Raman spectroscopy of carbonaceous particles of environmental interest

Anna Ferrugiari, Matteo Tommasini* and Giuseppe Zerbi

Introduction

In this paper, we deal with Raman spectroscopy, mainly applied as analytical tool, for the characterization of carbonaceous materials. We focus our interest on the atmospheric particulate of anthropogenic origin, which is an ubiquitous material found in atmospheric dust, smokes, car exhausts, deposited as powder or cohesive agglomerates on surfaces of any type, internalized in biological cells, and so on. Our work starts from the atmospheric particulate but wishes to lay the background for a wider application in the field of the characterization of carbonaceous materials in general. The anthropogenic atmospheric particulate can be approximately described as a material consisting of microparticles and nanoparticles made up by an inorganic fraction, an organic fraction soluble in organic solvents consisting mostly of relatively small aromatic molecules, and by an intractable and insoluble organic fraction (elemental carbon) associated to large polycyclic aromatic hydrocarbons (PAH) forming sheets of polycondensed benzene rings, which easily organize in ordered or disordered aggregates.[1] The experimental observation, variously accounted for in the literature (for instance, Sparkes *et al.*[2]), is that the Raman spectra of carbonaceous samples from different sources show non-negligible differences. The purpose of this work is to attempt to rationalize the Raman spectroscopic observations on many samples of carbonaceous particles of different origin in terms of the molecular properties of the graphitic components. Since many years, vibrational infrared spectroscopy has provided analytical and organic chemists with the 'infrared fingerprints of molecules'. A widespread procedure nowadays is to locate in the infrared spectrum absorption bands (or spectral patterns), which can be associated with molecular vibrations localized on specific chemical functional groups: books, tables, and data banks of 'infrared group wavenumbers' are on sale and are popular in laboratories.[3–5] In the past 10 years, the explosive technological and theoretical development of Raman spectroscopy is playing an exclusive new role in vibrational spectroscopy and starts providing complementary data on localized molecular motions so that data banks of Raman group wavenumbers started to appear, e.g. Socrates.[5] Moreover, for π-conjugated molecular systems with specific electronic and geometrical structures, the concept of collective vibrations involving the whole molecule (or large portions of it) is progressively being introduced by Raman spectroscopy. The Raman lines considered in this work are associated to collective molecular vibrations generated by carbon atoms in aromatic systems. The samples studied are mostly related to materials of interest in the works on cultural heritage and environmental science and range from simple commercial carbonaceous particles to standard particulate matter (PM) and to dusts and cohesive materials collected from various works of art in northern Italy.

The interest on the Raman spectrum of carbonaceous materials traces back to the work of Tuinstra and Koenig,[6] and it has grown to include many forms of graphitic carbon species of practical interest[7] and soots.[8] Because soot is ubiquitous, the Raman spectroscopy of carbon started to become a relevant research topic also in environmental science.[2,9–14] and it appeared as an analytical technique in cultural heritage[15] Another interesting application of Raman spectroscopy in the characterization of natural carbon has been also reported in geology: Raman spectroscopy allowed establishing a useful correlation with the maximum temperature reached during the formation of metasedimentary rocks embedding carbon.[16]

Our analysis of Raman data will begin by considering the effects of ball milling on pure microcrystalline graphite. This is a simple and effective way to introduce a controlled amount of defects in the samples, and it has been adopted many times in the literature.[17–24] These data form the basis for the interpretation of the Raman data on samples of environmental origin reported in the second part.

In this work, we wish to show how, by adopting statistical tools for data analysis and a molecular-based interpretation of the signals, one can employ Raman spectroscopy for distinguishing among carbonaceous samples of different origin.

*Correspondence to: Matteo Tommasini, Dipartimento di Chimica, Materiali e Ingegneria Chimica – Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano 20133, Italy.
Email: matteo.tommasini@polimi.it

Dipartimento di Chimica, Materiali e Ingegneria Chimica – Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano 20133, Italy

Raman signature of graphitic materials: the molecular origin of the G and D lines

We consider in this work the Raman spectra of environmental carbonaceous materials of different origins, and we limit our analysis to the first-order Raman spectra, which feature two characteristic lines, namely one near 1600 cm^{-1} (usually named as the G line) and one near 1300 cm^{-1} (usually named as the D line). The theoretical investigation of the G and D lines of graphitic materials (in particular graphene) has been and still is at the center of an evergreen interest driven by chemical methods, i.e., a molecular approach, which moves from PAHs as model compounds,[23,25–29] and by physical methods, mainly based on the theoretical techniques of solid-state physics.[30–32] The most unusual properties of the G and D lines, which required rationalization, are the following: (1) for a fixed laser excitation, the wavenumbers, intensities (absolute and/or relative), and widths of the G and D lines may change among different graphitic samples and (2) the D line shows a marked wavenumber and intensity dispersion with the wavelength of the excitation line. Necessarily, these changes are related to the molecular structures of the samples under study so that Raman spectroscopy becomes a valuable structural probe of graphitic materials.

In the following, we attempt to summarize the molecular interpretation of the G and D lines, which we think is more appropriate for the description of samples of graphitic materials lacking long-range crystalline order, as soots and carbonaceous particulate of environmental interest. Raman spectra of several PAH molecules generously shared with us by Prof. K. Muellen[27] have formed the background knowledge of our molecular approach,[28] which is applied in this work to the interpretation of the Raman spectra of carbonaceous (graphitic) particles of environmental interest. When PAH molecules are relatively small (i.e., generally, those which can be obtained by solvent extraction of PM particles), their infrared spectra show specific group wavenumbers, which allow their analytical characterization.[33] On the contrary, infrared spectroscopy does not give any contribution to the analytical chemistry of pure graphitic materials

