum user involvement.

3. Materials and methods

3.1. Experiment

3.1.1. BioAeRosol Detector (BARDet)

The detailed information concerning construction and parameters of the instrument used for the experiments was presented in our previous work (Kaliszewski et al., 2016). In general, the ambient air is continuously drawn through the nozzle. It is focused with sheath flow of filtered air. Particles in the focused air pass through the BARDet's chamber where they are interrogated by a 16mW CW laser beam generated by a diode laser operating at 375 nm wavelength (CUBE, Coherent). The backward and forward scattered signals are detected with two PMT's (H6780, Hamamatsu) mounted at the 35°and 145° angle to the laser beam axis.

The fluorescence of particles is measured at a 90° angle to the laser beam with 32 channel PMT (A10766, Hamamatsu). The longpass filter with cutting edge at 400 nm (Edmund Optics) separates

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

Atmospheric Measurement Techniques Discussions

the fluorescence signal from scattered light. The multichannel PMT measures fluorescence in 18 active channels in the range of 415.4-643.5 nm. The channels are grouped in 7 bands. The remaining channels are not used. The band configuration is presented in Table 1.

Table 1. Configuration of bands in the multichannel PMT.

BARDet's Fluorescence Bands	Bandwidth [nm]
B1	415.4 – 429.3
B2	443.1 – 456.8
B3	470.5 – 484.2
B4	497.8 – 524.9
B5	538.3 – 565.0
B6	578.3 – 604.6
B7	617.6 – 643.5

3.1.2. Biological Aerosols

For the tests, dry powders of harmless substances were used, since they did not need a specialized aerosol protection chamber. The samples used for this study are listed in Table 2. To perform numerous experiments, disposable vials were used, one for each aerosol sample. It prevented cross contamination between measured samples. The aerosols were generated from modified 50 ml Falcon tubes placed on the vortex. The vials in the lower part contained two connectors for silicon tubes. Vortexed particles were entrained and formed an aerosol cloud inside the Falcon tube. The aerosolized particles were aspirated from the vial to BARDet's aerosol inlet. Each tube contained about 50 mg of the dry powder sample. During aerosol generation filtered air was supplied into the vial to compensate the BARDet's flow. The concentration of the aerosols was adjusted with vibration frequency of vortex. The measurement started after the aerosol reached homogeneous concentration. The experimental setup is shown in figure 1.

Table 2. List of all substances used in experiment.

	Abbreviation	Name	Source	Group
1	FM7	Fluoromax microspheres 7 um	Thermo scientific	standard 1
2	Rib	Riboflavin	Sigma-Aldrich	standard 2
3	BGP	Bermuda grass pollen	Duke Sci. Corp.	pollen
4	CP	Corn pollen	Duke Sci. Corp.	
5	CA	*Corylus avellana* pollen	Own collectionon	
6	LP	*Lycopodium* pollen	Fluka	
7	PPP	*Poa pratrensis* pollen	Sigma-Aldrich	
8	RP	Ragweed pollen	Duke Sci. Corp.	
9	SCP	*Secale cereale* pollen	Sigma-Aldrich	

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

Atmospheric Measurement Techniques Discussions

10	SP	Spruce pollen	Own collection	
11	AA	*Abies alba* pollen	Own collection	
12	UDP	*Urtica dioica* pollen	Own collection	
13	PSP	*Pinus sylvestris* pollen	Own collection	
14	PNP	*Pinus nigra* pollen	Own collection	
15	LPP	*Lycopodium* pollen (Poland)	Own collection	
16	PMP	Paper mulberry pollen	Duke Sci. Corp.	
17	ATP	*Artemisia tridentata* pollen	Sigma-Aldrich	
18	AAP	*Artemisia absynthium* pollen	Sigma-Aldrich	
19	CPP	*Chenopodium* pollen	Own collection	
20	BWF	Buck wheat flour	Regular shop	flour
21	PF	Potato flour	Regular shop	
22	RF	Rice flour	Regular shop	
23	TF	Tapioca flour	Regular shop	
24	WF	Wheat flour	Regular shop	
25	Trp	Tryptophan	Sigma-Aldrich	amino acids and proteins
26	Phe	Phenylalanine	Sigma-Aldrich	
27	BSA	Bovine Serum Albumin	POCH Poland	
28	OVA	Ovalbumin	POCH Poland	
29	Ambio	*Bif. animalis, S. boulardii, S. thermophilus, L. casei, L. bulgaricus*	Pharmacy	bacteria in medium
30	LCB	*Lactobacillus bulgaricus*	Pharmacy	
31	LF	*Bifidobacterium animalis, L. acidophilus*	Pharmacy	
32	BA	Bacteriological Agar	Sigma-Aldrich	medium
33	BAB	Blood Agar Base	Sigma-Aldrich	
34	LB	Luria broth	Sigma-Aldrich	
35	NB	Nutrient broth	Sigma-Aldrich	
36	BTSTG	*Bacillus thuringiensis* spores technical grade	Agricultural	Bacterial spore with admixtures
37	SB	*Saccharomyces boulardii*	Pharmacy	funghi with admixtures
38	SC	*Saccharomyces cerevisiae*	Regular shop	
39	LS	*Lycoperdon* spores	Own collection	fungal spores
40	JGSS	Johnsons grass smut spores	Duke Sci. Corp.	smut spore (fungal spore)
41	BGSS	Bermuda grass smut spores	Duke Sci. Corp.	
42	ACFTD	AC Fine Test Dust	Duke Sci. Corp.	other
43	NT	Nivea talc	Regular shop	
44	PPD	Printer paper dust	Regular shop	
45	PTD	Paper towel dust	Regular shop	
46	Cin	Cinnamon	Regular shop	
47	Cel	Celulose	Sigma-Aldrich	
48	GGL	Grinded Green Leaves	Own collection	

