Responses to reviewers

We would like to thank the Reviewers for evaluation of our manuscript. The detailed answers to the referee's questions are as follows:

Reviewer #1

This manuscript details the use of an Artificial Neural Network, or ANN, to attempt to better identify bio-aerosol. Bio-aerosol has been a topic of contemporary interest in the atmospheric sciences and neural networks have gained prominence as a data reduction and analysis technique. This is therefore a paper that could be of interest to the AMT readership. There are however several large missing sections, e.g. aerosol justification and characterization, that should be addressed before it is publishable.

1. The writing of the paper is a bit too familiar and there are many unquantifiable terms, e.g. "Society is awaiting anxiously for system that could inform them in real-time about a real danger that is suspended in the air." – this would be a rather improved paper if this type of writing could be toned down as in "There is a need for real-time information about ambient particulate matter."

- The sentence was corrected.

2. In addition, the paper could benefit from a through read from a native English speaker with a focus on removal of incorrect and non-scientific terms. Examples, but by no means comprehensive: "really promising", "very high performance", past tense of grind is ground, not grinded, etc.

- The language correction was performed.

3. The name of the technique to which the ANN is applied, BARDet, should be stated in the abstract.

- The name of the device was added to the abstract.

4. The central issue with this paper is there needs to be a description of the aerosol generation method and the produced size distribution of each sample; some are solids, some are liquids. Were sizes comparable? Concentrations? Ideally this is a sub-section of 3.1.2.

- All aerosols were generated from powders only as it was described in section 3.1.2. The sizes depend on dimensions of particles. An information about particle's sizes was added to the table 2.

Going farther, why were these samples chosen? Some seem rather important e.g. pollens, while others are unclear. Paper towel? Multiple broths? It is upon the authors not to simply present so may aerosol types but instead (1) care-fully and completely characterize the aerosol investigated – not only what they look like to the BARDet - and (2) to argue why they are being investigated (do they have any atmospheric importance which is the theme of the paper)?

- The following explanation was added to the 3.1.2 section: “In order to achieve reliable aerosol classification the ANN’s needs to be trained using possibly large number of measurement data. Therefore, various particle types, that can be easily aerosolized, were tested. Samples like pollens, fungi, bacteria, spores and leaves scraps are present in the atmosphere. Biofluorophores like riboflavin, cellulose, aminoacids and proteins were also characterized since they are components of biological materials. The group of bacterial growth media was investigated due to their strong influence on bacteria fluorescence especially if they are not sufficiently washed. This can occur in case of intentionally released bacterial aerosols. Due to technical limitations the other than pharmaceutical samples could be aerosolized in this study. The aerosols of flours, and fluorescent non-biological substances like paper dust, AC fine Test Dust and talc were analyzed since they can occur especially in indoor and public places. The non-fluorescent particles were not a subject of the research since they can be automatically discarded as non-biological applying given fluorescence threshold.”

5. Going a step further, although there are 48 aerosol types suggested, in practive the confusion matrix says the separation is based on 7 broader classes. If this is indeed the case (as it appears) then (1) the abstract should reflect separation of 7 classes, not the 48 stated (2) Table 1 should state what fits into each class, since this is the central concept.

- In the manuscript we have stated as follows:
  “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.”
- In practice separation is done not by one confusion matrix (ANN) but by all of them in sequence (22 ANN’s combined in a decision tree). For example, if ANN classifies unknown substance into any of 22 groups it means that decision process is not ended but from that moment another ANN classifies this substance. That’s why there are substances which only needs one ANN to make a classification (e.g. FM7), but there are also such which needs 6 ANN (e.g. BWF) to complete the task. The main difference between this two examples is that 98.5% of all FM7 particles are classified correctly, but BWF has only 54.8% detected particles. However in both cases system recognize aerosol type every time with no mistake.

6. The statistic in Table 4 need to be placed in the abstract and repeated in the summary, these are the central results.

- Table 3, previously Table 4 do not represents the central result. It is only 1 of 22 nodes of a decision tree. The most important fact is that each one aerosol type can be recognized. In the abstract we added as follows: “In both cases the system recognized aerosol type with no mistake.”

For example, in Tables 4 and 5 it appears that there can be confusion on the 50th centile level. This is not altogether great separation and should be explicitly stated for the reader from the outset.

- It was stated in the text. However, we hope that modified explanation will be helpful (Lines 451-456).