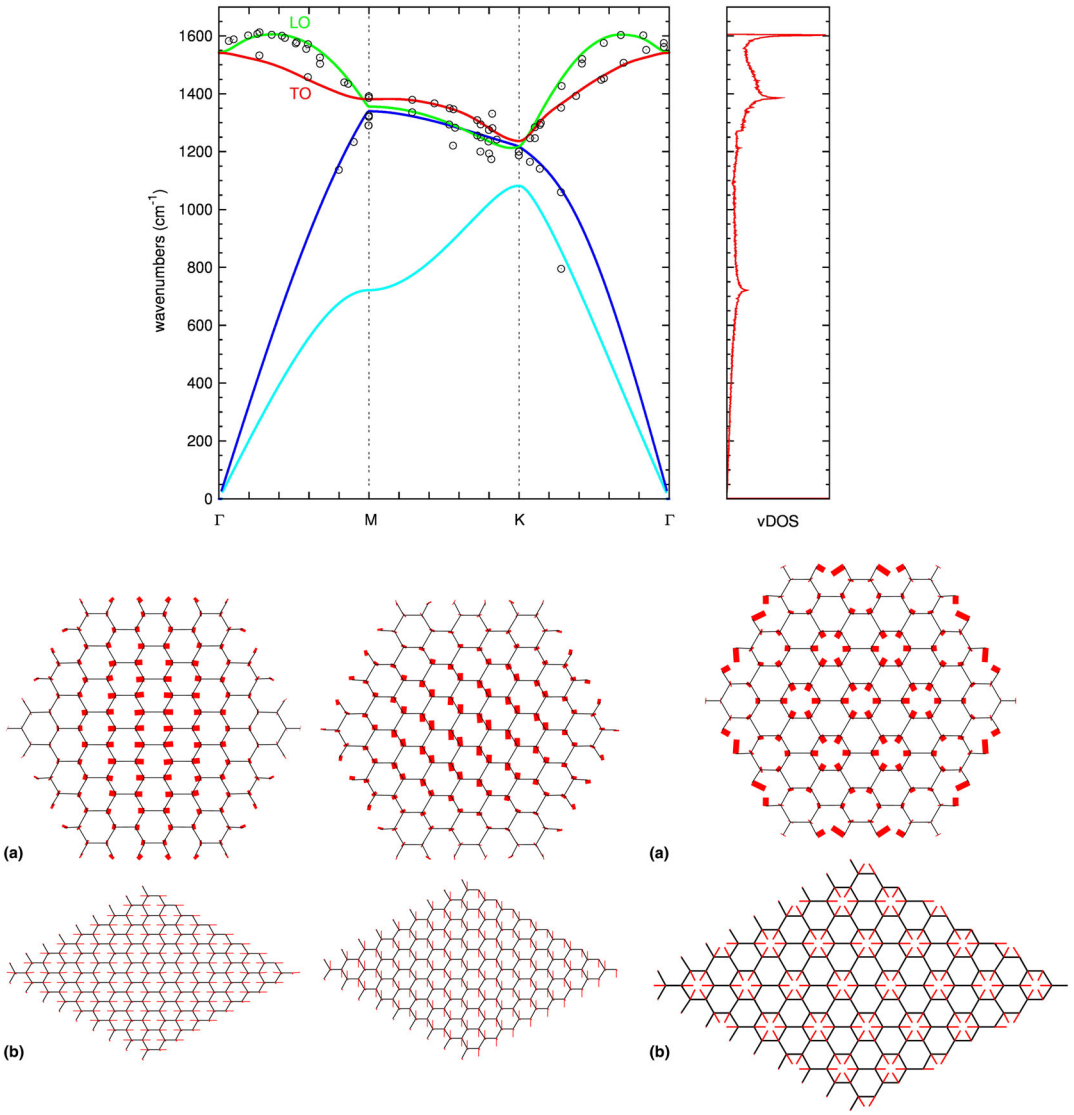


Figure 1. (top) Phonon dispersion curves and associated vibrational density of states (vDOS) computed for graphene according to Mapelli *et al.*[25] Dots are experimental data taken from Maultzsch *et al.*[34] (bottom) Nuclear displacements associated to the D and G Raman peaks are displayed for graphene (after Mapelli *et al.*[25]) and compared with related displacements in polycyclic aromatic hydrocarbons (after Di Donato *et al.*[27]).