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

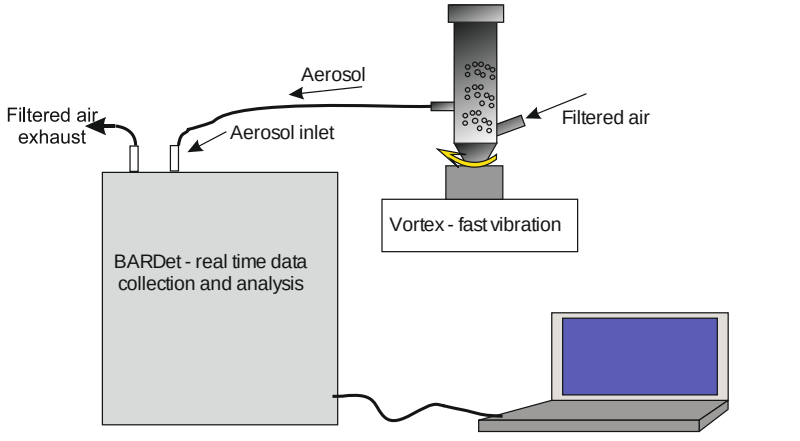


Figure 1. Setup of aerosol generation, data recording and analysis.

3.1.3. Data acquisition method and pre-processing

The fluorescence of each particle was recorded in 7 bands. It creates a time series of the signals which has to be pre-processed before further analysis. There are two steps of gathering data. First one is performed by internal BARDet's software, which is responsible for controlling the instrument and the acquisition of raw signals. Then data is forwarded to a pre-processing module of analysis software. Its first task is to extract valuable signals from the noise (three sigma rule). Then a normalization procedure is required. It is realized first by subtracting the average value of signal and then it normalizing to its standard deviation. The main goal was to analyze shape of emission spectrum (not signal strength).

An Important aspect of the data acquisition process was monitoring the rate of generation of aerosol, which should be stable (not too high or spontaneous). Finally, we gathered a total of 114 779 spectral characteristics of 48 aerosols which gives in average almost 2400 fluorescence signals per substance. It is important to note that fact because of its statistical value for the further analysis.

3.2. Data analysis

3.2.1. ANN (Artificial Neural Network)

3.2.1.1. Basics

There are many types of Artificial Neural Networks (ANN), but in this paper only the backpropagation algorithm is demonstrated because it is one of the most practical. The main concept of this algorithm is based on a model of neuron that has two tasks. It aggregates signals (1) and then processes them by an activation function (2), which, in this research, is a sigmoid. The result of such single processing is a new signal z_j propagated to other neurons (Figure 2).

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

Atmospheric Measurement Techniques Discussions

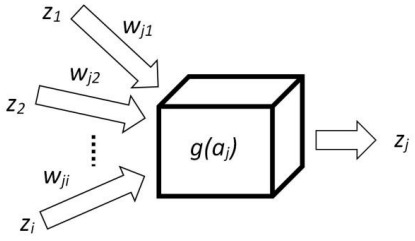


Figure 2. Mathematical model of single neuron cell.

$$a_j = \sum_i w_{ji} z_i \tag{1}$$

a_j- aggregated signal, w_{ji}- weight that connects neuron *i* with *j*, z_i- signal (input).

$$g(a_j) = \frac{1}{1 + e^{-\beta a_j}} \tag{2}$$

$g(a_j)$ – sigmoidal function, β- parameter (steepness) of sigmoid curve.

The structure of neural network is formed by layers of neurons: input, hidden and output. In this research input neurons are fluorescence spectrum and output neurons represent substances. In hidden layers (one and two hidden layers were examined) mostly actual computations are done. The schematic representation of neuron layers is presented in Figure 3.

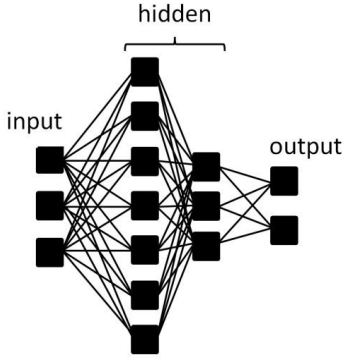


Figure 3. Typical topology of artificial neural network.

The described algorithm is the supervised learning method that requires training data for a teaching process. This allows one to calculate an error between the showed target and the ANN response. Every problem is related to minimizing output error which is calculated as Mean Squared Error (3).

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

Atmospheric Measurement Techniques Discussions

$$E = \frac{1}{2}\sum_{k=1}^{c}(y_k - t_k)^2 \quad (3)$$

E – Mean Squared Error, t_k- observed value (target), y_k- calculated response, k-output neuron, c – number of output neurons.

Gradient descent method is used to find a minimum of error function. Error is dependent on network weights Δw_{ji} which might be adjusted (4). In order to update weights correctly, the first one needs to propagate error backward by calculating partial derivatives δ_j (5) (Figure 4). All mathematical details are well described by Ch. M. Bishop book (Bishop, 1995).