The 48 types and 114k number of spectra, which are the data set, belong only in the methods section; while these seem rather impressive they are not results. The authors should therefore replace the sentences which repeat these values in abstract and summary with the separation ability.

- We are agree with reviewer that number of data are not a result. Therefore they were removed from the abstract and summary.

7. Table 3 is overly simplistic for a table; this can be stated in a single sentence. Please remove.

- The sentence was added and the table was removed (Lines 382 – 384).

8. In the summary : “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.” – this is again an unquantified statement. It is also at odds with “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.” Please provide the separation ability and then let the reader judge is this is a highly effective analysis.

- We provided for the reader only two examples that shows good and poor separation in accordance for individual particle within only these two groups (group 0 and group 21). Probably it was not emphasized clearly enough in the manuscript that system recognize aerosol type (all of them) with no mistake every time and that was main goal to achieve in presented analysis.
- In the lines 581-583 we added as follows:
  “However, in all other cases the system was able to recognize a particular aerosol accurately with no mistake, but a representative number of several dozens of particles in a cloud was necessary.”

9. Why weren’t non-biological materials tested?

- The materials and methods section was improved. We justified the use of tested samples. We also changed confusing title in 3.1.2 “Bioaerosols” for “Aerosols”
- The non-biological materials were tested:
  Fluoromax microspheres 7 um
  Nivea talc
  Printer paper dust
  Paper towel dust
  AC Fine test dust (This one can contain also biological particles)
- The most of non-biological materials like gypsum, syloid, desert sand are non-fluorescent and there is no any problem to differentiate them from biological particles.

Reviewer #3

In this paper the authors present a method for bio-aerosol classification using labelledlaboratory data. The authors are correct in noting the need to improve and document such methods for improved bio-aerosol research. However before publication is considered, I feel the following points should be addressed. Presently it is unclear how anyone might replicate these results.

Minor points:

The formatting of references is wrong? Please check with the Copernicus guidelines and change from (xx)(xx) format to (xx;xx;xx...)

- The formatting of references was corrected.

There is a range of grammatical issues that need revising before publication. I have listed some below but would suggest the authors re-read the paper and change accordingly, removing any vague descriptions that require support with numerics or information to enable replication of experimental conditions. E.g:

Line 76: 'This paper focuses on the application of ANN for real time discrimination of bio-aerosols basing on single particle fluorescence characteristics.' Please change 'basing' to 'based'

- Corrected

Line 108: 'The concentration of the aerosols was adjusted with vibration frequency of [the] vortex.

- Corrected

Line 176: In order to determine whether it is time to stop teaching,.

This is too informal. I would suggest rewriting in terms of the fitting process.

- In our opinion "teaching" process is appropriately used phrase and is widely applied in ANN related literature. We used "overfitting" in context of data not the learning process.

Specific Points:

In table 2 the authors use the term 'own collection'. I'm a little concerned this does not provide enough information to enable replication of results. Where was the sample obtained? How old? Also the terms 'regular shop' and 'pharmacy' raise similar concerns.

Which Pharmaceutical brand?

- The description and full information on the samples was added to the table 2.

Would it be possible to present size and shape information for each specie in a separate table?

- The missing data were added to the table 2.

Line 119: Please list the bands of florescence recorded. You have done so in Table 1 but you should reference this table in the text on this line to avoid confusion.

- The table has been referenced in the text just above.

Line 127: '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). 'Please define how this is quantified. What is 'too high'? How would this experiment be repeated?

- The BARDet's measurement window is 20us, but the data are integrated and recorded every 2 ms. It gives up to 100 averaged aerosol characteristics per 2 ms. It does not strongly influence the result if one aerosol type is measured, however, we tried to avoid such measurements.
- The sentence in the manuscript was clarified (Lines: 335-337) as follows: "The data acquisition process started after stabilization of aerosol generation rate which was measured by the device. It was important to not exceed one particle per 2 ms of data integration time at 20 us measurement window."

Line 130: 'It is important to note that fact because of its statistical value for the further analysis'. What statistical value?

- The sentence was removed.

Section 3.2.1.2: What comparisons have been made, if any, between the bespoke implementation of the ANN in this work with what should be identical performance in existing software packages? How do we know the implementation of the bespoke ANN is correct? Please provide evidence.

- The presented ANNs were not compared to existing packages. We believe that our implementation of ANNs is correct since they produce correct results on approximated mathematical functions.

