

Classification of Aerosol Particles using Inductive Conformal Prediction

Linn Karlsson

*Department of Environmental Science
Stockholm University
Stockholm, Sweden*

LINN.KARLSSON@ACES.SU.SE

Henrik Boström

*School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology
Stockholm, Sweden*

BOSTROMH@KTH.SE

Paul Zieger

*Department of Environmental Science
Stockholm University
Stockholm, Sweden*

PAUL.ZIEGER@ACES.SU.SE

Editor: Alexander Gammerman, Vladimir Vovk, Zhiyuan Luo, Evgeni Smirnov and Giovanni Cherubin

Abstract

Aerosol particles are small airborne particles suspended in air affecting the climate and human health. Different types of particles come from different sources and impact the environment in different ways, which is why a reliable particle classification is of interest. In this study, inductive conformal prediction is applied to a dataset of laboratory-generated aerosol particles, consisting of ten particle subclasses that can be grouped into four parent classes for classification. The performance of the inductive conformal predictor (ICP) is evaluated on particle subclasses that were not included in training or calibration. The ICP appears to give accurate predictions in some cases, namely if the unknown particle is similar to the known ones in the parent class. The precision of the underlying model is not high enough to reject *all* unknown particles for any subclass at the chosen significance levels, but the ICP manages to reject them at a higher rate if they are sufficiently different from the training and calibration samples. Overall, the performance is not straightforward to evaluate and it seems to depend on the heterogeneity and size of the classes of particles. Further investigations using a simpler data and model set-up would be beneficial, and data and sampling standardisation should be considered more carefully if the model is to be applied to field measurements.

Keywords: inductive conformal prediction, particle classification, single-particle measurements, aerosols, atmosphere

1. Introduction

Aerosol particles impact air quality and human health, e.g. by causing respiratory problems, and also affect the climate through interactions with clouds and radiation. The effect of a given aerosol particle depends both on its size, shape and chemical composition (Sein-

feld and Pandis, 2016). Identifying which types of particles, e.g. salt, dust, bioaerosols, etc., are present in the atmosphere is therefore important for characterising the sources of particles and understanding the details of how they affect our environment. Bioaerosols, such as pollen and fungal spores, contain materials that can absorb and re-emit radiation. Because of this property, fluorescence spectroscopy techniques are often used for real-time measurements of bioaerosols (Huffman et al., 2019). A common technique is ultraviolet light-induced fluorescence, whereby individual aerosol particles are irradiated with ultraviolet light and the resulting fluorescent signals can be used to distinguish between different bioaerosol types (Huffman et al., 2019). In addition to the fluorescent properties, instruments often record additional information about the size and morphology of the particles to help the discrimination (Ruske et al., 2018).

Real-time, single particle measurements generate large amounts of data, and machine learning techniques are important tools to analyse this data, since they can provide more accurate results than manual analyses when it comes to identifying bioaerosol types using spectral data (Huffman et al., 2019). In previous studies, both unsupervised and supervised methods, ranging from k -means clustering to artificial neural networks, have been used with varying results. Ruske et al. (2017) evaluated a variety of algorithms for classifying laboratory generated aerosol particles, and were able to distinguish between broad particle classes (pollen, bacteria, fungal spores, and non-fluorescent material) with high accuracy. They found that supervised methods performed better than unsupervised methods, and that gradient boosting and neural networks achieved the highest performance out of the tested algorithms (Ruske et al., 2017).

When sampling ambient aerosol data, the true class labels of the particles are not known. Hence, we have to rely on laboratory-generated data of known particles as training data. The performance of the classification models will be highly dependent on the quality of the training data (Ruske et al., 2017) as well as how the training samples were generated, e.g., what aerosolisation method was employed (Huffman et al., 2019). When we apply trained models to ambient aerosol data, how certain can we be that the classification is correct? And how will a supervised learning model classify types of aerosol particles that were not present in the training data?

In this study, inductive conformal prediction (Vovk et al., 2005) will be used for uncertainty quantification of a particle classifier, and the behaviour of the conformal predictor on previously not observed particle types will be investigated. Aerosol particles in the atmosphere can come from a multitude of sources, and it would not be feasible to produce laboratory data for every possible aerosol material for training a predictive model. Hence, a model trained on laboratory data and later applied to ambient aerosol data will almost certainly encounter types of material that were not present in the training, and it is therefore important to characterise the behaviour of the predictor under such circumstances. In principle, an individual aerosol particle can also consist of a mixture of materials, but it is beyond the scope of this study to investigate what effects this could have on the classification.