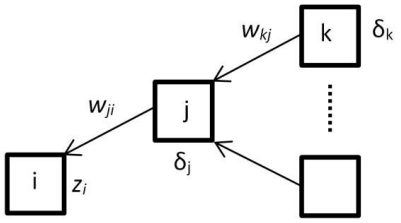


Figure 4. Model of backward error propagation.

$$\Delta w_{ji}(t) = -\eta\delta_j z_i + m\Delta w_{ji}(t-1) \quad (4)$$

η- learning rate, m - momentum, t - iteration.

$$\frac{\delta E}{\delta w_{ji}} = \frac{\delta E}{\delta a_j}\frac{\delta a_j}{\delta w_{ji}} = \delta_j z_i \qquad \delta_j = g'(a_j)\sum_k w_{kj}\delta_k \quad (5)$$

The learning rate factor determines the size of the steps while momentum parameter helps to skip local minimum by adding a fraction of the weight correction from the last step.

After the correction of all weights of ANN, the output error is examined and the procedure starts again unless an error level is low enough and there is no overfitting. All data are divided into three different sets: training, test and validation. For calculations during the learning process, only the first two are used. In order to determine whether it is time to stop teaching, one has to observe an error in the test set. There will be a moment when this error comes to be constant or starts increasing due to the overfitting of training data (Figure 5). The validation data set may be useful for confronting different models or just to verify the current model on completely separate set of data.

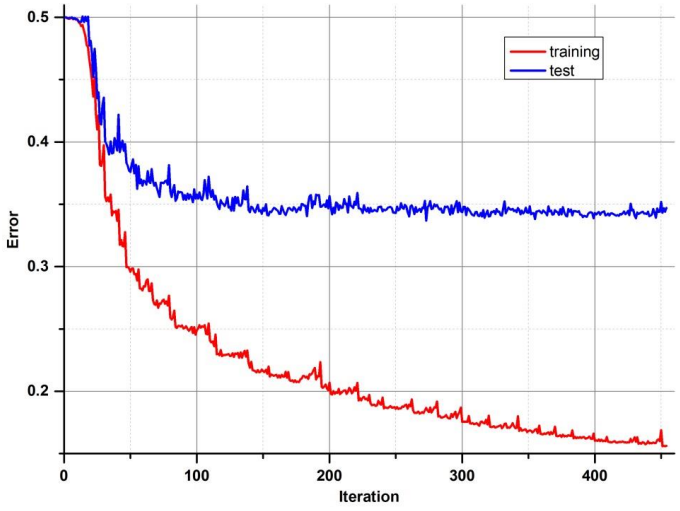


Figure 5. Example of error minimizing during training process.

3.2.1.2. Implementation of ANN for BARDet

There are statistical commercial software packages available that provide ANN modules as one of the methods to analyze the data. It is worthwhile noting that customized software was developed for the purpose of this research. This approach helped to understand ANN in depth and let to the development of software that is not only responsible for data pre-processing and network training, but also (mainly) for solving a real time classification problem.

Ruske et al. in their studies (Ruske et al., 2017) compared various algorithms to analyze single particle data and noted that ANN requires much more user input. However, we present the method to overcome this inconvenience by automatizing the process and implementing procedures that simplifies and improve analysis.

The main disadvantage of ANN is the fact that it is a parametrized algorithm. How well it works depends strictly on a proper choice of the best possible factors, which may be different for each problem. There are two types of factors that influence the ANN outcome. The first one corresponds to the architecture of ANN which comprises: number of layers, neurons and activation function parameter. The second one determines the learning process: momentum and learning rate. The last one can be tuned during the learning process to make it much faster. The “bold driver” procedure was chosen for that purpose. It continuously increases the learning rate unless an error is higher from that before the change. If it is, the algorithm radically decreases the learning rate and obtains weights from the last step again. Teaching ANN is a stochastic process caused by using randomly chosen initial weights. It was found that the best procedure for this investigation would be to make all optimization processes that way. Therefore, parameters of ANN, responsible both for structure and learning process, are randomly selected until the desired result is reached. In fact, the calculations are done automatically and simultaneously for several models due to multi core oriented software. The benefits of this approach are: time saving and high effectiveness of finding the best model. The last one is especially important, because the goal is to create a model that produces the best results, which doesn’t necessary mean creating a more complicated network (more neurons or

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

Atmospheric Measurement Techniques Discussions

layers).

3.2.2. Model evaluation

The main goal of analysis described in this paper is to find a solution to the bio-aerosol classification problem. When a training process ends, a final model is created: a network, which has a unique structure and a set of weights. One can create many of them and make a comparison only by a final error. It is not the best solution, because the goal is to distinguish patterns in data consistently, not to produce a network with a minimal error. That is why there is a need to make a final analysis of the results and evaluate the model in accordance with the best classification performance.

The standard method for visualization of results is a confusion matrix which will be necessary for Receiver Operating Characteristics (ROC) analysis (Fawcett, 2006). It simply shows what fraction of population for each class is predicted correctly or not. Each element from the data set increments one of the fields: TP, TN, FN and FP (Table 3). If it belongs to a diagonal (TP, TN), it was classified correctly.

Table 3. Structure of confusion matrix.