because they give origin only to two extremely weak adsorption bands near 1580 cm^{-1} (E_{1u}) and 868 cm^{-1} (A_{2u}),[30] hardly observable with ordinary instruments. Let us consider graphite as a crystalline material consisting of extremely large sheets of condensed benzene rings with a homogeneous delocalization (conjugation) of π electrons, which causes all CC bonds to be almost identical with the same bond order and bond length. The vibrational assignments are obtained from lattice dynamical calculations on such 'infinite' two-dimensional lattice as reported in Figure 1. Theory predicts the existence of a doubly degenerate vibrational mode (phonon) at 1580 cm^{-1} Raman active (G band), which can be described as the collective (longitudinal or transversal) in-phase CC stretching mode. The collective ring-breathing mode is calculated near 1300 cm^{-1}, but it is predicted to be Raman inactive for a perfect two-dimensional lattice.[25,30] This is precisely what is observed in the Raman spectrum of samples of graphite with long-range honeycomb lattice structure. As soon as the sample becomes damaged, in other words as soon as the ideal sheet is broken into smaller flakes, we turn to more common real materials where the graphite flakes, albeit very large on a molecular scale, are finite in dimension and with irregular boundaries (for Scanning Tunneling Microscope pictures, see, for instance, Tommasini *et al.*[35]). They may contain all sorts of electronic and structural defects affecting the π system such as holes,[33] sp^3 or sp carbon atoms, free radicals, dangling bonds, distortions from planarity, grafted chemical functional groups, C–H bonds, and so on as also speculated by Radovic and Bockrath.[36] Because all samples of carbonaceous materials show a Raman line (or a group of overlapping lines) near 1300 cm^{-1} with the help of theoretical calculations on smaller models of graphene (PAH molecules), it has been shown that the activated D peak near 1300 cm^{-1} originates from the in-phase breathing vibration of the benzene rings of the core of the flake coupled with the modes on the edges of the flake (Figure 1; Negri *et al.*[37]). The existence of edges implies that we move from a fully aromatic infinitely large PAH to flakes where π electrons are necessarily confined, thus leading to a relaxed electronic structure, which tends to bond alternation.[28] In chemical terms, the contribution by localized Kekulé structures becomes more relevant. In order to put on firmer basis the role of the edges in the generation of the D line, we have recorded with three excitation lines (785, 514, and 458 nm) the Raman spectra of samples of graphite powder (Aldrich) subjected to grinding by ball milling for increasing time (up to 1200 min). As shown in Figure 2, the main G line near 1589 cm^{-1} does not significantly shift with increasing grinding time, while the intensity of the D line centered near 1330 cm^{-1} peaks at the same wavenumber but grows in intensity relative to the intensity of the G line. The intensity of the D line increases (relative to G) with increasing of the grinding time, i.e., along with the fact that smaller and smaller flakes are produced. This process has been followed by measuring the ratio $R(\lambda) = (I_D/I_G)(\lambda)$ as function of the grinding time and has been discussed in detail by Tommasini *et al.*[23] in conjunction with Electron Paramagnetic Resonance (EPR) measurements, demonstrating an interesting correlation between $R(\lambda)$ and EPR T_1 spin lattice relaxation time of conduction electrons. Tommasini *et al.*[23] also showed the relevance of the use of several excitation lines in a Raman experiment for the characterization of the structural changes occurring in graphite after a progressive ball milling, as reported in Figure 2. These Raman data show how different excitation lines give different increasing trends of the $R = I_D/I_G$ ratios as the ball-milling time increases.[23] This was explained through the concept of 'ramanophores' formed by ball milling in the nanostructured material and associated to edges and defects. Different ramanophores are resonantly excited by the different lasers used in the Raman experiment, and they are responsible for both the wavenumber dispersion of the D line and its relative intensity with respect to the G line, which is mainly due to bulk (i.e., structurally ordered) graphitic material. Other independent Raman experiments[38–40] have shown that the D line is related to the existence of a disordered matter topologically located at the edges of the flakes (for instance, Figure 3). Referring to the schemes reported in Figure 4, from the molecular viewpoint, one can think that the disorder at the very edge of graphitic regions may be described as polycondensed rings along random topologies and with many possible chemical substitutions (Figure 4). An extended graphene flake (Figure 4) can be described as consisting

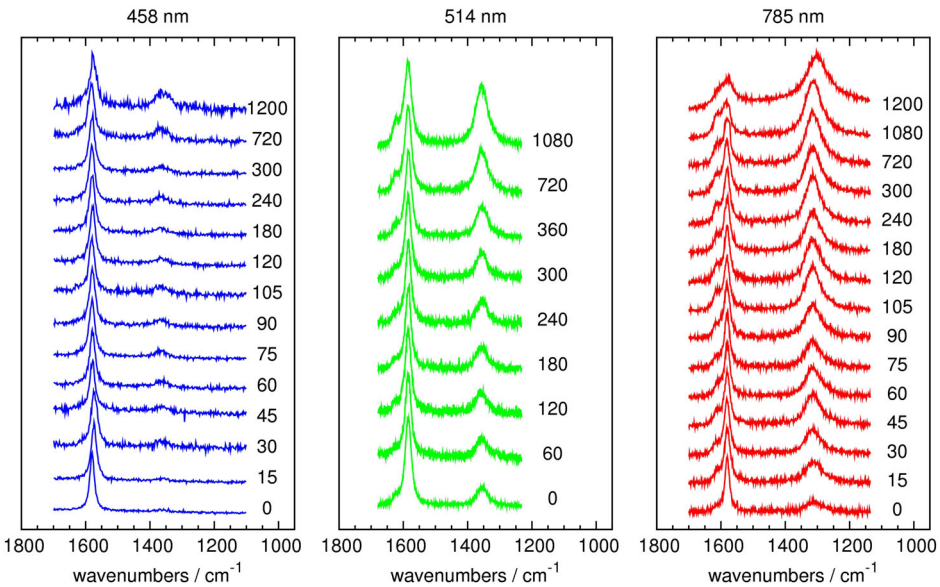


Figure 2. Raman spectra of ball-milled graphite powders recorded with different excitations. The grinding times (minutes) are reported for each spectrum.

