Sensor applications

1D-CNN Network Based Real-Time Aerosol Particle Classification With Single-Particle Mass Spectrometry

Guanzhong Wang[1], Heinrich Ruser[1]*, Julian Schade[2,3,5], Johannes Passig[3,4,5], Thomas Adam[2,4], Günther Dollinger[1], and Ralf Zimmermann[3,4,5]

[1]*Institute for Applied Physics and Measurement Technology, University of the Bundeswehr Munich, 85577 Neubiberg, Germany*
[2]*Institute of Chemistry and Environmental Engineering, University of the Bundeswehr Munich, 85577 Neubiberg, Germany*
[3]*Joint Mass Spectrometry Centre, Chair of Analytical Chemistry, University of Rostock, 18059 Rostock, Germany*
[4]*Joint Mass Spectrometry Centre, Helmholtz Zentrum München, 85764 Neuherberg, Germany*
[5]*Department Life, Light & Matter, Interdisciplinary Faculty, University of Rostock, 18059 Rostock, Germany*
**Member, IEEE*

Manuscript received 24 August 2023; accepted 31 August 2023. Date of publication 14 September 2023; date of current version 25 October 2023.

Abstract—Single-particle mass spectrometry (SPMS) is a measurement technique that aims to identify the chemical composition of individual airborne aerosol particles (PM 1 or PM 2.5) in real time. One-dimensional (1-D) spectral data of aerosol particles generated by SPMS carry rich information about the chemical composition associated with the sources of the particles, e.g., traffic and ship emissions, biomass burning, etc. Accurate classification of aerosol particles is essential to understand their sources and effects on human health. This letter investigates the application of SPMS and 1-D-convolutional neural network (1D-CNN) in aerosol particle classification. The proposed 1D-CNN achieved a mean classification accuracy of 90.4% with 13 particle classes. According to the experimental results, the combination of SPMS and 1D-CNN enables real-time collection, analysis, and classification of airborne aerosol particles to be used for highly responsive automated air quality monitoring.

Index Terms—Sensor applications, 1D-convolutional neural network (1D-CNN), aerosol particle, real-time air quality monitoring, single-particle mass spectrometry.

I. INTRODUCTION

The mass spectrum of individual aerosol particles captured by single-particle mass spectrometry (SPMS) represents the intensity distributions of positive and negative ions, sorted according to their mass-to-charge ratios (m/z). These mass spectra can be considered as 1-D vectors, see Fig. 1. The possible source of particles can be assessed by the different combinations of chemical components in the mass spectra. Traditional unsupervised classification methods such as ART-2a [1], [2] and k-means [3], [4] rely on manually postprocessing (selecting and merging) of their clustering results, causing the classification to be performed offline and time-consuming. Additionally, the amount of data can be extremely large: SPMS instruments have the capacity to analyze the chemical composition of up to some thousand particles per minute. With conventional, partially manual clustering methods, such scale of data generated during a continuous measurement campaign of several weeks or months is very difficult to handle and an analysis in real time is unfeasible. For the automated classification of SPMS data, several supervised methods were attempted, such as random forest [5], support vector machine [6], and multilayer perceptron [6]. However, the selection of hand-crafted features relies distinctively on field expertise [5].

In contrast, convolutional neural networks (CNNs) as architectures for deep learning algorithms are capable of automatically extracting meaningful features from the input data during the training process [7]. Originally, CNNs were specifically designed to process 2-D image data, in order to automatically learn to recognize patterns in complex data structures. Some studies, however, have successfully applied CNN to classify also 1-D data, like vibration signals for real-time machinery fault detection [8], [9] or electrocardiogram signals to monitor heart activities [10], [11]. In comparison, to the best of our knowledge, there is no reference in the literature for the classification of aerosol particle mass spectra using CNN. In order to apply CNN on the 1-D spectra, we treat each element in the 1-D vector as an image pixel, thus obtaining a "long" image of one pixel in height. For training the 1D-CNN classifier, we created a benchmark dataset containing ~37 000 *in situ* collected particle samples and provided manual annotations dividing

