Black Carbon characterization with Raman spectroscopy and machine learning techniques: first results for urban and rural area

CEST 2023

18th International Conference on Environmental Science and Technology

Athens, Greece, 30 August to 2 September 2023

Black Carbon characterization with Raman spectroscopy and machine learning techniques: first results for urban and rural area

DRUDI L.[1,*], GIARDINO M.[2,3], JANNER D.[2,3], POGNANT F.[4], MATERA F.[4], SACCO M.[5], BELLOPEDE R.[1].

[1]Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

[2]Department of Applied Science and Technology (DISAT), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

[3]Consorzio Interuniversitario Nazionale per la Scienza e Tecnologia dei Materiali (INSTM), Via G. Giusti 9, 50121 Florence, Italy

[4]Evironment Direction, Regione Piemonte, Via Principe Amedeo, Turin, Italy

[5]Arpa Piemonte, North West Piedmont Department, Via Pio VII, 9, 10135, Turin, Italy

*corresponding author:
e-mail: lia.drudi@polito.it

Abstract

Among the chemical substances of Particulate Matter (PM), there is a considerable quantity of black carbon (BC), which is linked to adverse public health effects and climate change. This study aims to develop an innovative method for the source apportionment of BC inside the PM, using Raman spectroscopy and machine learning techniques.

Different BC sources, including biomass ashes and vehicle emissions, and different PM samples from air quality monitoring stations have been analyzed with a Raman spectrometer. The PM samples used in the present study are collected from two different locations: an urban environment (Turin, Italy) and an alpine valley context (Oulx, Italy).

To each obtained spectrum, which presents the characteristic G and D bands, a five-band fitting has been applied to gather information that can lead to the identification of the different BC sources. Machine learning techniques, including the K-Nearest Neighbors (KNN) algorithm, have been applied to calculate the cluster resolution through a value of accuracy. Finally, the same algorithm, trained on the BC emission sources' data, tries to associate each BC in the PM to its source. In particular, a large amount of BC from diesel engine car exhaust emissions is found in all the considered PM samples.

Keywords: Fossil fuel, Biomass burning, Source apportionment, D Band, G Band

1. Introduction

Black Carbon is a well-known pollutant produced fossil fuel, biomass and biofuels combustion. It is the second strongest contributor to global warming only after carbon dioxide (Ramanathan & Carmichael, 2008). Still the impact of BC on the climate is different from the other greenhouse gases: it has a short lifetime in the atmosphere (Bond et al., 2013), of about 1 week (Mingjiang et al, 2014), and it has the capacity to absorb solar radiation. This causes a modification of the atmospheric radiative properties resulting in a substantial warming climate effect. Furthermore, BC can affect the meteorological condition by modifying the air relative humidity causing a decrease in precipitation (Ramanathan & Carmichael, 2008). A secondary impact on the climate is the deposition of BC on ice and glaciers that reduce their albedo effect and consequently enhance their melting (Ramanathan & Carmichael, 2008) (Bond et al., 2013).

PM adverse effects are widely established and depend on size, chemical composition and concentration. Likewise, BC physical and chemical properties affect human health. In addition, it can also carry other toxic substances, like polycyclic aromatic hydrocarbons (PAHs), that cause significant negative effects on human health ranging from normal respiratory transient to cardiovascular problems, morbidity and mortality (Ali et al., 2021).

The most widely used optical instrument for BC measurements is the aethalometer which can estimate the concentration in real-time by measuring the light transmittance (Mingjiang et al., 2014). Furthermore, using different wavelengths, a model can be applied to perform a source apportionment to distinguish BC from biomass burning and fossil fuel combustion (Mousavi et al., 2019). It has been reported that biomass burning contributes to BC contributions more in rural and suburban areas than in urban centers; moreover, in the cold season the BC contribution is also higher due to the increase of wood burning from residential heating (Mousavi et al., 2019).

A new way to perform and improve the source apportionment of BC can be with the use of Raman

spectroscopy. In fact, in the last decade, Raman spectroscopy has been widely used to investigate the single particle composition of aerosol PM. In particular, this technique can provide an unambiguous identification of the different PM chemical species correlating their characteristic vibrational modes to the resulting Raman spectra (Doughly & Hill, 2020). Among the PM chemical substances, BC can be easily recognized by its darkness and the resulting Raman spectra, which, as most of the Carbon-based material, is characterized by two overlapped bands: the G-band (1580 cm^{-1}) and the D band (1360 cm^{-1}). The G band is due to the stretching vibration of pairs of sp^2 carbon atoms, while the D band arises when defects or impurities are present. (Russo & Ciajolo, 2015). The shape of these two bands, or their deconvolution, can provide information about the type of fuel used, the fuel/oxygen ratio, and the treatment temperature (Ge et al., 2019). Raman spectra combined with a statistical analysis method promises to be an effective tool for source apportionment of atmospheric BC particles.