In the next section, we describe the dataset, the conformal predictor, including the underlying predictive model, and the experimental procedure. In section 3, we present results from applying the conformal predictor to both unmodified data (including all particle types) and to data where certain classes of particles have been excluded from the training

set, but included in the test set. Finally, in section 4, we discuss the main findings and outline directions for future work.

2. Method

2.1. Dataset

We have used the dataset from [Ruske et al. \(2017\)](#), which contains single-particle measurements of different kinds of materials, e.g. dust, salt, pollen, bacteria, and fungal spores, that were aerosolised¹ in the laboratory and subsequently sampled with a multiparameter bioaerosol sensor (MBS). The dataset contains 10 subclasses of particles, that can be grouped into four parent classes: *bacteria*, *fungal spores*, *pollen*, and *non-biological* particles (see Table 1). The bacteria and salts were aerosolised from liquids or suspensions while the other samples were aerosolised from dry material ([Ruske et al., 2017](#)). The *fungal* parent class only contains one subclass of particles, while the other parent classes contain three subclasses each. The parent classes are the ones that are used as class labels for the classification.

Table 1: Overview of the aerosol particle dataset. For more details see [Ruske et al. \(2017\)](#).

Subclass	Parent class	# samples	% of total
<i>BG</i> * spores (washed, in distilled water)	bacteria	1417	7.47
<i>BG</i> * spores (unwashed, in L-broth)	bacteria	1831	9.66
<i>E. coli</i> ** cells (unwashed, in L-broth)	bacteria	1991	10.50
Puffball spores	fungal	2607	13.75
Fuller’s earth	non-biological	3238	17.08
1%NaCl (aqueous solution)	non-biological	4502	23.75
Phosphate-buffered saline	non-biological	1388	7.32
Aspen pollen	pollen	466	2.46
Poplar pollen	pollen	469	2.47
Paper mulberry pollen	pollen	1048	5.53

Bacillus atropheus* *Escherichia coli*

For each measured aerosol particle, the MBS records the size, the time of flight, the fluorescence in 8 different wavelength channels, and the light scattering pattern (related to particle shape; recorded using two linear detector arrays yielding 512 shape features each). This results in a dataset with 1034 potential features to use for classification. As an alternative to using all 1024 shape measurements, the instrument provides various statistics calculated from the light scattering patterns, e.g., variance, skewness, kurtosis per array, which, if used instead, reduce the number of shape features to 22. More details about the MBS and the dataset can be found in [Ruske et al. \(2017\)](#).

Not all of the 1024 shape pixels carry information; the edges of the detector arrays essentially measure only noise after background subtraction. To prevent their models from learning noise rather than actual particle features, [Ruske et al. \(2017\)](#) redid the background

1. Dispersed as an aerosol by means of, e.g., a nebuliser or compressed air

subtraction so that all data samples got the same noise peaks. We investigated whether it would have been enough to remove the array pixels that only contained noise; however, this was not the case. The experimentation also showed that the use of the shape statistics (22 features) resulted in virtually the same accuracy as using the full array data (1024 features) based on five-fold cross-validation. Using the shape statistics comes with several benefits: it reduces the number of features relative to the number of training samples, it appears robust to the noise issue, and it should be more easily generalisable to other MBS instruments, as different instruments do not necessarily have exactly the same shape pixel alignment. In addition, the statistics introduce domain knowledge that the classifier would not have otherwise, because the statistics describe the signal distribution in the detection arrays rather than treating each pixel as an isolated feature.

2.2. Algorithms

2.2.1. CONFORMAL PREDICTION

Conformal prediction concerns predicting with confidence (Vovk et al., 2005). In the case of classification, a conformal predictor returns a set of class labels rather than a single label, where we have $1-\sigma$ confidence that the true class label is included in the output set (for some significance level σ), assuming that the instances are independent and identically distributed (IID). Conformal prediction can be used together with any underlying predictive model. It also requires a non-conformity function C , which is used to assign a score of "strangeness" to each instance and class label pair (\mathbf{x}, y) . In principle, this can be any function, but it is usually related to the predicted probabilities from the underlying model. An example, which is the function used in this study, is

$$C(\mathbf{x}, y) = P_{(\mathbf{x}, y)} - 1 \quad (1)$$

where $P_{(\mathbf{x}, y)}$ is the predicted probability of the underlying model that the instance \mathbf{x} belongs to class y . This score ranges from -1 to 0 , where lower scores indicate that the instance is more "strange".