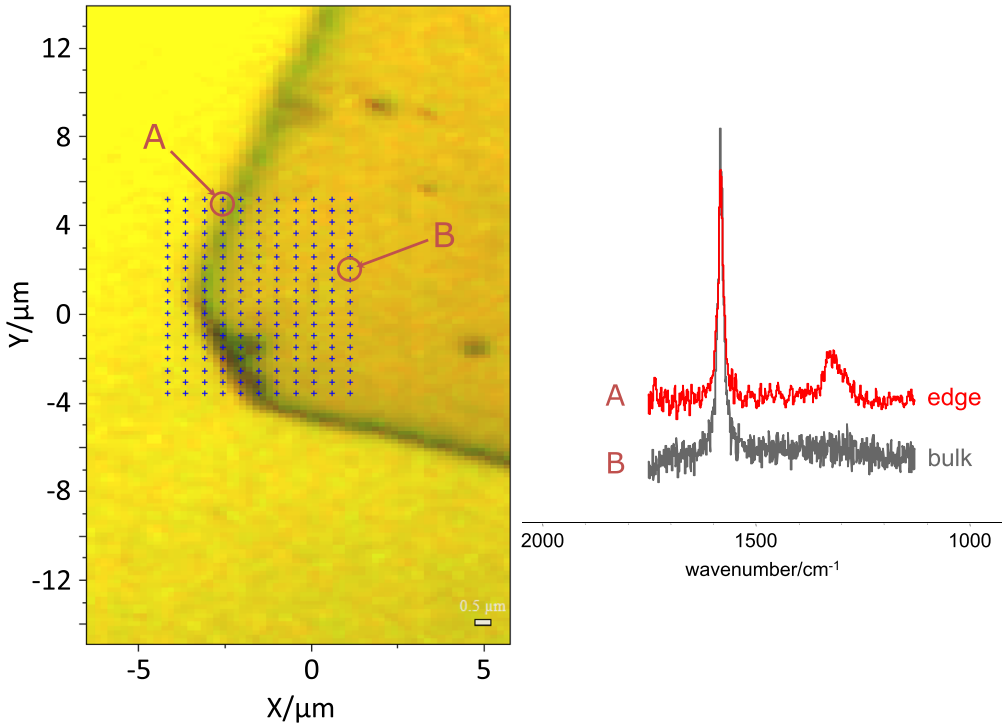


Figure 3. Micro-Raman imaging of highly oriented pyrolitic graphite deposited on Si substrate by direct contact easily allows identifying that the G line originates from the core of the graphitic 'platelets' while D originates from the edges. This is a well-known fact, independently proved by many authors with micro-Raman imaging on monolayered or bilayered graphene sheets.[38–40]

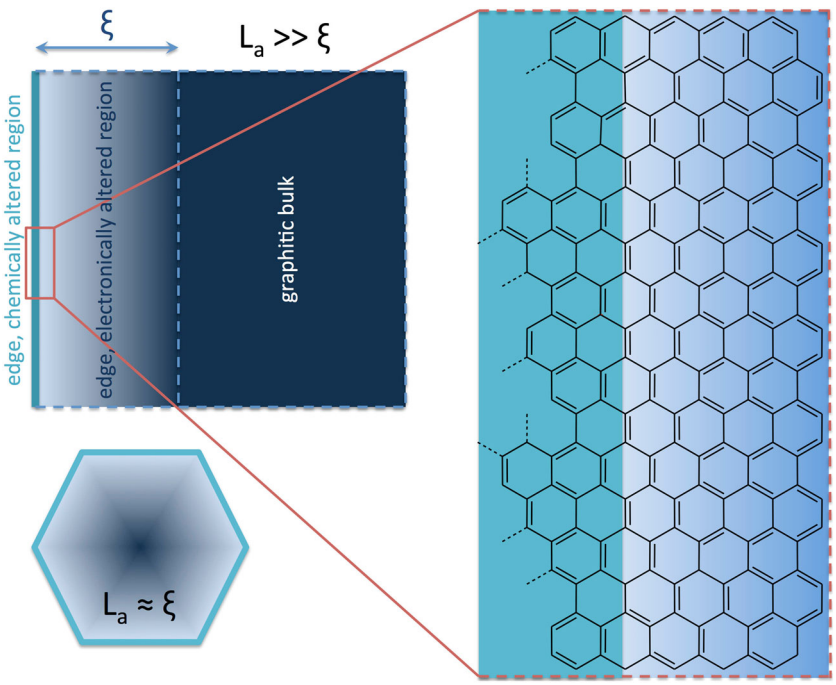


Figure 4. Cartoon depicting the origin of the Raman D peak scattering from an electronically altered region (confined; Tommasini *et al.*[23] and Castiglioni *et al.*[28]). Two regimes are reported: (top) graphitic flakes for which the electronically altered region is small compared with the crystallite size and (bottom) the molecular regime for which the whole graphitic particle is electronically altered compared with bulk graphite. On the right is a sketch of the chemically altered region (ultimate edge) inspired by the work of Radovic and Bockrath.[36] Broken bonds are indicating a variety of possible chemical groups.

of a bulk (core of size L_a), located sufficiently far from the edge, whose electronic and vibrational structures are very similar to those of an infinitely graphene sheet and Raman signature is the G line. This electronically and vibrationally unperturbed region obeys the selection rules of a graphene crystal for which the D line is forbidden by symmetry.[41] However, close to the edge, within a region of thickness $\xi << L_a$, the electronic and vibrational structures of the honeycomb lattice are perturbed by the breakdown of the translational symmetry. This gives rise to a relaxation of the lattice associated to a symmetry breaking, which activates the Raman D mode.[41,42] At the very edge of the graphene flake, many kinds of chemical structures and defects could be envisaged.[36] The concept of the inhomogeneous broadening of the D peak due to contributions from a distribution of ramanophores will be the molecular basis for the interpretation of the statistical analysis of the multiwavelength Raman spectra recorded in the present work. The $R(\lambda)$ ratio thus becomes an indication of the amount of ramanophores excited at wavelength λ. It is clear that by considering how a given sample responds under different excitations (i.e., by measuring $R(\lambda_1), R(\lambda_2), R(\lambda_3)$, etc.), one can probe spectroscopically the distribution of ramanophores in the material, which could provide a fingerprint of its structure at the molecular level. Starting from the pioneering work by Tuinstra and Koenig,[6] the Raman intensity ratio of the D and G signals (I_D/I_G) has been shown to be related with the inverse of graphite crystallite size, $1/L_a$. Thus, the larger the ratio I_D/I_G, the smaller is the average size of the flake, which originates the D band. Later, the Tunistra–Koenig relationship has been reconsidered.[43–49] Nevertheless, unless a precise determination of L_a is desired, it is clear that the I_D/I_G parameter *per se* can be considered a reliable structural probe for the mere characterization of graphitic carbonaceous materials. In other words, graphitic materials with different structures are expected to be distinguishable based on I_D/I_G ratios, especially if the data from several excitation lines are examined simultaneously. This is precisely the approach we follow for the Raman analysis of the carbonaceous particles of environmental origin considered in this work.