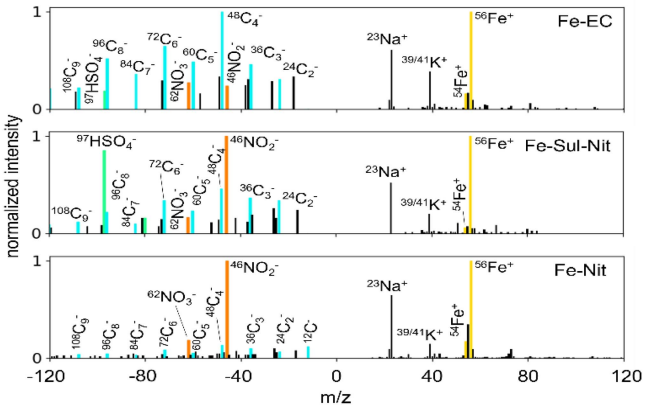

Fig. 1. Mass spectra of three subclasses of iron-containing particles (Fe-EC, Fe-Sul-Nit, and Fe-Nit). The ion markers with similar chemical properties are displayed in the same colors.

Corresponding author: Heinrich Ruser (e-mail: heinrich.ruser@unibw.de).
Associate Editor: K. Ozanyan.
Digital Object Identifier 10.1109/LSENS.2023.3315554

TABLE 1. Overview of the Dataset (# Means the Number of Samples) and the Most Significant Ion Markers for the Labeling Process

Class	#	Source	Ion markers
Elemental carbon (EC)	816	traffic emissions [2, 17]	***EC***: $^{12}C^{\pm}$, $^{24}C_2^{\pm}$, ..., $^{120}C_{10}^{\pm}$
Organic carbon - EC (OC-EC)	3383	biomass burning [2, 17]	***OC***: $^{27}[C_2H_3]^+$, $^{37}[C_3H]^+$, $^{39}[C_3H_3]^+$, $^{43}[C_4H_7]^+$, $^{51}[C_4H_3]^+$, $^{63}[C_5H_3]^+$, etc; ***EC***
K-rich	3300	biomass burning [2, 17]	$^{39/41}K^+$
Ca-rich	3238	lubricating oil of ship engines [2, 18]	$^{40}Ca^+$, $^{56}[CaO]^+$
V-rich	943	ship fuel emissions [18, 19]	$^{51}V^+$, $^{67}[VO]^+$; $^{54/56}Fe^+$; $^{60}Ni^+$
Mn-rich	2904	industrial emissions [3]	$^{55}Mn^+$
Fe-EC	3306	ship fuel emissions [18, 19]	***Fe***: $^{54/56}Fe^+$, $^{73}[FeOH]^+$; ***EC***
Fe-Nit	3050	ship fuel emissions [18, 19]	***Fe***; ***Nitrate***: $^{46}[NO_2]^-$, $^{62}[NO_3]^-$
Fe-Sul-Nit	2992	ship fuel emissions [18, 19]	***Fe***; ***Sulfate***: $^{80}[SO_3]^-$, $^{96}[SO_4]^-$, $^{97}[HSO_4]^-$; ***Nitrate***
Fe-Nit-EC	3451	ship fuel emissions [18, 19]	***Fe***; ***Nitrate***; ***EC***
Fe-dominant	3423	ship fuel emissions [18, 19]	***Fe***; negative signals are empty or very weak
Sea salt	3300	sea salt [2, 19]	***Salt***: $^{23}Na^{\pm}$, $^{39}[NaO]^+$, $^{62}[Na_2O]^+$, $^{63}[Na_2OH]^+$, $^{35/37}Cl^-$; ***Sulfate***; ***Nitrate***
Salt-Fe	3300	mixed state [19]	***Fe***; ***Salt***

the measured spectra into 13 classes, according to their significant ion markers; see Table 1. We developed a fast and highly accurate 1D-CNN network for automatic aerosol particle classification, to illustrate the potential for continuous online analysis of SPMS data.

II. METHODOLOGY

A. Single-Particle Mass Spectrometry

Several techniques of aerosol mass spectrometers can obtain chemical information from airborne particles in real time [12], [13]. Among these techniques, SPMS stands out for characterizing individual particles, thus revealing the mixing state of the particle ensemble. In SPMS, individual particles are hit by intense laser radiation that desorbs and ionizes at least a fraction of the particle within a single laser pulse (laser desorption/ ionization, LDI). This ionization technique is highly nonlinear; ionization efficiencies of particle compounds vary widely and depend also on the particle's main components and morphology. However, LDI produces sufficient ion numbers even from single nanoparticles and is effective for refractory compounds like metals [14]. The SPMS instrument, its working principle, and parameters have been described in detail previously [15]. Fig. 2 shows its basic design [16].