In inductive conformal prediction, the training data are split into a proper training set, used to train the underlying model, and a calibration set used to calculate the reference non-conformity scores. The non-conformity scores of the calibration instances are only calculated relative to their true class labels. Each new instance to be classified can then be given a score for each possible class label, and those scores are then compared to the calibration scores to assess how strange it would be if the instance belonged to the class in question. Instances can either be compared to the full calibration set, or we can choose to have one calibration set per class (a *Mondrian* approach; Vovk et al. (2003)). The p -value for an instance-label pair, $p_{(\mathbf{x}, y)}$, is determined by the fraction of calibration instances that are more than or equally non-conforming as this instance. The prediction for the instance at significance level σ is a (potentially empty) set of class labels $\{y_1, \dots, y_n\}$ where each label y_i fulfils $p_{(\mathbf{x}, y_i)} > \sigma$.

2.2.2. UNDERLYING MODEL

In this study, a gradient boosting classifier, as implemented in the Python package *scikit-learn* (v. 0.21.3) (Pedregosa et al., 2011), is used as the underlying model, since it was one

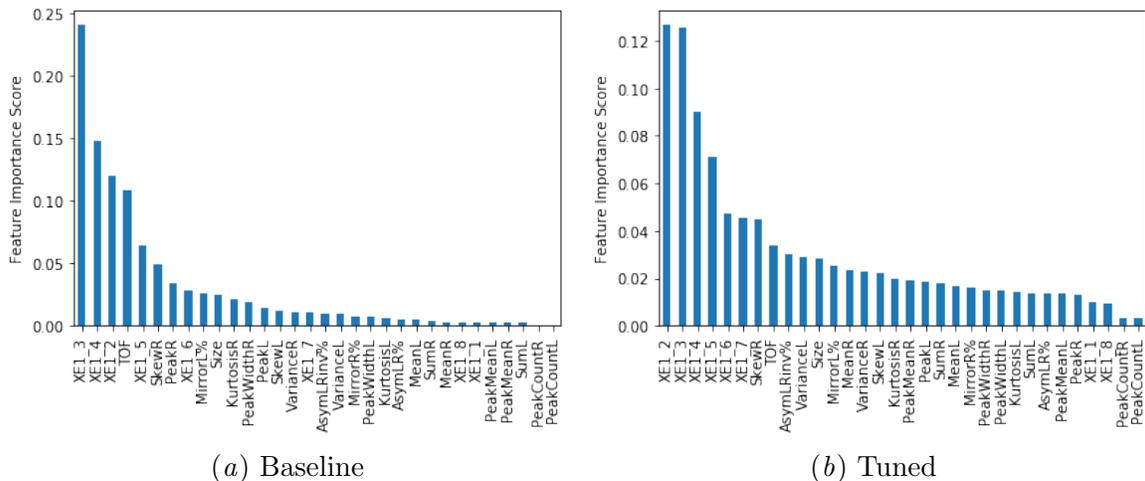


Figure 1: Feature importance scores for the gradient boosting classifier with (a) default hyperparameters (baseline model) and (b) tuned hyperparameters. Note the different y-axis scales.

of the classifiers that were reported to work best on this dataset (Ruske et al., 2017). With default parameter settings, it was possible to reproduce the previously reported results. The hyperparameters were further tuned following Jain (2016), which resulted in slightly higher accuracy and AUC scores. The parameter tuning also led to the feature importances of the model being more evenly balanced (see Fig. 1) which is expected to be beneficial since this lets the algorithm use more of the available information (Jain, 2016). Maximising the accuracy of the underlying model does not necessarily produce a more efficient conformal predictor (Johansson et al., 2013); however, in this case the tuned model did appear to improve the efficiency slightly.

2.3. Procedure

The dataset was split into proper training, calibration, and test sets at a ratio of 2:1:1. Using 50% of the dataset for training produced accuracies comparable to Ruske et al. (2017) (who used the full dataset), while at the same time leaving enough data for calibrating and testing the model. The dataset is quite unbalanced, so a stratified split was used. The parent classes from Table 1 (see above) were used as the class labels for training the model. A Mondrian approach was used to guarantee the error level for each parent class (Vovk et al., 2003; Shi et al., 2013) (cf. Fig. 2).

As the aim of the study is to investigate how unknown particle types would be classified, the following experimental procedure was employed: the model was re-trained and re-calibrated after removing one subclass at a time, and then evaluated on the part of the test set containing the removed subclass. Strictly speaking, this violates the IID assumption because the test instances are not sampled from the same distribution as the training and calibration instances. While this means that the error level can no longer be guaranteed, it does not mean that the p -values become meaningless — we should still be able to make a qualitative assessment of the degree of confidence in the model’s predictions.