Experimental

Raman spectra were recorded with the following spectrometers: (1) a double monocromator Dilor XY equipped with a microscope Olympus B142 with 10×, 50×, and 100× magnification and (2) Labram HR800 with microscope Olympus BX41 with 10×, 50×, and 100× magnification. Both spectrometers employ cooled Charge-Coupled Device (CCD) light detectors. Several laser lines have been used: 458 and 514 nm from an Ar^+ laser and 785 nm from a diode laser. The laser power at the samples ranged from 10 to 20 mW. The recording of the Raman spectra of some of the samples collected as dust or cohesive deposits from surfaces of the monuments considered sometimes was made difficult by strong fluorescence. A mild washing of the sample with dilute solution of nitric acid allowed overcoming this difficulty, strongly reducing the fluorescence background. The experimental parameters recorded from the Raman spectra of each sample were as follows: wavenumber and intensity of the G and D lines as function of the three exciting line (λ_{exc} = 785, 514, 457 nm). For each spectrum, the ratio $R(\lambda) = (I_D/I_G)(\lambda)$ has been measured, where the intensities were measured as band areas with suitable and reasonable choices of the base lines. In the analysis of the many spectra recorded, the reality of the samples collected needed consideration. Indeed, all samples are made up by solid particles forming a heterogeneous mixture of micro-objects of different sizes cohesively or incohesively organized in disordered superstructures. The conclusions that can be derived from the analysis of the experimental data can be representative only if the data are averaged over a reasonably large number of spectra for each sample. Hence, for each sample, several spectra were recorded, each time moving to another site of the same sample. This issue has been taken, as much as possible, in consideration in our work. We have recorded the Raman spectra with λ_{exc} = 785, 514, or 457 nm of approximately 50 samples from different sources (Table 1) that were purposely chosen in the attempt of correlating the measured properties with the environmental conditions the particles were exposed to. Based on their origin, the samples examined in this work can be grouped as follows (the labels given to the samples will help the discussion of the results):

(a) Pure microcrystalline graphite obtained from Aldrich (pristine and subjected to successive grinding with ball milling for different lengths of time reaching 1080 min; Figure 2);
(b) Active carbon (for pharmaceutical use), glassy carbon (purchased from Aldrich), and Printex Degussa (toner for laser printer);
(c) Soot collected from aero-disperse particles collected on fiberglass filters (S1-9);
(d) Particles of PM 1, PM 2.5 and PM 10 collected on teflon/fiberglass filters (PM1, PM2.5, PM10). All these samples were kindly provided by Regional Agency for the Environmental Protection;
(e) Six samples of dust (CAD) collected in the *Cattedrale di Aosta* in 2005 by the technical staff of the *Sovraintendenza per i Beni Ambientali ed Architettonici della Regione Valle d'Aosta*;
(f) Five samples of crusts (TAC) collected from the *Teatro Romano* of the city of Aosta in 2006 by the technical staff of the *Soprintendenza per i Beni Ambientali ed Architettonici della Regione Valle d'Aosta*;
(g) Fifteen samples (AAC, AAD) of crusts and dusts collected directly by one of the authors (A. F.) of this paper (on the 13th of April 2007), all in one day with no winds, on the north and south sides of the *Arco di Augusto*, Aosta (northwest of Italy), from the vertical and horizontal surfaces removing first the dust from the surface with a brush and then using the chisel for obtaining microsamples of the crust from the same surface that was previously brushed;
(h) Seven samples (AVD, AVC) of dusts and crusts collected from the *Arena di Verona* (north center of Italy) directly by one of the authors (A. F.) of this paper (on the 5th of May 2007) using the same techniques as the one mentioned earlier. The samples were taken both from an area of the monument exposed to the urban traffic and one from the interior of a protected arcade;
(i) Five samples (LVD, LVC) of dusts and crusts collected from the *Libreria Marciana* of Venice (northeast of Italy) directly by one of the authors (A. F.) of this paper (on the 10th of October 2007) using the same techniques as the one mentioned earlier; and
(j) Micro-Raman spectra of highly oriented pyrolitic graphite mechanically transferred on silicon substrate by simple contact (Figure 3).

For a better understanding of the Raman spectra for all the samples, we have recorded the Raman spectra with three excitation lines, namely λ_{exc} = 785, 514, 457 nm). Particular attention has been paid for the measuring of I_D/I_G. Further details and data analysis can be found in the work by Ferrugiari.[50]

Table 1. I_D/I_G ratios of the set of samples subjected to multiwavelength Raman analysis (discussed in the text)

Label	Description	785 nm	514 nm	457 nm
AC	Active carbon	3.78	1.52	1.10
GC	Glassy carbon	2.63	0.74	0.80
PD	Printex Degussa	3.31	1.27	0.88
S1	Diesel soot	3.96	1.09	0.52
S2	Ethylene soot	3.44	1.22	1.20
S3	Ethylene soot	3.97	1.90	1.14
S4	Benzene soot	3.39	0.81	1.24
S5	Soot from hydrocabron mixture (C1–C4)	2.74	1.02	0.77
S6	Acetone soot	2.09	1.11	0.77
S7	Decane soot	2.25	1.01	0.92
S8	Toluene soot	2.20	1.00	1.02
S9	Xylene soot	2.11	0.91	0.80
PM1	Particulate matter < 1 μm	4.40	0.95	0.48
PM2.5	Particulate matter < 2.5 μm	3.61	0.98	0.56
PM10	Particulate matter < 10 μm	3.70	0.88	0.51
TAC6	Crust from Aosta theater	2.75	1.43	1.00
TAC13	Crust from Aosta theater	3.35	1.37	0.96
TAC18	Crust from Aosta theater	3.63	1.03	1.40
TAC22	Crust from Aosta theater	2.87	0.94	0.75
TAC35	Crust from Aosta theater	2.00	0.64	0.71
CAD4	Dust from Aosta cathedral	2.63	0.65	0.41
CAD5	Dust from Aosta cathedral	2.67	0.69	0.41
CAD6	Dust from Aosta cathedral	2.43	0.66	0.30
CAD7	Dust from Aosta cathedral	2.89	0.82	0.29
CAD8	Dust from Aosta cathedral	2.69	0.74	0.58
CAD9	Dust from Aosta cathedral	2.81	0.67	0.26
AAC4	Crust from Aosta arc	2.32	1.12	0.65
AAC6	Crust from Aosta arc	2.70	0.68	0.77
AAC10	Crust from Aosta arc	3.45	0.52	0.76
AAC12	Crust from Aosta arc	3.17	0.65	0.91
AAC16	Crust from Aosta arc	3.16	0.66	0.46
AAC18	Crust from Aosta arc	3.80	0.60	1.37
AAD1	Dust from Aosta arc	2.97	0.44	0.44
AAD3	Dust from Aosta arc	3.19	1.00	1.15
AAD9	Dust from Aosta arc	3.32	0.72	1.21
AAD11	Dust from Aosta arc	3.20	0.74	0.87
AAD13	Dust from Aosta arc	3.54	0.79	0.44
AAD15	Dust from Aosta arc	3.27	0.54	1.18
AAD17	Dust from Aosta arc	3.22	1.28	0.74
AAD19	Dust from Aosta arc	3.35	0.71	0.28
AAD23	Dust from Aosta arc	3.67	0.80	0.76
AVD1	Dust from Verona arena	3.17	0.66	0.28
AVD3	Dust from Verona arena	2.45	0.52	0.19
AVD5	Dust from Verona arena	3.03	0.95	0.35
AVD7	Dust from Verona arena	2.58	0.63	0.59
AVD11	Dust from Verona arena	2.44	0.77	0.29
AVD15	Dust from Verona arena	2.54	0.79	0.39
AVC8	Crust from Verona arena	3.61	1.08	0.46
LVD1	Dust from Venice 'Marciana' library	3.31	0.56	0.49
LVD3	Dust from Venice 'Marciana' library	2.86	0.38	0.47
LVD5	Dust from Venice 'Marciana' library	3.09	0.92	0.97
LVD7	Dust from Venice 'Marciana' library	3.71	0.85	0.56
LVC4	Crust from Venice 'Marciana' library	3.63	0.54	0.62