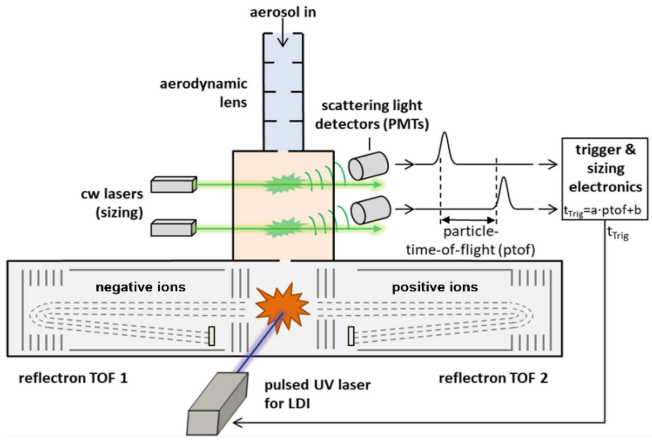

Fig. 2. Aerosol particles enter the SPMS through an aerodynamic lens and pass through a pair of cw laser beams. Based on the time-of-flight between the light scattering detectors, a size-dependent particle velocity is calculated and the arrival time for an individual particle in the ion source and trigger the UV laser for LDI is derived. Both positive and negative ions are recorded in a bipolar TOFMS setup. Particle sizes range from 150 nm (due to the wavelength of the sizing diodes) to 2.5 μm (due to the pulse energy). The parameters of the SPMS device were optimized to gain a high 50% hit rate.

B. Dataset

The analysis of the generated mass spectra provides insight into the chemical composition, mixing state, and likely origin of aerosol particles in the air. The data used in this study were obtained from a measurement campaign carried out from July 1 to August 3, 2022, close to the port of the coastal town of Rostock, Germany, and main shipping lanes, in order to investigate the composition and possible sources of aerosol particles, especially from the emissions of ship engines. Considering the relevance of iron-containing particles to ship emissions, we particularly focused on the aging degree of these particles, i.e., emission distance. Therefore, we classified iron-containing particles in more refined classes, while particles from other sources were not divided into subclasses.

We created a dataset containing 37 406 samples and manually labeled them into 13 different particle classes, representing the most abundant aerosol particles in the atmosphere at the collection site. In the process of manual labeling, supported by expert knowledge and verified twofold, we usually labeled particles based on the highest ion peak and major characteristic ion compositions in both the positive and negative ion mass spectrum. Table 1 lists the most significant ion markers for distinguishing different particles.

Exemplarily, Fig. 1 illustrates the mass spectra of several particles (containing 240 elements, with mass-to-charge ratios m/z from –120 to +120) which are challenging to differentiate, since it would be difficult to set thresholds for signal intensities to determine the presence or absence of certain combinations of ions. Moreover, some mass-to-charge ratios have multiple meanings, e.g., 39 for K^+, $[NaO]^+$ or $[C_3H_3]^+$, 51 for V^+ or $[C_4H_3]^+$, 56 for $[CaO]^+$ or Fe^+, etc. Therefore, the labeling of the particles based on the combination of ions and their intensities was done by expert analysis.

C. Network Architectures

Before feeding the data into the network, the spectrum data is preprocessed, separately normalizing the intensities to the maximum