		Predicted class	
		positive	negative
True class	positive	True Positives (TP)	False Negatives (FN)
	negative	False Positives (FP)	True Negatives (TN)

The ROC graphs are very simple but useful tools for discovering whether a classifier is worth using or if it makes a random classification. It is based on two rates from confusion matrix: hit rate (6) and false alarm rate (7).

$$hit\ rate\ (true\ positive\ rate) = \frac{TP}{TP + FN} \tag{6}$$

$$false\ alarm\ rate\ (false\ positive\ rate) = \frac{FP}{FP + TN} \tag{7}$$

Each discrete classifier has a threshold level that assigns an element to a positive or negative class. The points of ROC graph (Figure 6) represent the classifier for many thresholds. The most desired curve reaches the highest true positive rate with the lowest false positive rate (convex line). The random classifier, in turn, has a hit rate equal to a false alarm rate despite threshold variation (diagonal line). To identify ROC analysis with one coefficient, the area under the curve (AUC) may be used. The higher value of AUC results in better performance (0.5-means random, 1-excellent).

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

Atmospheric Measurement Techniques Discussions

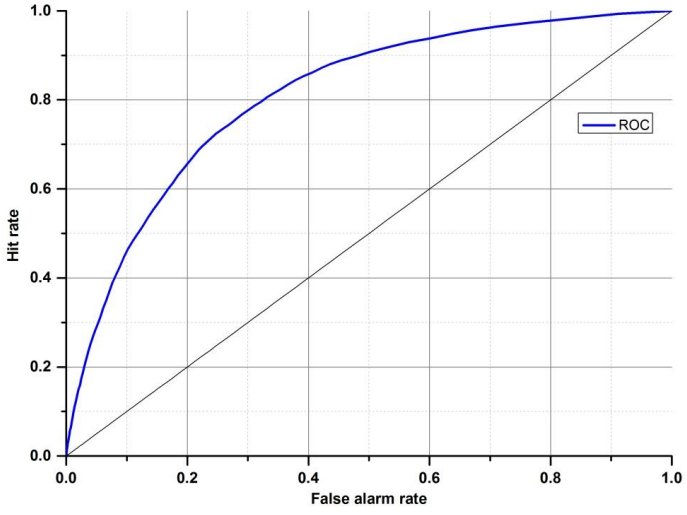


Figure 6. ROC graph with an example of classifier (blue).

The confusion matrix and ROC analysis described above were defined for two class problems (positive, negative). There is a straightforward way to expand it for the multi-class problem. One needs to take a desired class versus all other classes. Then there is a possibility to compare how good the classifier for specific classes within one model is.

4. Results

4.1. ANN performance

First attempts were made to distinguish all substances using only one neural network model. The tests revealed that it is impossible due to the huge number of samples (48 aerosols) and only a few of them presented significantly different fluorescence spectra. A more practical approach to this problem would be to create several groups (considering information about aerosols), but we did not want to make any classes *a priori*. Although the demonstrated ANN type needs a training, which requires a set of known classes, further tests showed that there is a possibility to find similarities between substances through the analysis of confusion matrices. It was achieved after many trials of matching substances, which were not well separated, into new groups and checking if they are good enough on ROC graphs. Consequently, this procedure was also applied to those new groups.

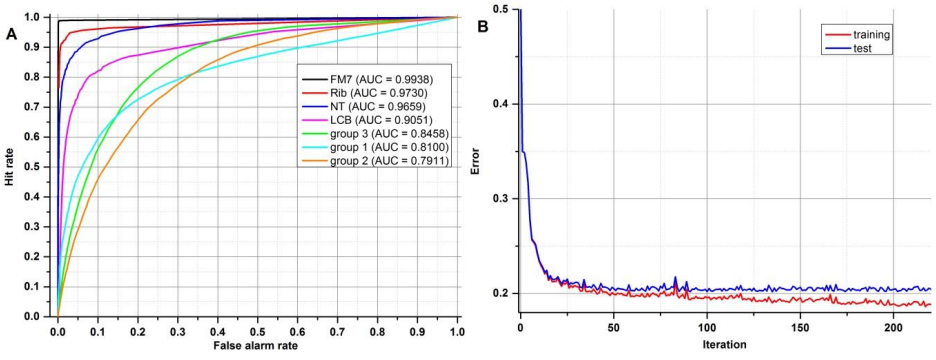
All examples demonstrated below were calculated on the test data sets, not training data. In the first presented network (Figure 7), which try to classify all of 48 substances (group 0), four aerosols reached very high accuracy of separation (AUC>0,9). The best separation was achieved for fluorescent microspheres (FM7). In this case 98.5% of all FM7 particles were correctly classified. Very high separation efficiency was achieved for riboflavin (Rib), NT (Talc) and LCB (*Lactobacillus bulgaricus*). The remaining aerosols were divided into 3 separate groups that gather the most similar substances (group 1-3) (Table 4). The subsequent groups up to 21 represent individual ANNs leading to the final classification of the aerosol.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

Atmospheric Measurement Techniques Discussions

265 Table 4. Confusion matrix of all aerosols.
266

		predicted						
		FM7	Rib	NT	LCB	group 3	group 1	group 2
true	FM7	98.5	0	0	0.3	0.1	0	1.1
	Rib	0.1	91	0.5	3.1	1.2	0.6	3.4
	NT	0	0.1	86.5	0	9.3	0.3	3.8
	LCB	1	1.6	0.6	72.7	3.9	10.7	9.5
	group 3	0	0.7	6.6	0.6	63.3	12	16.8
	group 1	0.2	1	1	7.9	12.5	61.6	15.8
	group 2	0.1	1.2	3.8	6.6	17.6	13.2	57.4