Results and discussion

D peak wavenumber dispersion

Definitely all samples examined in the scattering range of the D line show wavenumber and intensity dispersion upon excitation with λ_{exc} = 745, 514, and 459 nm, following a well-known general trend of graphitic materials early reported by Pócsik[51] for microcrystalline graphite. Furthermore, as discussed earlier, enough evidence has been collected that the D line originates mostly from the normal vibrations of the edges of graphitic structures. Owing to the strongly and intrinsic heterogeneous nature of the samples examined, the scattering lines are broad, and necessarily, we

have to consider them as a convolution of the scattering of a set of 'ramanophores', which form the chemical and structural disorder mostly located at the edges of the graphitic structures generated during the formation process of the material. With this term, we wish to indicate the existence at the edges of one or more types of polycondensed aromatic molecular islands whose vibrational modes are vibronically coupled with the (delocalized) $\pi - \pi^*$ transition excited by the chosen laser line at λ_{exc}. In Figure 5, we report the position of the D line for the series of samples considered in this work. For samples of the same kind (e.g., AVDN, $N = 1, 3, 5, 7, 11, 15$), we report the average wavenumber over the available data points.

Because of the complexity of the samples, a detailed vibrational assignment cannot be proposed; however, from an overall comparison of the observed spectra, the following observations and suggestions can be offered:

(i) The resonance or near-resonance transition excited by $\lambda_{exc} = 745$ nm gives rise preferentially to scattering in the narrow wavenumber range 1310–1320 cm^{-1} in all the samples examined in this work. This signal has been investigated in detail in ball-milled graphite samples (Figure 2), which reveals the existence of a group of ramanophores whose relative concentration increases when the size of the 'flake' decreases, eventually reaching the onset of saturation (Tommasini *et al.*[23]).

(ii) Excitation with $\lambda_{exc} = 514$ nm brings up a broad scattering for all samples in the wavenumber range 1345–1360 cm^{-1} as if resonance or near resonance was reached across an energy gap larger than that reached with excitation in the red ($\lambda_{exc} = 745$ nm).

(iii) Excitation with higher energies with a blue line ($\lambda_{exc} = 459$ nm) shows a scattering from 1350 to 1380 cm^{-1}, which extends over a range significantly wider than those in (i) and (ii). We observe that for the series of samples collected personally and with particular care by one of the co-authors (A. F.) at Venice, Verona, and Aosta, the samples as powders (D) scatter at wavenumbers higher than those of the same materials but in cohesive form (C). Moreover, the wavenumber difference of the scattering with $\lambda_{exc} = 514$ and 459 nm within each sample is larger than what is experimentally observed for all the remaining kinds of samples. We may infer that, under the influence of the environment, the chemical nature of particles at a given site may change over time. Tentatively, this may be associated with photochemical reactions, thermal relaxations, or chemical reactions caused by residual free radicals. Consequently, the electronic and vibrational properties of these sp^2 carbon materials are necessarily affected.