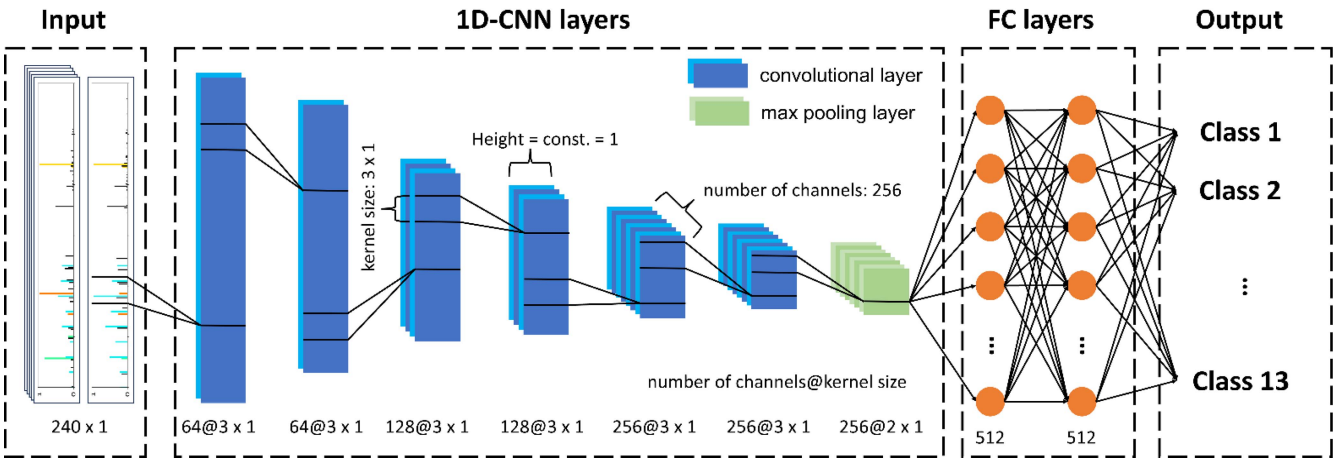

Fig. 3. 1D-CNN architecture. As input data serve the preprocessed and normalized 1-D mass spectra. Six convolutional layers are used to extract the local correlations within the data. Two fully connected (FC) layers perform the classification of the mass spectra into 13 classes.

TABLE 2. Average Classification Results Over the Five Folds on Testing Set

Accuracy	Precision	Recall
90.4 ± 0.9	0.90 ± 0.02	0.89 ± 0.02

intensities of the negative and positive ions. The best performing network hyperparameters were found by a grid search strategy. Fig. 3 illustrates the network architecture for the aerosol particle classification, which includes six convolutional layers to learn and extract the local correlations. The extracted features capturing the essential information for accurate classification will be imported to two fully connected (FC) layers, which will be activated by ReLU function and transferred to SoftMax layer for the classification. It is important to note that for 1D-CNN, the output of each layer (from the input layer to each convolutional, pooling and FC layer) is always a 1-D vector. For the training of the network we used the cross-entropy as loss function to measure the dissimilarity between the predicted and actual labels, which will be minimized by the Adam optimizer. Also by grid search, the optimal learning rate was set to start at 0.00001 and to decrease by 10% every 100 epochs, for a total of 500 epochs, and the batch size was set to 2048. During the training process, batch normalization and dropout techniques were used in the network to prevent overfitting.

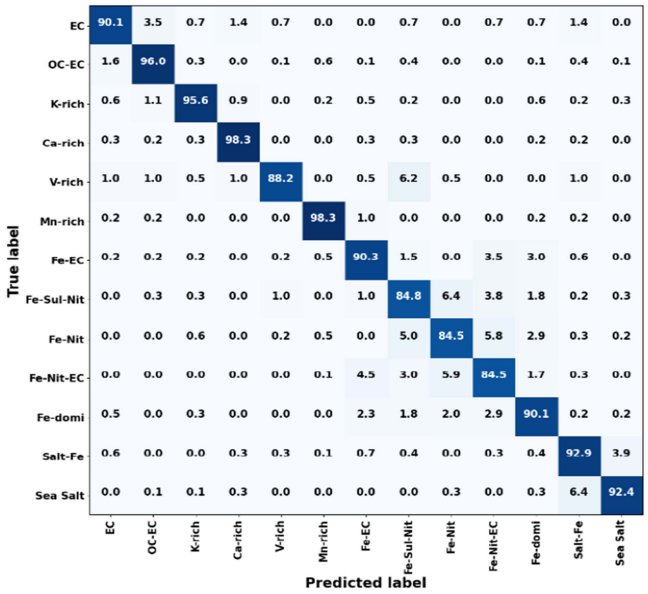

Fig. 4. Normalized confusion matrix displaying the rates (in percent) for the classification of ∼7500 mass spectra in testing set into 13 classes with the trained 1D-CNN classifier.