267

268 Figure 7. (A) ROC and (B) error progress of ANN that classifies all samples.

269 Table 5 and Figure 8 show results achieved for two substances that have very similar spectrum
270 and calculated AUCs are not much higher than in a random classifier. This example clearly shows why
271 we are not always able to classify each particle of aerosol with 100% accuracy. However, just a
272 representative number of measured particles allows the proper prediction of aerosol types within a
273 few seconds. This is easy to observe during real time detection, because counts allocated in
274 confusion matrix tend to reach a stable state quite quickly.
275
276 Table 5. Confusion matrix of two substances that have very similar spectra.

		predicted	
		BWF	Cel
true	BWF	54.8	45.2
	Cel	45.6	54.4

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

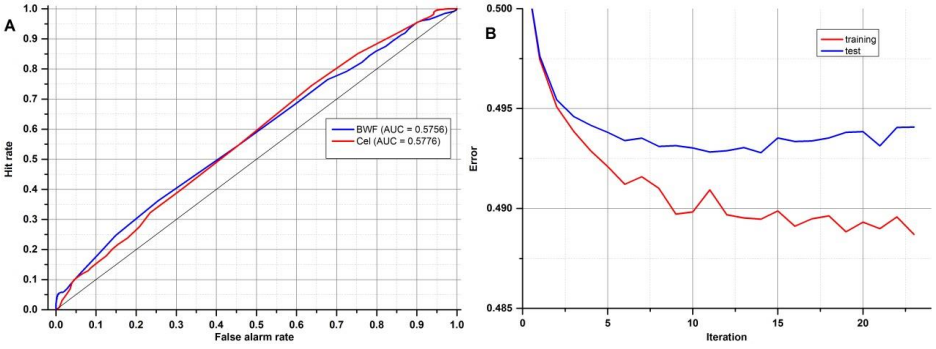


Figure 8. ROC (A) and error progress (B) of ANN that classifies two very similar samples.

4.2. Classification tree

Finally, to achieve the best possible classification, the decision tree was created (Figure 9). It comprises not one but 22 models. It is difficult to present confusion matrices and ROC graphs for all neural networks in this paper, so only the most interesting one has been discussed. Here, each node represents a network that classifies a group of aerosols. The aerosols on the left side of the diagram show the most distinct differences, thus they are easy to classify (Level 0). In the right direction (Level 1-5) this task is much more demanding due to similar spectrum and the separation is less probable in accordance to single particles, although still it is very useful from a practical point of view for aerosol cloud discrimination.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-86
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 20 April 2018

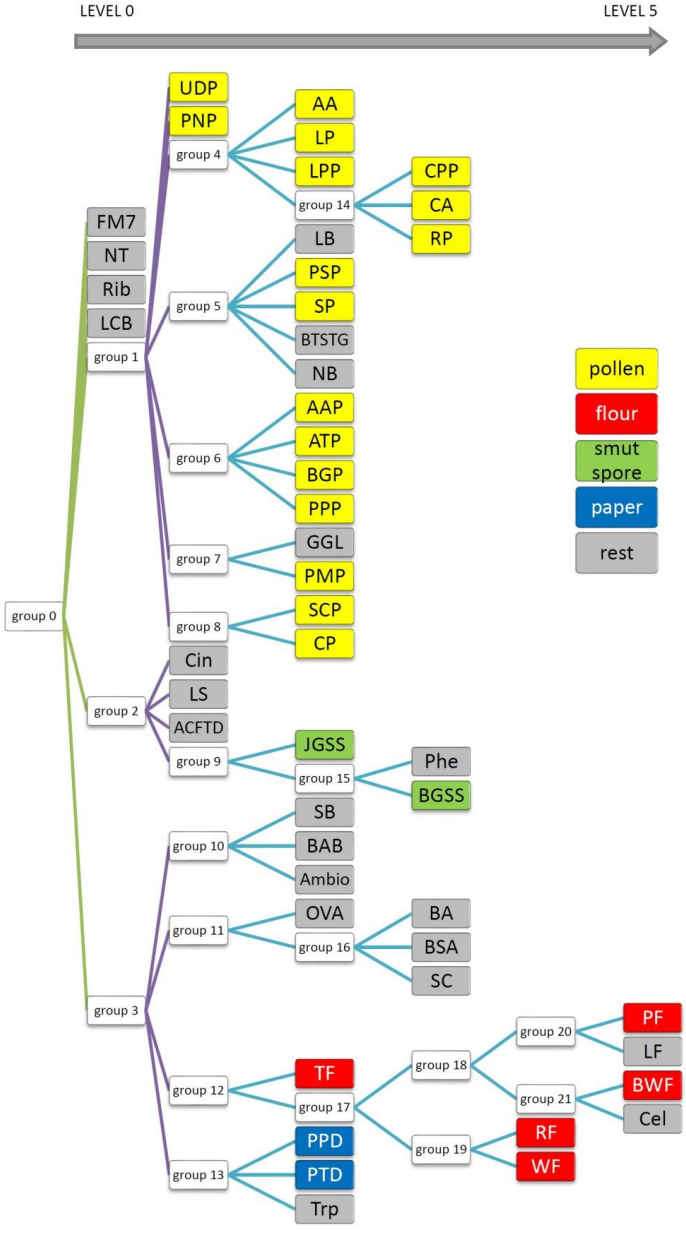


Figure 9. Decision tree consists of 22 ANN separating 48 substances.