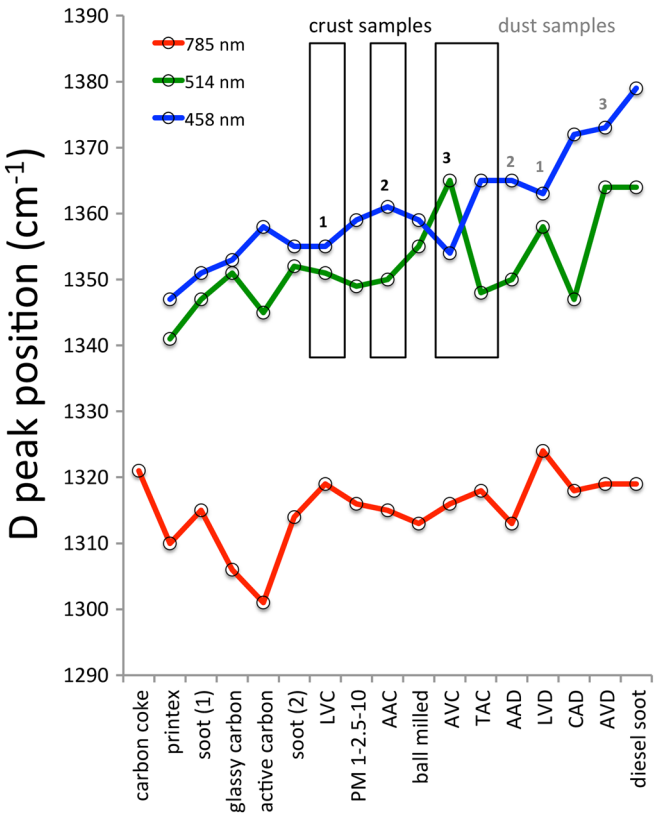


Figure 5. Dispersion of the D line (cm^{-1}) for the samples considered in this work (discussed in the text). Based on the position of the D line, for a given sample origin (Venice, Verona, and Aosta), dust samples, compared with crust samples, show different ramanophores in green and blue, not in red (i.e., (1) LVD > LVC; (2) AAD > AAC; and (3) AVD > AVC).

Statistical analysis

Two different statistical analyses have been considered, namely Ward's[52] hierarchical cluster analysis (HCA) and principal component analysis (PCA) as implemented in R statistical package.[53] The results are reported in Figures 6 and 7. The results from HCA are graphically represented by the dendrogram reported in Figure 6. HCA attempts to group multivariate data – i.e., the set of I_D/I_G Raman intensity ratios at the three excitation wavelengths for the various samples considered – by minimizing the sum of distances among the samples within the cluster (we consider the Euclidean distance between n-dimensional vectors, with $n = 3$). Therefore, the samples characterized by close enough values of I_D/I_G ratios at the three excitations fall within the same cluster.

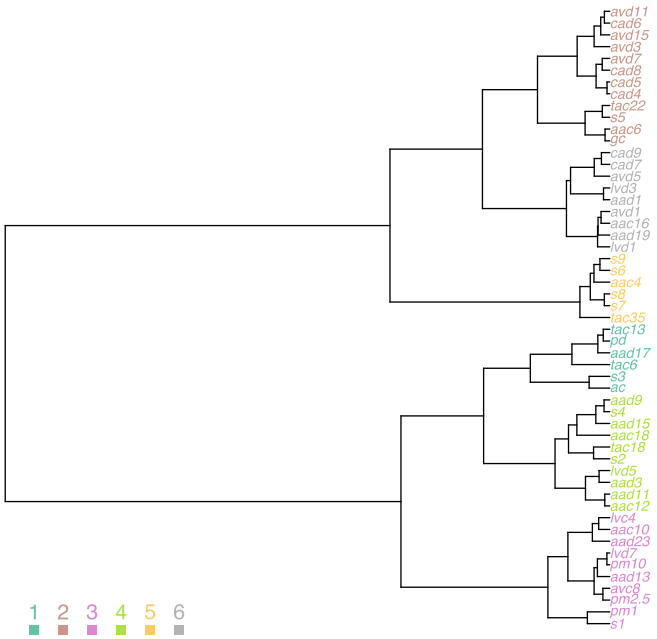
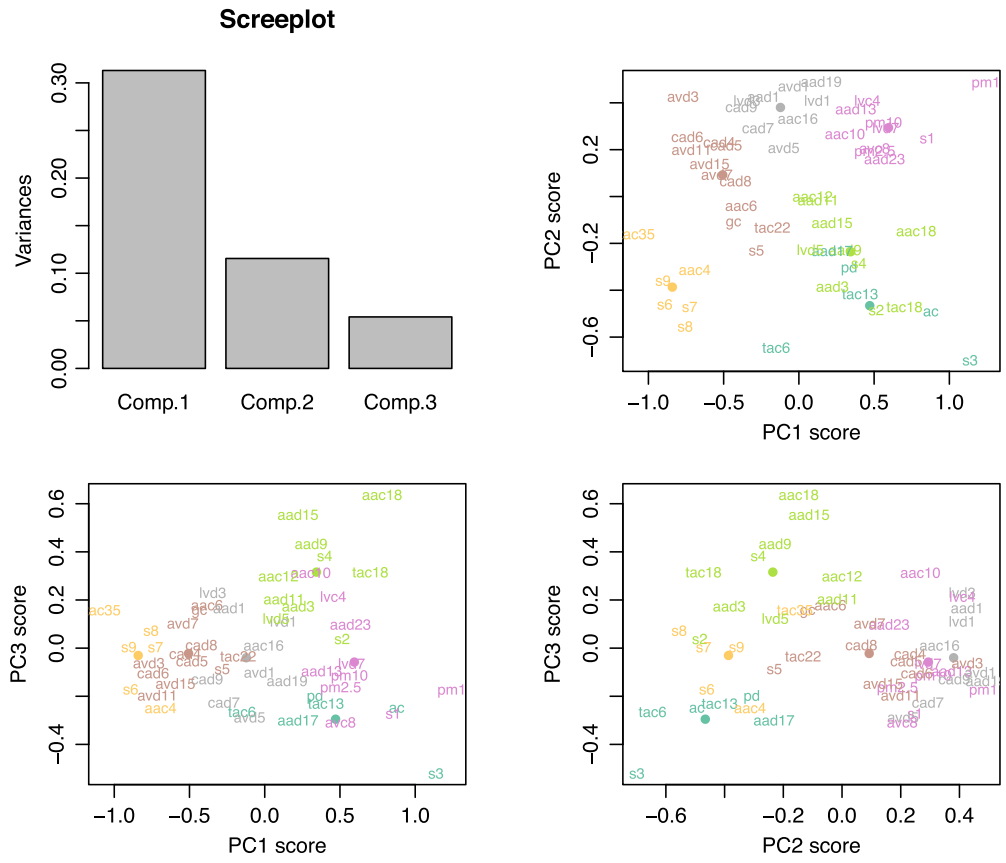


Figure 6. Dendrogram relative to the hierarchical cluster analysis of the dataset reported in Table 1.



cluster #	R785	R514	R458	PC1	PC2	PC3
1	3.40	1.46	0.97	0.47	-0.46	-0.30
2	2.61	0.73	0.52	-0.51	0.09	-0.02
3	3.73	0.85	0.57	0.60	0.29	-0.06
4	3.35	0.82	1.15	0.34	-0.24	0.31
5	2.16	0.96	0.81	-0.84	-0.39	-0.03
6	3.06	0.65	0.37	-0.12	0.38	-0.04
dataset	*3.08*	*0.87*	*0.71*	*0.00*	*0.00*	*0.00*

Figure 7. Graphical representation of the principal component analysis results. The coordinates of the average over the data points belonging to each cluster from hierarchical cluster analysis (centroids) are reported in the bottom panel in order to guide the reading of the results. The loadings along the three PCs are also given as histogram plot.