III. RESULTS AND DISCUSSION

To the best of our knowledge, there is no labeled dataset of atmospheric aerosol particles publicly available. Hence, we divided our dataset into two parts. 80% were used for cross validation (80% training and 20% validation), and the remaining 20% (not used during training and validation), for testing. Because our dataset is imbalanced, we used stratified five-fold cross validation to evaluate the performance of the network. Thus, the samples were randomly divided into five folds, four for training and one for validation, and the proportion of different classes in each fold is the same as the proportion of the whole dataset. The average performance (with accuracy, precision, recall, and confusion matrix as popular metrics) of the proposed network on the test set is presented in Table 2. The confusion matrix in Fig. 4 displays recognition rates above 84% for each of the 13 classes, with rates higher that 95% for four classes.

Cross validation, used to avoid overtraining of models on a particular dataset in experiments, improves the reliability of the models. In addition, we observed that using a downsampling pooling layer and activation function after each convolutional layer to reduce the dimension of the feature map did not help to improve the performance. The best performance was obtained using only one Max Pooling layer and ReLU activation function only once, after the last convolutional layers. Moreover, we obtained better results using smaller convolutional kernels (we have tested 3×1, 5×1, and 7×1 convolutional kernels).

The results show that the use of an 1D-CNN network to extract features in mass spectrum can accomplish automated and accurate classification of SPMS data. By contrast, a comparison with (unsupervised) clustering algorithms is not fair, since those methods depend heavily on manual postprocessing (i.e., selecting and merging the generated cluster centers). With our limited amount of labeled mass

spectra, traditional supervised ML methods (like SVM, RF, and MLP) [6] give similar results to 1D-CNN. Due to the unique ability of CNN to automatically extract and learn features, we focus on CNN only. In addition, we tested the robustness on a small extra dataset, and CNN showed significantly better accuracy than traditional ML methods. The important issue of robustness is still under development and will be published later.

To demonstrate the performance of the deep learning approach, 13 particle classes were defined from the measurement data. However, for many applications the aerosol particle classes are usually much larger in number. To enable the model to identify more particle classes, the annotated dataset will have to be continuously expanded. Furthermore, due to usually largely different abundances of various particles in the air, to create a balanced dataset is difficult and data augmentation methods need to be investigated to create synthetic data in order to counteract the effects of data imbalance on the models. Addressing these limitations and challenges will be critical to further improve performance and reliability.

IV. CONCLUSION

We proposed a 1D-CNN architecture to classify ambient air-based aerosol particles, measured by an SPMS instrument. Our results show for the first time, that 1D-CNN can be trained to efficiently learn and capture the relevant features from aerosol mass spectra for this challenging classification task. Using the real-time measurement capability of SPMS and the powerful feature extraction and classification abilities of deep learning, there is high potential to meet the demands for real-time analysis and classification of on-site collected data for the purpose of air quality monitoring. In principle, the classification capability is only dependent on the availability and quality of labeled mass spectra.

The results show very promising potential. The trained 1D-CNN accurately predicts the 13 particle classes proposed in this article. The long-term goal of using SPMS is to monitor the chemical composition and class of individual aerosol particles in the environment automatically and in real time. As the dataset of mass spectra from particles continues to largely expand during continuous observations, the proposed method paves the way for early warning and effective profiling of harmful pollutants.

ACKNOWLEDGMENT

This work was supported by dtec.bw — Digitalization and Technology Research Center of the Bundeswehr (Project "LUKAS") through European Union — NextGenerationEU.

REFERENCES

[1] X.-H. Song, P. K. Hopke, D. P. Fergenson, and K. A. Prather, "Classification of single particles analyzed by ATOFMS using an artificial neural network, ART-2A," *Anal. Chem.*, vol. 71, no. 4, pp. 860–865, Feb. 1999, doi: 10.1021/ac9809682.

[2] M. Dall'Osto and R. Harrison, "Chemical characterisation of single airborne particles in Athens (Greece) by ATOFMS," *Atmospheric Environ.*, vol. 40, no. 39, pp. 7614–7631, Dec. 2006, doi: 10.1016/j.atmosenv.2006.06.053.