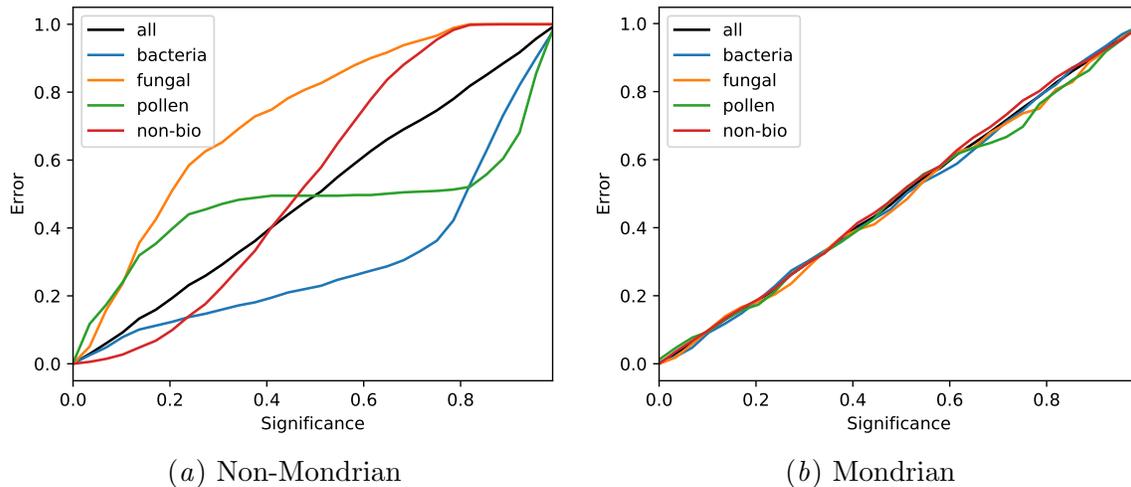


Figure 2: Validity of the inductive conformal predictor for the individual parent classes using (a) non-Mondrian calibration and (b) Mondrian calibration.

In conformal prediction, the error is normally the proportion of prediction sets that do not contain the true class label. However, in our case, if the instance to be classified belongs to an unknown parent class, an empty prediction set should be considered correct. If the instance belongs to a known parent class but an unknown subclass, e.g., a new type of bacteria, then both an empty set and a set containing the presumed parent class could be considered correct.

3. Results

3.1. The inductive conformal predictor

In a well-calibrated inductive conformal predictor (ICP), the error rate is close to the chosen significance level and the accuracy of the singleton predictions (OneAcc) is higher than the accuracy of the underlying model (Johansson et al., 2013). Figure 3a shows that the resulting ICP is valid. However, the multi-class prediction rate (MultiC) decreases fairly slowly, which suggests that the dataset is quite difficult and that some particle classes are perhaps not very distinct from each other (as far as the model can tell). A confusion matrix (Fig. 3b) shows that the *bacteria* and *non-biological* classes can be predicted with high precision by the underlying model, while the *fungal* and *pollen* classes are more difficult.

3.2. Classifying unknown particle types

As described above, each particle subclass was predicted (with respect to the parent classes) using the complete model (in Fig. 3) as well as using a model where the subclass had not been included in the training or calibration sets of the ICP; that is, a model for which the subclass was an unknown particle type. Table 2 shows how the error and efficiency scores of the predictor changed for each subclass. Now, we have to be careful since (a) the validity is only guaranteed for samples from the *parent class* distribution (not necessarily individual