The present study investigates, through Raman spectroscopy and machine learning techniques, the BC characteristics coming from different emission sources to develop a possible source apportionment tool. Hence, the goal is to identify the emission source for several BC particles within a PM sample and estimate their contribution quantitatively.

2. Materials and methods

The analysis is divided into two subsequent stages to evaluate the BC source emission in PM. In the first stage, samples from different BC emission sources are analyzed to collect representative spectra. Instead, in the second stage, PM samples are collected from two different air quality stations by the Regional Environmental Protection Agency (Agenzia Regionale per la Protezione Ambientale Piemonte) to assign at each BC particle its emission source.

Two distinct locations were chosen for the sampling to represent both urban and rural contexts to highlight the different BC sources. The first station is located in the center of the metropolitan city of Turin in "Piazza Conti di Rebaudengo" near a busy road and in an industrial and residential area. The city of Turin is a high-density urban center, the second largest in the Po valley, and one of the most polluted Italian cities. The poor air quality is due to the city itself and its orographic context. The Alps and hills that surround the city do not allow pollutant dilution. This phenomenon is enhanced in winter because this region (Po valley) is subject to the formation of an inversion layer resulting in a higher air pollution concentration (Pernigotti et al., 2012). The second location considered is the municipality of Oulx, in the Alps, 80 km from Turin, with a population of about 3000 people. The air quality station is placed near a state road in a residential area.

The measurements were performed in winter, during February, when the BC concentration is usually higher, mostly due to the increase in biomass burning for domestic heating (Mousavi et al, 2019).

2.1. Apparatus and measurement parameters

The BC and aerosol PM particles were analyzed using a Renishaw inVia Raman spectrophotometer coupled with a LEICA confocal microscope. Measurements were performed using a 532 nm diode pumped solid state (DPSS) laser, set to a power of 150 μW (0,5 % of the maximum power) to avoid beam damage of the sample, and focusing the beam via a 50x LWD objective (NA 0.4).

An 1800 lines/mm grating was used in static mode and positioned to cover the spectral range from 115 cm^{-1} to 1881 cm^{-1} is chosen. The acquisition time was set to 12 seconds and each spectrum was integrated over 5 accumulations.

2.2. Emissive sources studies: Single sources of BC and the two PM samples

BC in the PM is mainly produced by biomass burning and traffic. The considered urban sources of BC are:

- Wood/biomass ashes

- Soot from vehicle engine (gasoline, diesel, LPG and methane)

In Table 1 the number and typology of BC spectra analyzed both for BC emission sources and for BC particle from PM are shown.

Table 1. Number and typology of BC spectra

BC emission source	**Number of spectra**
Biomass burning	54
Gasoline vehicle emission	56
Diesel vehicle emission	63
LPG vehicle emission	60
Methane vehicle emission	59
BC particle from PM	**Number of spectra**
Rebaudengo	28
Oulx	28

2.3. Data analysis

Once the Raman spectra are obtained from both the source emissions and the PM sample, the following post-processing is applied: screening, smoothing, baseline removal and deconvolution.

For the screening process both measurements with high fluorescence and measurements that detect other species are excluded. Specifically, the presence of hematite's presence could alter D peak's intensity because it is characterized by a broad band at around a Raman shift of 1320 cm^{-1} (Doughly & Hill, 2020). Because of that, all the measurements that show other peaks (such as iron oxide and calcite) than the D and G bands are removed from this analysis.

The smoothing is performed using the Savitzky-Golay algorithm to improve the signal to noise ratio (SNR). The baseline removal is performed using a straight line

that connects the points at a Raman shift of 1000 cm^{-1} and 1778 cm^{-1}, as performed by Laumer et al. (2016).

The deconvolution is made using five bands (G, D1, D2, D3 and D4), as proposed by Sadezky et al. (2005), that is the combination of one Gaussian (D3) and four Lorentzian curves (G, D1, D2 and D4).

Each curve is characterized by the position and maximum intensity of the peak, the area and the Full Width at Half Maximum (FWHM).

Hence, for each measurement, 17 parameters are returned as a single value (position of the peak) and as ratio (intensity, area and FWHM) where the parameter associated with the G band is always at the denominator. (Feng et al, 2019). Two additional parameters (R2 and R3) are calculated as indicated by Feng et al. (2019).