<u>Major points:</u>
It is difficult to contextualise the input data being used. Please provide a visualization of some example spectra.

- An exemplary characteristics were added as a figure 2.

To the best of the reviewers understanding, each particle will be classified at multiple levels of the decision tree. For example each particle will be classified as FM7, Rib, NT, LCB, or group 1 etc. and then should the particle be identified as group 1, the particle will then get classified again as UDP, PNP, group 4 etc.

- Yes. In practice separation is done not by one confusion matrix (ANN) but by all of them in sequence (22 ANN's combined in a decision tree). For example, if ANN classifies unknown substance into any of 22 groups it

means that decision process is not ended but from that moment another ANN classifies this substance. That's why there are substances which only needs one ANN to make a classification (e.g. FM7), but there are also such which needs 6 ANN (e.g. BWF) to do that. Main difference between this two examples is that 98.5% of all FM7 particles are classified correctly but BWF has only 54.8% detected particles. However in both cases system recognize aerosol type every time with no mistake.

For example, should a particle from group 2 be misclassified and placed into group 1, which will happen about 12% of the time, how does this error propagate down the tree? Will it be evenly distributed amongst UDP, PNP, group 4 etc. or will it be heavily weighted towards one class?

- Error should be distributed according to confusion matrix of the group where particle is classified. There are 22 groups/ANN's/confusion matrices. In paper only 2 were presented as an examples.

With the exception of the level 0 ANN, I assume that each of the ANNs are trained only on a subsection of the data. This needs to be clarified. For example the ANN for group 1, is trained in absence of the data from group 2 etc.

- It is done exactly like that.
- To clarify the text to the reader the following sentence in lines 504-516 was added: "In practice separation is done not by one confusion matrix (ANN) but by all of them in sequence (22 ANN's combined in a decision tree). For example, if ANN classifies unknown substance into any of 22 groups it means that decision process is not ended but from that moment another ANN classifies this substance. However, each new ANN is trained using only subsection of the data excluding the data from other groups."

On line 245 it is stated that it is impossible to produce a single neural network to perform classification of all 48 classes. Need to be clear whether this means that it is impossible because of the number of classes, or that it is possible to create a single neural network but the classification error is unreasonably high.

- Our intention was to reporting that it is impossible to distinguish all substances using one ANN, not to create such single ANN.
- In the manuscript it was as follows: "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."
- To clarify the text in lines 487-488 where additional explanation was added: "…that allow accurate characterization. The remaining substances are then misclassified. Therefore, we decided to use a…."

Would it be possible to produce a contour confusion matrix plot for the full 48 classes, for a single ANN and for the approach suggested in the manuscript, or to provide adjusted rand score or percentage of particles correctly classified to demonstrate

whether better classification can be attained using the tree of ANNs as opposed to a single ANN?

- Such network and comparison has been made but authors decided not to present such single ANN, just mentioned about it in the text. Also presentation of 48 substances ANN would be hard to follow due to large number of data.

How was the decision tree created? I.e. how it was decided which individual classes would be placed into group 1 through 3?

- The process of creation of decision tree was described in the manuscript as follows: “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.”
- New groups had been tested by creating for them new ANN’s and checking by ROC graphs which one separates substances better. Many of them had been trained before the best ones were found. The Final ANN’s were learned after dozens of trials.

The authors have indicated on line 203 that the hyper-parameters of the ANNs have been randomly selected until the desired/best result is reached. In terms of reproducibility, it would be helpful to specify the range of parameters which were tested and which of these options produced the best results. Also did each of the 22 networks utilise the same hyper parameters, or was this optimisation conducted for each of the 22 networks?

- It is impossible to reproduce learning process. Even if exactly the same parameters are chosen the learning process will generate each time different result according to randomly chosen initial weights. The range of parameters is typical for backpropagation algorithm and is well documented in the literature. Therefore, authors decided to perform random parameters procedure demonstrated in the paper.

There is no discussion on either data or software availability. The authors need to consider the default Copernicus publishing rules and provide text that would allow others to request access to both the data and software. If this is restricted, it should be stated with the reasons why. https://www.atmospheric-measurement-techniques.net/about/data_policy.html

- The following sentence was added in the manuscript ”The experimental aerosol data can be provided upon request. The software for automatic data analysis cannot be commonly provided at this moment since it is a subject of negotiations with a company.”