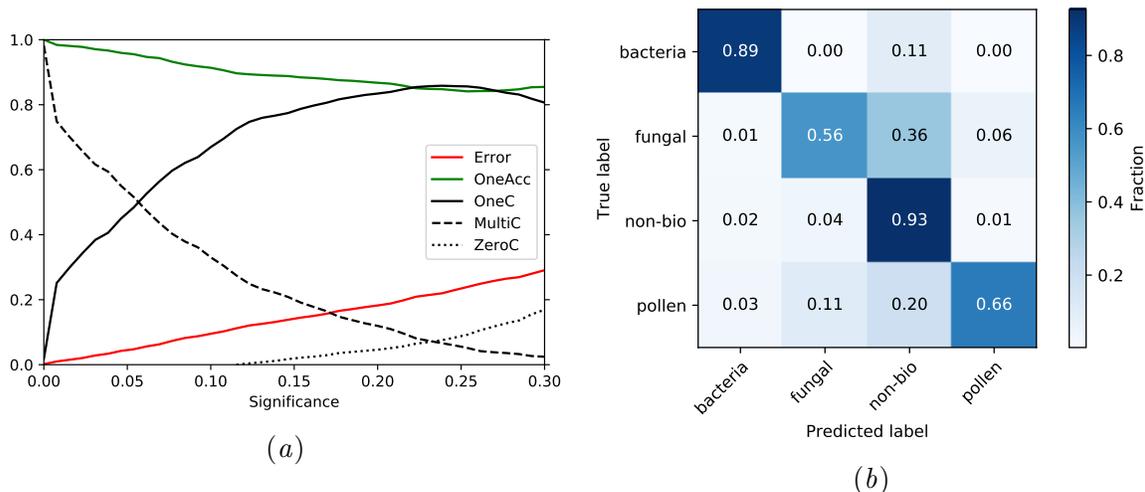


Figure 3: (a) Characteristics of the inductive conformal predictor trained, calibrated and tested on complete (unmodified) dataset. Error: fraction of output sets that do not contain the correct class label. OneAcc: accuracy of the singleton predictions. OneC: fraction of singleton predictions. MultiC: fraction of multi-class prediction sets. ZeroC: fraction of empty prediction sets. (b) Confusion matrix of the predictions from the underlying model, normalised with respect to the true labels.

subclasses) and (b) the error level is defined differently in the two cases (since rejection of all parent classes can be considered correct when the subclass is unknown). Nevertheless, there are some patterns that can be seen.

For all subclasses except one, the number of empty prediction sets (ZeroC) increased slightly. This is encouraging, since it suggests that the ICP is able, at least to some extent, to realise that the particle type is unknown. The particle type *puffball* had the highest ZeroC change, which could be expected because it was the only type of fungal spore in the dataset. However, its MultiC decreased and its singleton prediction rate (OneC) increased, which suggests that it was often confused for another class. Similar patterns can be seen for *aspen pollen* and *poplar pollen*. In fact, the pollen and fungal classes had the lowest precision of the four classes, and confusion matrices suggest they were often confused with non-bioaerosol (cf. Fig. 3).

Except for the case of *phosphate-buffered saline*, a small change in error rate seems to correlate with a small change in the other efficiency measures. Although the error has been defined differently when a subclass is unknown, a small change across the board still suggests that the predictions have not changed much. In other words, the ICP also appears to be able to correctly classify unknown subtypes of a known parent class in some cases. The *bacteria* class looks particularly consistent.

Four subclasses show a relatively large increase in the error: *fuller's earth*, *puffball*, *aspen pollen* and *poplar pollen*. Interestingly, these are also subclasses for which the fuzziness of the predictions is significantly *reduced* when they are removed from training and calibration (the opposite behaviour is seen for the other subclasses; cf. Fig. 4). Predictions of those four classes were also the fuzziest of all subclasses for the full model, which reflects the precision

Table 2: How the conformal predictor performance changed for each subclass when it was removed from the training and calibration, evaluated at significance level 0.2. Negative values are coloured red and values with an absolute change > 0.1 are bolded.

Subclass	Error	OneAcc	OneC	MultiC	ZeroC	AvgC
<i>BG</i> spores (washed)	0.011	-0.090	-0.085	0.011	0.073	-0.062
<i>BG</i> spores (unwashed)	-0.041	-0.030	-0.013	0.004	0.009	-0.004
<i>E. coli</i> (unwashed)	-0.048	-0.078	-0.012	0.000	0.012	-0.012
Puffball spores	0.632	-0.786	0.046	-0.224	0.178	-0.402
Fuller’s earth	0.317	-0.309	-0.012	-0.014	0.026	-0.043
1%NaCl	0.015	-0.094	-0.121	0.070	0.051	0.022
Phosphate-buffered saline	-0.052	-0.037	-0.118	0.003	0.115	-0.112
Aspen pollen	0.172	-0.083	0.138	-0.147	0.009	-0.155
Poplar pollen	0.333	-0.288	0.231	-0.282	0.051	-0.333
Paper mulberry pollen	-0.011	0.007	-0.042	0.046	-0.004	0.050

scores in the underlying model. In the case of *aspen pollen* and *poplar pollen*, it is worth noting that their natural size range is outside the instrument size range, so the samples in this dataset are pollen fragments, whereas intact *paper mulberry* particles may be within the MBS size range, which might explain the lower precision/higher fuzziness for the aspen and poplar subclasses (Simon Ruske, personal communication 2020-03-29).