2.4. Machine learning

Machine learning techniques have been applied to identify data patterns and carry out predictive analysis on PM samples. The written algorithm is supervised as the input data (parameter values) are already associated with the correct output (type of BC emission source). With ML it has been possible to recognize in a bi-dimensional diagram, which has the parameters as axes, clusters of the different BC emission sources. The algorithm used for the classification is the K-Nearest Neighbors (KNN) which allows the class determination of one data according to the mutual position. Specifically, the data class is assigned as the most common class from the K data around the original data. The value of K used is 3.

All the data are divided into two datasets to evaluate the cluster resolution: 80% of the data from the different BC emission sources as the training dataset and 20% as the test dataset. The training dataset is used to create with the KNN algorithm the prediction of every single value in the test dataset. Then, the predicted class is confronted with the actual class for each value in the test dataset. From this comparison, prediction accuracy is calculated in percentage points. The possible division between the training and test datasets can affect the accuracy; therefore, several casual divisions have been considered and an average accuracy is obtained.

3. Results

3.1. Analytical determination of accuracy

In order to train the algorithm, two different combinations have been explored. The first combination takes into account all the different BC emission sources. The highest accuracies obtained are equal to 74% and 70%. A reduction of the number of classes was carried out to improve the value of obtainable accuracy. For the Turin province, 90% of vehicles are powered by diesel and gasoline while LPG, methane (and electric) accounts only for 10% (ACI website). Hence, the second combination only considers three emission sources: gasoline and diesel vehicle emissions and biomass burning. The accuracy obtained is 90% for 6 different couples of parameters. As an example, two couples of parameters are shown in Figure 1.

3.2. Class prediction

Using the trained algorithm on all the data from the considered BC emission source, a class prediction can be obtained for each BC particle. For each combination, BC particle and couple of parameters, a class can be assigned. The class is assigned only if most of the couples considered indicate the same class; otherwise, the BC particle is classified as not identified.

In Table 2 and Table 3 the class prediction of BC for Turin and Oulx air quality stations obtained for the different algorithm prediction (Combination 1 and 2) are shown.

Table 2. Class prediction of BC in Turin Rebaudengo

Combination	Diesel	Biomass	Others	Not identified
1	61%	0%	0%	39%
2	71%	25%	0%	4%

Table 3. Class prediction of BC in Oulx

Combination	Diesel	Biomass	Others	Not identified
1	61%	0%	0%	39%
2	57%	35%	0%	7%

4. Discussion and conclusion

The results from the station of Turin Rebaudengo (Table 2) show a considerable amount of BC from diesel car emissions for both combinations. Concerning the second combination, the class of almost all the non-identified BC particles can be determined thanks to the higher accuracy gained during the algorithm training. The same consideration on the two different combinations can be applied to the result from Oulx (Table 3).

The high share of BC from Diesel vehicle emissions can be associated with the high traffic volume for both stations as they are placed in front of busy roads, especially for Turin Rebaudengo. Even if the percentage of diesel and gasoline vehicles is more or less the same for the Turin province (ACI website), it has been reported that the Diesel vehicles produce much more BC than gasoline vehicles (Bond et al., 2013). For that reason, the non-identification of gasoline vehicle emissions even for the second combination can be explained. As expected, the percentage of biomass is higher in the rural context compared to the urban context as expected (Mousavi et al., 2019). Furthermore, the percentage of biomass burning BC found can be coherent with the mean value of 33% obtained by applying the aethalometer model in the same winter in another station in the city of Turin (Ballato & Sacco, 2022).

Due to the climate and health effects of BC, providing a method to classify and investigate the source of atmospheric BC aerosol can be useful for obtaining effective information and applying meaningful air quality management . Raman spectroscopy and machine learning techniques show promising results for the BC source apportionment and can identify the different percentages of biomass and diesel vehicle emissions within a PM sample.