At first glance one could see that FM7 and Rib are very well recognized, but that was expected, because these are standards of fluorescence. Surprisingly, NT and LCB aerosols were also separated from the others (Level 0 network). Further analysis of the tree structure identifies a correlation between samples and their real categories, especially it is noticeable for Pollens, which are allocated on a separate branch of that tree and all stems from group 1. Most of them were classified on the third level. Interestingly all grass pollens (AAP, ATP, BGP, PPP) belong to the same group 6. Similarly

both *Lycopodium* pollens from different regions of the word show close correlation, however *Abies alba*, which is a tree, was classified to the same group. Flours, Smut Spores and Papers are dispersed between different levels but particular groups belong to the same branch of the tree. However, some of samples, are scattered on the whole tree area and do not correspond to any group.

It should be noted that the result is a system of 22 ANNs that works simultaneously. In comparison to the training process, which is rather time consuming and has to be empirically optimized, this cluster of learned ANN's delivers very high performance. Input data is processed by a single ANN in milliseconds. This performance makes neural network a great tool as a splitting node in the classification tree. Comparing to our previous results, where Principal Component Analysis was applied to analyze data from BARDet (Kaliszewski et al., 2016), the ANN allowed much better discrimination between various bio-aerosols.

5. Summary

In this paper the possibility of an application of the Artificial Neural Network (ANN) for a real time classification of biological aerosols was investigated. The spectral characteristics of bio-aerosols were collected using the BARDet instrument. Finally, the database consisted a large data set of 114 799 samples (particle characteristics) of 48 substances. It ensured that application of the ANN was fully justified. Finally, we trained 22 neural networks and combined them into a decision tree, which was laborious and time consuming. However, trained ANN's characterized single particles in real time. Tests revealed that only several substances have such characteristic fluorescence spectra that allows correct classification of almost each particle. However, in all other cases the system was able to recognize a particular aerosol cloud. Further approximation was based on decision tree analysis where each node corresponded to a separate learned ANN. The best sets of ANN's for each group of similar aerosols were discovered utilizing confusion matrices and ROC analysis. Our intentions were to make a complete system which detects and classifies substances without creating groups *a priori*. This attitude helped to create a powerful analytical tool that works automatically and the results of classification are immediately available on the operator's screen.

This study proved that it is possible to create a tool for a highly effective analysis of bio-aerosols using multiple ANNs combined into decision tree. Our approach allowed automation and speed up of analysis, which reduced time and the amount of needed computing power. In a future study we will extend the database to obtain possibly vast variety of samples including bacteria and fungi. Finally, the actual performance of the system will be determined under real environmental conditions.

Acknowledgments

Presented work was supported by grant from The National Centre of Research and Development (Poland), project: "Mobile laboratory for environmental sampling and identification of biological threats" (O ROB 0031 01/ID/31/1).

Bibliography

Agranovski, V., Ristovski, Z., Hargreaves, M., Blackall, P. J. and Morawska, L.: Performance evaluation of the UVAPS: Influence of physiological age of airborne bacteria and bacterial stress, J. Aerosol Sci., 34(12), 1711–1727, doi:10.1016/S0021-8502(03)00191-5, 2003.

339 Antowiak, M. and Chałasińska-Macukow, K. C. H. a: Fingerprint identification by using
340 artificial neural network with optical wavelet preprocessing, , 11(4), 327–337, 2003.

341 Bhangar, S., Huffman, J. A. and Nazaroff, W. W.: Size-resolved fluorescent biological aerosol
342 particle concentrations and occupant emissions in a university classroom, Indoor Air, 24(6),
343 604–617, doi:10.1111/ina.12111, 2014.

344 Bishop, C. M.: Neural Networks for Pattern Recognition, Oxford University Press, Inc., New
345 York, NY, USA., 1995.

346 Blais-Lecours, P., Perrott, P. and Duchaine, C.: Non-culturable bioaerosols in indoor settings:
347 Impact on health and molecular approaches for detection, Atmos. Environ., 110, 45–53,
348 doi:10.1016/j.atmosenv.2015.03.039, 2015.

349 Borecki, M., Korwin-Pawlowski, M. L. and Beblowska, M.: A Method of Examination of
350 Liquids by Neural Network Analysis of Reflectometric and Transmission Time Domain Data
351 From Optical Capillaries and Fibers, IEEE Sens. J., 8(7), 1208–1214,
352 doi:10.1109/JSEN.2008.926182, 2008.

353 Choi, K., Ha, Y., Lee, H. K. and Lee, J.: Development of a biological aerosol detector using
354 laser-induced fluorescence and a particle collection system, Instrum. Sci. Technol., 42(2),
355 200–214, doi:10.1080/10739149.2013.855639, 2014.

356 Davidson, C. I., Phalen, R. F. and Solomon, P. A.: Airborne particulate matter and human
357 health: A review, Aerosol Sci. Technol., 39(8), 737–749, doi:10.1080/02786820500191348,
358 2005.

359 Deguillaume, L., Leriche, M., Amato, P., Ariya, P. a., Delort, A. M., Pöschl, U., Chaumerliac, N.,
360 Bauer, H., Flossmann, a. I. and Morris, C. E.: Microbiology and atmospheric processes:
361 chemical interactions of Primary Biological Aerosols, Biogeosciences Discuss., 5(1), 841–870,
362 doi:10.5194/bgd-5-841-2008, 2008.

363 Fawcett, T.: An introduction to ROC analysis, Pattern Recognit. Lett., 27(8), 861–874,
364 doi:https://doi.org/10.1016/j.patrec.2005.10.010, 2006.