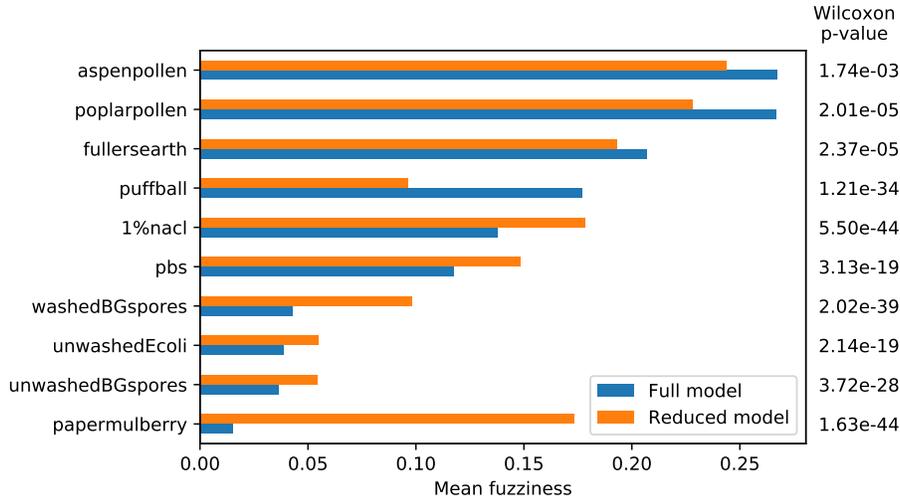


Figure 4: Fuzziness (average of the sum of all p -values except the highest) of the conformal predictor for the different subclasses. Full model means a model trained and calibrated on all subclasses, reduced model means the models trained and calibrated without each respective subclass. The Wilcoxon p -value is included to indicate if the fuzziness is significantly different when the subclass is not included in training compared to when it is.

In an attempt to better understand the behaviour of the ICP, an hierarchical agglomerative clustering (HAC) on the subclass means (Fig. 5) was performed as a quick way to

illustrate the relationship between the different particle types. A couple of things become apparent: 1) the *pollen* and *non-bioaerosol* classes are quite heterogeneous, and 2) *paper mulberry* is very different from all the other particle types.

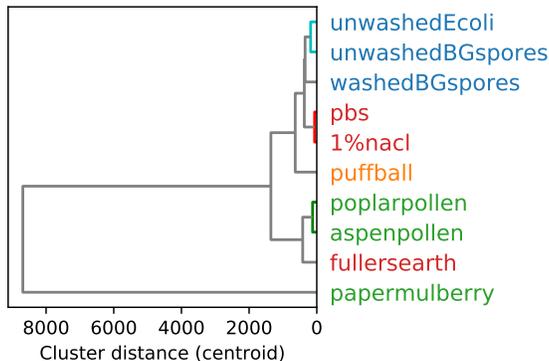


Figure 5: Dendrogram of centroid-linkage hierarchical agglomerative clustering where the initial clusters were represented by the mean of each subclass. The colours of the labels indicate which subclasses were grouped together in this project (cf. Table 1).

When dealing with heterogeneous classes, we encounter the same issue as with a non-Mondrian calibration — the error levels can be quite different for each subclass. This can be seen for the *pollen* class in Fig. 6a. In a more homogeneous class, like *bacteria* in this dataset, the individual errors are more similar (Fig. 6b), and we already saw in Table 2 that subclasses of bacteria could be predicted reasonably well even if they had been excluded from training and calibration. Why *paper mulberry* can be accurately predicted even when it was excluded from the training data is more perplexing, but probably it is different in such a way that it is similar to extreme samples of the other pollens (paper mulberry is generally the largest and most fluorescent of the particle types in the dataset).

The heterogeneity of the classes is one factor that can help explain the behaviour of the ICP but, since the dataset is unbalanced, the number of samples likely also plays a role. Because a Mondrian approach was used, the calibration sets of the minority classes will also be smaller which could make the predictions more fuzzy. For example, the pollen class is the one with the smallest number of samples (see Table 1). However, the non-bioaerosol class, which includes *fuller’s earth*, has the highest number of samples, so the heterogeneity and the error distribution within the parent classes are still important.

Some particle subclasses in the dataset were aerosolised from liquids or suspensions, namely the bacteria samples and the sodium chloride and phosphate-buffered saline samples, while the other samples were generated from dry material (Ruske et al., 2017). The particles were not passed through a dryer prior to being sampled by the MBS instrument (Ruske et al., 2017). From a measurement technical point of view, it is interesting to note that the wet samples cluster together in Fig. 5, which may indicate that the generation method or the sampling method (i.e. drying or no drying), or a combination of the two, impacts the measured properties of the particles, as suggested by previous studies (e.g. Huffman et al., 2019; Wang et al., 2010).