[3] J. Arndt et al., "Characterization and source apportionment of single particles from metalworking activities," *Environ. Pollut.*, vol. 270, Feb. 2021, Art. no. 116078, doi: 10.1016/j.envpol.2020.116078.

[4] R. M. Healy et al., "Sources and mixing state of size-resolved elemental carbon particles in a European Megacity: Paris," *Atmos. Chem. Phys.*, vol. 12, no. 4, pp. 1681–1700, Feb. 2012, doi: 10.5194/acp-12-1681-2012.

[5] C. D. Christopoulos, S. Garimella, M. A. Zawadowicz, O. Möhler, and D. J. Cziczo, "A machine learning approach to aerosol classification for single-particle mass spectrometry," *Atmos. Meas. Techn.*, vol. 11, no. 10, pp. 5687–5699, Oct. 2018, doi: 10.5194/amt-11-5687-2018.

[6] G. Wang et al., "Machine learning approaches for automatic classification of single-particle mass spectrometry data," EGUsphere, May 2023, pp. 1–21, doi: 10.5194/egusphere-2023-784.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. Assoc. Comput. Machinery*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[8] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *J. Sound Vib.*, vol. 388, pp. 154–170, Feb. 2017, doi: 10.1016/j.jsv.2016.10.043.

[9] X. Wang, D. Mao, and X. Li, "Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network," *Measurement*, vol. 173, Mar. 2021, Art. no. 108518, doi: 10.1016/j.measurement.2020.108518.

[10] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[11] J. Torres, J. D. Buldain, and J. R. Beltrán, "Arrhythmia detection using convolutional neural models," in *Proc. Int. Conf. Distrib. Comput. Artif. Intell.*, 2019, pp. 120–127.

[12] J. Laskin, A. Laskin, and S. A. Nizkorodov, "Mass spectrometry analysis in atmospheric chemistry," *Anal. Chem.*, vol. 90, no. 1, pp. 166–189, Jan. 2018, doi: 10.1021/acs.analchem.7b04249.

[13] K. A. Pratt and K. A. Prather, "Mass spectrometry of atmospheric aerosols—Recent developments and applications. Part II: On-line mass spectrometry techniques," *Mass Spectrometry Rev.*, vol. 31, no. 1, pp. 17–48, 2012, doi: 10.1002/mas.20330.

[14] J. Passig et al., "Resonance-enhanced detection of metals in aerosols using single-particle mass spectrometry," *Atmos. Chem. Phys.*, vol. 20, no. 12, pp. 7139–7152, Jun. 2020, doi: 10.5194/acp-20-7139-2020.

[15] J. Schade et al., "Spatially shaped laser pulses for the simultaneous detection of polycyclic aromatic hydrocarbons as well as positive and negative inorganic ions in single particle mass spectrometry," *Anal. Chem.*, vol. 91, no. 15, pp. 10282–10288, Aug. 2019, doi: 10.1021/acs.analchem.9b02477.

[16] J. Passig and R. Zimmermann, "Laser ionization in single-particle mass spectrometry," in *Photoionization and Photo-Induced Processes in Mass Spectrometry*, R. Zimmermann and L. Hanley, Eds., 1st ed. Hoboken, NJ, USA: Wiley, 2021, pp. 359–411, doi: 10.1002/9783527682201.ch11.

[17] R. M. Healy et al., "Source apportionment of $PM_{2.5}$ in Cork Harbour, Ireland using a combination of single particle mass spectrometry and quantitative semi-continuous measurements," *Atmos. Chem. Phys.*, vol. 10, no. 19, pp. 9593–9613, Oct. 2010, doi: 10.5194/acp-10-9593-2010.

[18] J. Passig et al., "Detection of ship plumes from residual fuel operation in emission control areas using single-particle mass spectrometry," *Atmos. Meas. Techn.*, vol. 14, no. 6, pp. 4171–4185, Jun. 2021, doi: 10.5194/amt-14-4171-2021.

[19] J. Passig et al., "Single-particle characterization of polycyclic aromatic hydrocarbons in background air in northern Europe," *Atmos. Chem. Phys.*, vol. 22, no. 2, pp. 1495–1514, 2022, doi: 10.5194/acp-22-1495-2022.