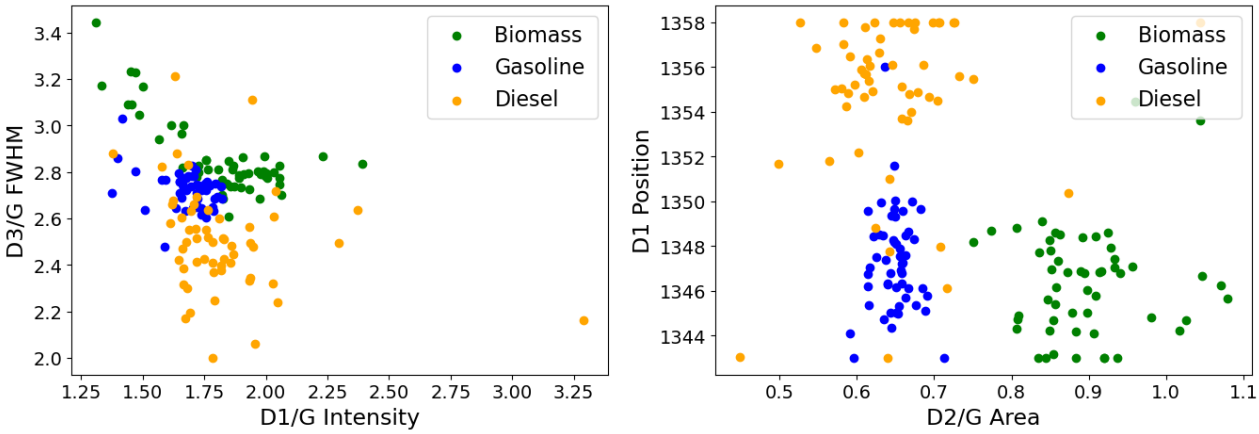

Figure 1. Couple of parameters with highest accuracy (combination 2)

References

ACI (2023), *Automobile Club d'Italia Open Parco Veicoli*, https://opv.aci.it/WEBDMCircolante/(Accessed: 20/03/23)

Ali M. U., Siyi L., Yousaf B., Abbas Q., Hameed R., Zheng C., Kuang X., Wong M. H. (2021), Emission sources and full spectrum of health impacts of black carbon associated polycyclic aromatic hydrocarbons (PAHs) in urban environment: A review, *Critical Reviews in Environmental Science and Technology*, **51**, 857-896.

Ballato E., Sacco M. (2022), Il black carbon in Piemonte (131-138), in *La qualità dell'aria in Piemonte*, Arpa Piemonte, https://www.arpa.piemonte.it/arpa-comunica/file-notizie/2022/rapporto-qaria-2021-lr.pdf

Bond T. C., Doherty S. J., Fahey D. W., Forster P. M., Berntsen T., DeAngelo B. J., Flanner M.G., Ghan S., Kärcher B., Koch D., Kinne S., Kondo Y., Quinn P. K., Sarofim M. C., Schultz M. G., Schulz M., Venkataraman C., Zhang H., Zhang S., Bellouin N., Guttikunda S. K., Hopke P. K., Jacobson M. Z., Kaiser J. W., Klimont Z., Lohmann U., Schwarz J. P., Shindell D., Storelvmo T., Warren S. G., C.S. Zender (2013), Bounding the role of black carbon in the climate system: A scientific assessment, *Journal of geophysical research: Atmospheres*, **118**, 5380-5552.

Doughty D. C., Hill S. C. (2020), Raman spectra of atmospheric aerosol particles: Clusters and time-series for a 22.5 hr sampling period, *Journal of Quantitative Spectroscopy and Radiative Transfer*, **248**, 106907.

Feng Y., Liu L., Yang Y., Deng Y., Li K., Cheng H., Dong X., Li W., Zhang L. (2019), The application of Raman spectroscopy combined with multivariable analysis on source apportionment of atmospheric black carbon aerosols, *Science of the Total Environment*, **685**, 189-196.

Ge H., Ye Z., He R. (2019), Raman spectroscopy of diesel and gasoline engine-out soot using different laser power, *Journal of Environmental Sciences*, **79**, 74-80.

Laumer J., Selvakumar R. S., O'Leary S. K. (2016), A Raman spectroscopic analysis of thin carbon films deposited onto curved Ti6Al4V substrates with and without silicon adhesion layers, *Diamond and Related Materials*, **70**, 59-64.

Mingjiang N., Jianxin H., Shengyong L., Xiaodong L., Jianhu a Y., Kefa C. (2014), A review on black carbon emissions, worldwide and in China, *Chemosphere*,**107**, 83-93

Mousavi A, Sowlat M. H., Lovett C., Rauber M., Szidat S., Boffi R, Borgini A., De Marco C., Ruprecht A.A., Sioutas C. (2019), Source apportionment of black carbon (BC) from fossil fuel and biomass burning in metropolitan Milan, Italy, *Atmospheric Environment*, **203**, 252-261.

Pernigotti D., Georgieva E., Thunis P., Bessagnet B. (2012), Impact of meteorological modelling on air quality: Summer and winter episodes in the Po valley (Northern Italy), *International Journal of Environment and Pollution*, **50**, 111-119.

Ramanathan V. and Carmichael G. (2008), Global and regional climate changes due to black carbon, *Nature Geoscience*, **1**, 221-227.

Russo C., Ciajolo A. (2015), Effect of the flame environment on soot nanostructure inferred by Raman spectroscopy at different excitation wavelengths, *Combustion and Flame*, **162**, 2431-2441.

Sadezky A., Muckenhuber H., Grothe H., Niessner R., Pöschl U. (2005). Raman microspectroscopy of soot and related carbonaceous materials: Spectral analysis and structural information, *Carbon*, **43**, 1731–1742.