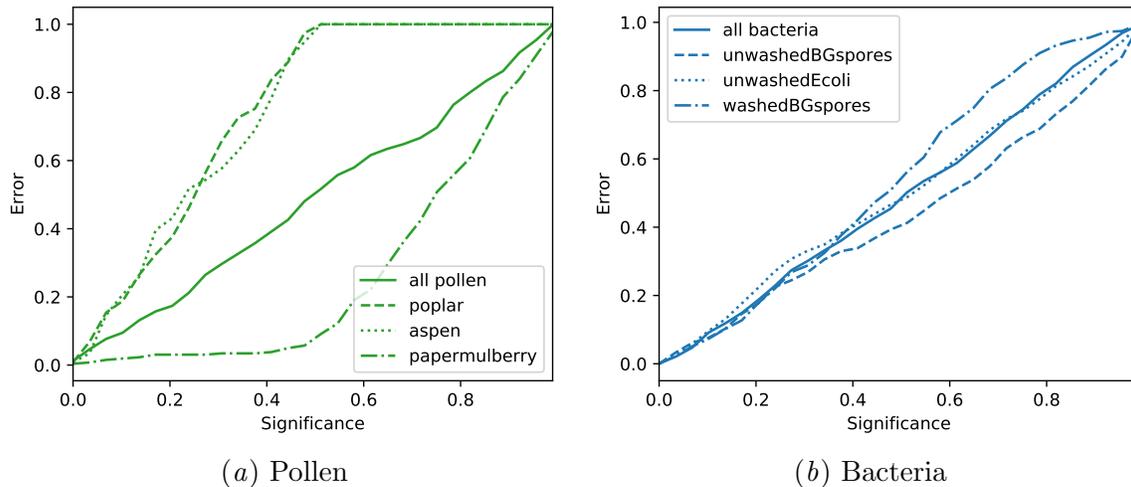


Figure 6: Validity of the inductive conformal predictor for (a) pollen (an example of a heterogeneous class) and (b) bacteria (an example of a homogeneous class).

4. Discussion

An investigation of how ICP deals with instances that belong to a class that was not included in the calibration or the training of the underlying model has been presented. Evaluating and interpreting the results has not been straightforward, since technically the premise of the exercise violates the IID assumption behind conformal prediction. That is, the ICP cannot be expected to be valid for unknown particles unless they are drawn from the same underlying distribution. Strictly speaking it cannot be expected to be completely valid for separate subclasses even if they were included in the training set, since the validity is guaranteed for the parent classes, so we are left with a similar issue as the one we try to solve by using a Mondrian approach (cf. Figs. 2 & 6). Even so, it appears that the ICP is close to valid and gives reasonable predictions for unknown particle types *if* they are similar to the particles that the ICP was trained and calibrated on, e.g. as for the relatively homogeneous bacteria class, which makes sense because this case is closer to IID. The ICP also showed signs of being able to reject unknown particles at a higher rate (increased ZeroC) for almost all subclasses.

Apart from the IID assumption, there are several other limitations to this study. The dataset itself is quite difficult to predict, due to an imbalance between classes both in terms of number of samples and heterogeneity within the parent classes. Although the AUC was used as a metric when training the model to try to account for the imbalance, it might have been easier to interpret the results if a better balance was achieved for the precision of the different classes. We only tried one underlying model (gradient boosting), but it is possible that another algorithm, such as a neural network (which also performed well on this dataset previously (Ruske et al., 2017)), may have performed better in combination with the ICP than the gradient booster did. However, the decision path in the underlying model probably changes slightly between the different training/calibration set configurations, which also complicates the comparison. To further study conformal prediction for unknown classes

and establish proper evaluation metrics, it may be instructive to use a less complex dataset and to use a simpler and more interpretable underlying model, such as a decision tree, to be able to examine how the decision path is affected. Other possible directions for future work include using Venn predictors (e.g. [Johansson et al., 2019](#)) to obtain intervals for class probabilities, thus allowing for cost–utility type analyses which may have useful applications within the air quality and health branch of aerosol science.