365 Fennelly, M. J., Sewell, G., Prentice, M. B., O'Connor, D. J. and Sodeau, J. R.: Review: The Use
366 of Real-Time Fluorescence Instrumentation to Monitor Ambient Primary Biological Aerosol
367 Particles (PBAP), Atmosphere (Basel)., 9(1), 2018.

368 Feugnet, G., Lallier, E., Grisard, A., McIntosh, L., Hellström, J. E., Jelger, P., Laurell, F., Albano,
369 C., Kaliszewski, M., Wlodarski, M., Mlynczak, J., Kwasny, M., Zawadzki, Z., Mierczyk, Z.,

370 Kopczynski, K., Rostedt, A., Putkiranta, M., Marjamäki, M., Keskinen, J., Enroth, J., Janka, K.,
371 Reinivaara, R., Holma, L., Humppi, T., Battistelli, E., Iliakis, E. and Gerolimos, G.: Improved
372 laser-induced fluorescence method for bio-attack early warning detection system, in
373 Proceedings of SPIE - The International Society for Optical Engineering, vol. 7116, Thales
374 Research and Technology, France., 2008.

375 Fröhlich-Nowoisky, J., Kampf, C. J., Weber, B., Huffman, J. A., Pöhlker, C., Andreae, M. O.,
376 Lang-Yona, N., Burrows, S. M., Gunthe, S. S., Elbert, W., Su, H., Hoor, P., Thines, E.,
377 Hoffmann, T., Després, V. R. and Pöschl, U.: Bioaerosols in the Earth system: Climate, health,
378 and ecosystem interactions, Atmos. Res., 182, 346–376,
379 doi:10.1016/j.atmosres.2016.07.018, 2016.

380 Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier Van Der Gon, H., Facchini, M. C.,
381 Fowler, D., Koren, I., Langford, B., Lohmann, U., Nemitz, E., Pandis, S., Riipinen, I., Rudich, Y.,
382 Schaap, M., Slowik, J. G., Spracklen, D. V., Vignati, E., Wild, M., Williams, M. and Gilardoni, S.:
383 Particulate matter, air quality and climate: Lessons learned and future needs, Atmos. Chem.
384 Phys., 15(14), 8217–8299, doi:10.5194/acp-15-8217-2015, 2015.

385 Gabey, A. M., Gallagher, M. W., Whitehead, J., Dorsey, J. R., Kaye, P. H. and Stanley, W. R.:
386 Measurements and comparison of primary biological aerosol above and below a tropical
387 forest canopy using a dual channel fluorescence spectrometer, Atmos. Chem. Phys., 10(10),
388 4453–4466, doi:10.5194/acp-10-4453-2010, 2010.

389 Gabey, A. M., Stanley, W. R., Gallagher, M. W. and Kaye, P. H.: The fluorescence properties
390 of aerosol larger than 0.8 μ in urban and tropical rainforest locations, Atmos. Chem. Phys.,
391 11(11), 5491–5504, doi:10.5194/acp-11-5491-2011, 2011.

392 Górny, R. L.: Filamentous microorganisms and their fragments in indoor air - A review, Ann.
393 Agric. Environ. Med., 11(2), 185–197, doi:10.1007/BF02677055, 2004.

394 Hernandez, M., Perring, A. E., McCabe, K., Kok, G., Granger, G. and Baumgardner, D.:
395 Chamber catalogues of optical and fluorescent signatures distinguish bioaerosol classes,
396 Atmos. Meas. Tech., 9(7), 3283–3292, doi:10.5194/amt-9-3283-2016, 2016.

397 Hill, S. C., Pinnick, R. G., Niles, S., Pan, Y.-L., Holler, S., Chang, R. K., Bottinger, J., Chen, B. T.,
398 Orr, C.-S. and Feather, G.: Realtime Measurement of Fluorescence Spectra from Single
399 Airborne Biological Particles, F. Anal. Chem. Technol., 3(4–5), 221–239,
400 doi:10.1002/(SICI)1520-6521(1999)3:4/5<221::AID-FACT2>3.3.CO;2-Z, 1999.

401 Huffman, J. A., Treutlein, B. and Pöschl, U.: Fluorescent biological aerosol particle
402 concentrations and size distributions measured with an Ultraviolet Aerodynamic Particle
403 Sizer (UV-APS) in Central Europe, Atmos. Chem. Phys., 10(7), 3215–3233, doi:10.5194/acp-

10-3215-2010, 2010.

Kaliszewski, M., Trafny, E. A., Lewandowski, R., Włodarski, M., Bombalska, A., Kopczyński, K., Antos-Bielska, M., Szpakowska, M., Młyńczak, J., Mularczyk-Oliwa, M. and Kwaśny, M.: A new approach to UVAPS data analysis towards detection of biological aerosol, J. Aerosol Sci., 58, 148–157, doi:https://doi.org/10.1016/j.jaerosci.2013.01.007, 2013.

Kaliszewski, M., Włodarski, M., Młyńczak, J., Leśkiewicz, M., Bombalska, A., Mularczyk-Oliwa, M., Kwaśny, M., Buliński, D. and Kopczyński, K.: A new real-time bio-aerosol fluorescence detector based on semiconductor CW excitation UV laser, J. Aerosol Sci., 100, 14–25, doi:10.1016/j.jaerosci.2016.05.004, 2016.

Kohlus, R. and Bottlinger, M.: Particle Shape Analysis as an example of knowledge extraction by neural nets, Part. Part. Syst. Charact., 10(5), 275–278, doi:10.1002/ppsc.19930100511, 1993.

Lakowicz, J. R.: Principles of Fluorescence Spectroscopy, Second., Kluwer Academic/Plenum Publishers., 1999.

Lee, B. U., Jung, J. H., Yun, S. H., Hwang, G. B. and Bae, G. N.: Application of UVAPS to real-time detection of inactivation of fungal bioaerosols due to thermal energy, J. Aerosol Sci., 41(7), 694–701, doi:https://doi.org/10.1016/j.jaerosci.2010.04.003, 2010.

Leśkiewicz, M., Kaliszewski, M., Mierczyk, Z. and Włodarski, M.: Comparison of Principal Component Analysis and Linear Discriminant Analysis applied to classification of excitation-emission matrices of the selected biological material, Biul. Wojsk. Akad. Tech., 65(1), 15–31, doi:10.5604/12345865.1197960, 2016.

Lim, D. V., Simpson, J. M., Kearns, E. A. and Kramer, M. F.: Current and developing technologies for monitoring agents of bioterrorism and biowarfare, Clin. Microbiol. Rev., 18(4), 583–607, doi:10.1128/CMR.18.4.583-607.2005, 2005.

Mauderly, J. L. and Chow, J. C.: Health effects of organic aerosols., 2008.

Miaskiewicz-Peska, E. and Lebkowska, M.: Comparison of aerosol and bioaerosol collection on air filters, Aerobiologia (Bologna)., 28(2), 185–193, doi:10.1007/s10453-011-9223-1, 2012.

Michaels, R. A.: Environmental Moisture, Molds, and Asthma—Emerging Fungal Risks in the Context of Climate Change, Environ. Claims J., 29(3), 171–193, doi:10.1080/10406026.2017.1345521, 2017.

435 Pan, Y. Le, Hill, S. C., Pinnick, R. G., House, J. M., Flagan, R. C. and Chang, R. K.: Dual-
436 excitation-wavelength fluorescence spectra and elastic scattering for differentiation of single
437 airborne pollen and fungal particles, Atmos. Environ., 45(8), 1555–1563,
438 doi:10.1016/j.atmosenv.2010.12.042, 2011.

439 Pan, Y. Le, Huang, H. and Chang, R. K.: Clustered and integrated fluorescence spectra from
440 single atmospheric aerosol particles excited by a 263- and 351-nm laser at New Haven, CT,
441 and Adelphi, MD, J. Quant. Spectrosc. Radiat. Transf., 113(17), 2213–2221,
442 doi:10.1016/j.jqsrt.2012.07.028, 2012.

443 Pinnick, R. G., Hill, S. C., Pan, Y. Le and Chang, R. K.: Fluorescence spectra of atmospheric
444 aerosol at Adelphi, Maryland, USA: Measurement and classification of single particles
445 containing organic carbon, Atmos. Environ., 38(11), 1657–1672,
446 doi:10.1016/j.atmosenv.2003.11.017, 2004.

447 Pöhlker, C., Huffman, J. A. and Pöschl, U.: Autofluorescence of atmospheric bioaerosols:
448 Spectral fingerprints and taxonomic trends of pollen, Atmos. Meas. Tech., 6(12), 3369–3392,
449 doi:10.5194/amt-6-3369-2013, 2013.

450 Pope III, C. A. and Dockery, D. W.: 2006 Critical Review: Health Effects of Fine Particulate Air
451 Pollution: Lines That Connect, J. Air Waste Manag. Assoc., 56(6), 709–742,
452 doi:10.1080/10473289.2006.10464485, 2006.

453 Pósfai, M. and Buseck, P. R.: Nature and Climate Effects of Individual Tropospheric Aerosol
454 Particles, Annu. Rev. Earth Planet. Sci., 38(1), 17–43,
455 doi:10.1146/annurev.earth.031208.100032, 2010.

456 Purnomo, H. D., Hartomo, K. D. and Prasetyo, S. Y. J.: Artificial Neural Network for Monthly
457 Rainfall Rate Prediction, IOP Conf. Ser. Mater. Sci. Eng., 180(1), 12057, 2017.

458 Ruske, S., Topping, D. O., Foot, V. E., Kaye, P. H., Stanley, W. R., Crawford, I., Morse, A. P. and
459 Gallagher, M. W.: Evaluation of machine learning algorithms for classification of primary
460 biological aerosol using a new UV-LIF spectrometer, Atmos. Meas. Tech., 10(2), 695–708,
461 doi:10.5194/amt-10-695-2017, 2017.

462 Shiraiwa, M., Selzle, K. and Pöschl, U.: Hazardous components and health effects of
463 atmospheric aerosol particles: reactive oxygen species, soot, polycyclic aromatic compounds
464 and allergenic proteins, Free Radic. Res., 46(8), 927–939,
465 doi:10.3109/10715762.2012.663084, 2012.

466 Taketani, F., Kanaya, Y., Nakamura, T., Koizumi, K., Moteki, N. and Takegawa, N.:
467 Measurement of fluorescence spectra from atmospheric single submicron particle using

laser-induced fluorescence technique, J. Aerosol Sci., 58, 1–8, doi:https://doi.org/10.1016/j.jaerosci.2012.12.002, 2013.

Trafny, E. A., Lewandowski, R., Stępińska, M. and Kaliszewski, M.: Biological threat detection in the air and on the surface: How to define the risk, Arch. Immunol. Ther. Exp. (Warsz)., 62(4), 253–261, doi:10.1007/s00005-014-0296-8, 2014.