The ICP presented here showed some promise in being able to handle unknown particle types, and studying its application to field data or data collected with another MBS instrument would be interesting for future work. Preliminary results from applying the ICP to laboratory-generated sea spray aerosol (expected to be mostly salt particles with few particles containing biological material) measured with another MBS instrument revealed that, at a significance level of 0.1, none of the prediction sets included the non-bioaerosol class (and none were empty). The model’s failure to recognise salt particles might be because these sea spray particles were generated differently than the salt particles in the main dataset, and they were passed through a dryer prior to being sampled. The aforementioned factors affect the size and morphology of the particles, which in turn affect the magnitude of both the fluorescence signal and the shape signal. Part of the issue might be solved by normalising the data with respect to particle size but, since size is often not the only factor, this is impossible without further information about individual particles ([Huffman et al., 2019](#)). Standardising the sampling set-up so that the training data are collected in the same manner as the data intended to be classified is most likely necessary in future work if the classifier is to be applied to other datasets.

Acknowledgments

We would like to acknowledge the valuable input from Simon Ruske, Martin Gallagher and David Topping (The University of Manchester, U.K.) and would like to thank them for providing the data as described in [Ruske et al. \(2017\)](#). This work was partially financed by the Knut-and-Alice-Wallenberg Foundation within the ACAS project (Arctic Climate Across Scales, project no.2016.0024) and the Swedish Foundation for Strategic Research within the CDA project(Continuous Deep Analytics, project no. BD15-0006).

References

- J. Alex Huffman, Anne E. Perring, Nicole J. Savage, Bernard Clot, Benoît Crouzy, Fiona Tummon, Ofir Shoshanim, Brian Damit, Johannes Schneider, Vasanthi Sivaprakasam, Maria A. Zawadowicz, Ian Crawford, Martin Gallagher, David Topping, David C. Doughty, Steven C. Hill, and Yongle Pan. Real-time sensing of bioaerosols: Review and current perspectives. *Aerosol Science and Technology*, 0(0):1–31, sep 2019. ISSN 0278-6826. doi: 10.1080/02786826.2019.1664724.
- Aarshay Jain. Complete machine learning guide to parameter tuning in gradient boosting (GBM) in Python, Feb 2016. URL <https://www.analyticsvidhya.com/blog/2016/02/complete-guide-parameter-tuning-gradient-boosting-gbm-python/>.

- Ulf Johansson, Henrik Bostrom, and Tuve Lofstrom. Conformal Prediction Using Decision Trees. In *2013 IEEE 13th International Conference on Data Mining*, pages 330–339. IEEE, dec 2013. ISBN 978-0-7695-5108-1.
- Ulf Johansson, Tuve Lofström, Henrik Linusson, and Henrik Boström. Efficient venn predictors using random forests. *Machine Learning*, 108(3):535–550, 2019.
- F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- Simon Ruske, David O. Topping, Virginia E. Foot, Paul H. Kaye, Warren R. Stanley, Ian Crawford, Andrew P. Morse, and Martin W. Gallagher. Evaluation of machine learning algorithms for classification of primary biological aerosol using a new UV-LIF spectrometer. *Atmospheric Measurement Techniques*, 10(2):695–708, 2017. ISSN 18678548. doi: 10.5194/amt-10-695-2017.
- Simon Ruske, David O Topping, Virginia E Foot, Andrew P Morse, and Martin W Gallagher. Machine learning for improved data analysis of biological aerosol using the WIBS. *Atmospheric Measurement Techniques*, 11(11):6203–6230, nov 2018. ISSN 1867-8548. doi: 10.5194/amt-11-6203-2018.
- John H. Seinfeld and Spyros N Pandis. *Atmospheric Chemistry and Physics : From Air Pollution to Climate Change*. John Wiley & Sons, Incorporated, New York, 3 edition, 2016.
- Fan Shi, Cheng Soon Ong, and Christopher Leckie. Applications of Class-Conditional Conformal Predictor in Multi-class Classification. In *2013 12th International Conference on Machine Learning and Applications*, volume 1, pages 235–239. IEEE, dec 2013. ISBN 978-0-7695-5144-9. doi: 10.1109/ICMLA.2013.48.
- Vladimir Vovk, David Lindsay, Ilia Nouretdinov, and Alex Gammerman. Mondrian confidence machine. Workingpaper, 3 2003. On-line Compression Modelling project, <http://vovk.net/kp>, Working Paper 4.
- Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer-Verlag, Berlin, Heidelberg, 2005. ISBN 0387001522.
- Z. Wang, S. M. King, E. Freney, T. Rosenoern, M. L. Smith, Q. Chen, M. Kuwata, E. R. Lewis, U. Pöschl, W. Wang, P. R. Buseck, and S. T. Martin. The Dynamic Shape Factor of Sodium Chloride Nanoparticles as Regulated by Drying Rate. *Aerosol Science and Technology*, 44(11):939–953, sep 2010. ISSN 0278-6826. doi: 10.1080/02786826.2010.503204.