The grouping procedure in Ward's algorithm is carried out in a hierarchical fashion and produces a partitioning of the dataset described by a tree diagram (Figure 6). At the right end of the tree, one sees the extreme case where each sample lies in its own cluster. Moving to the left of the tree, the grouping among the samples is carried out at subsequent steps, progressively putting together samples that are more similar. To be of practical use, the best clustering along the tree is chosen by selecting a number k of clusters that is not too large or too small. In the present case, $k = 6$ represents a reasonable compromise between a too severe criterion for judging diversity among samples and an excessively loose one. It is satisfactory to observe that samples with similar origin can be effectively grouped within the same cluster (see Figure 6 and information on the samples given in Table 1). For instance, this is the case of the soot samples produced by combustion of different substances in our laboratory (cluster #5). More loosely, samples from Aosta and Verona tend to be grouped in a few clusters.

As a further step, we have coupled HCA with PCA. This allows quantitatively characterizing how samples are interrelated. The results from PCA are summarized in Figure 7 adopting the same color scheme obtained from HCA. PCA is carried out through the covariance matrix of the dataset. The principal components are related to the eigenvectors of the covariance matrix and describe independent directions in the space of the variables along which the greatest changes among the samples are observed. In PCA, these changes are computed with respect to the average over the full dataset. The loadings (PCA eigenvectors) describe the relative importance of each variable in describing the changes along the dataset. The difference between the centroid (i.e., average) of each cluster and the average of the dataset can be projected on the PCA eigenvectors to give the representative positions in

the principal component space of the six clusters of data points. This information is tabulated in the bottom of Figure 7 beside the graphical representation of the three principal component eigenvectors. PC1 mainly represents a change along the R785 variable with a minor R514 component. PC2 represents a change along both R514 and R458. PC3 represents a change where R514 and R458 vary in opposite fashion. In more 'molecular' terms, taking as a reference the average over the full dataset (with all zero coordinates in principal component space), a displacement along PC1 describes the increase of ramanophores resonating mainly at 785 nm, while a displacement along PC2 describes the increase of ramanophores resonating at 514 and 458 nm. Finally, PC3 describes a relative increase of ramanophores resonating at 514 nm associated with a decrease of those resonating at 458 nm. Keeping in mind this interpretation of PCs, one can better distinguish the different characteristics of the clusters. For instance, cluster #5 has less ramanophores resonating at 785 nm (and also less ramanophores resonating at 514 and 458 nm) than the average of the dataset (both PC1 and PC2 scores are negative). On the contrary, cluster #3 shows the opposite behavior, with a larger content of ramanophores resonating at 785, 514, and 458 nm (both PC1 and PC2 scores are positive).

In conclusion, we have shown that multiwavelength Raman spectroscopy of carbonaceous particulate can unveil differences among samples, which are ultimately traced back to their different structure at the molecular level. Even though a precise correlation with the underlying chemical structures still represents a daunting problem for the vibrational spectroscopy of such disordered and heterogeneous samples, we believe that by means of the methods of multivariate statistics (e.g., HCA and PCA), it is possible to provide a rationale for the classification of otherwise scattered experimental observations, which are very difficult (if not impossible) to scrutinize by human inspection. We have found that the I_D/I_G descriptor, widely adopted in the Raman spectroscopy of graphitic systems, when considered as a function of the excitation wavelength (λ), becomes a sensitive probe of the different distribution of resonating (molecular-like) species in the samples. Hence, multivariate analysis of $I_D/I_G(\lambda)$ ratios can disclose differences among samples, which may go beyond the morphological level and go closer to a chemically relevant description of this family of samples, which are difficult to analyze by the common methods of analytical chemistry. We hope that the proposed Raman characterization and data analysis approach may foster the studies on environmental carbon, such as airborne soot, an active area of research[9–11,13,14] where Raman spectroscopy is expected to play a role. Interestingly, an automated Raman analysis of naturally occurring carbon in powdered geological and environmental samples has been recently reported.[12] Automated approaches are well-suited to be complemented with further data analysis schemes, such as the one described in the present work. Furthermore, very recently, the multiwavelength Raman spectroscopy of the G and D lines of soots originated from the combustion of different fuels in controlled conditions has shown the possibility of unveiling differences among samples at molecular level.[54] This strongly motivates the use of Raman in the fields of combustion and environment.

Acknowledgements

We acknowledge Dr. L. Brambilla and Dr. G. Fustella (Politecnico di Milano) for their help in the experimental aspects of the work and the related discussion.

References

[1] B. van Setten, M. Makkee, J. Moulijn, *Catal. Rev. Sci. Eng.* **2001**, *43*, 489–564.
[2] R. Escribano, J. Sloan, N. Siddique, N. Sze, T. Dudev, *Vib. Spectrosc.* **2001**, *26*, 179–186.
[3] R. N. Jones, C. Sandorfy, in *Chemical Applications of Spectroscopy* (Ed: A. Weissenberger), Techniques of Organic Chemistry, vol. 9, Wiley, New York, **1956**.
[4] L. Bellamy, *The Infrared Spectra of Complex Molecules*, Springer, Dordrecht, **1975**.
[5] G. Socrates, *Infrared and Raman Characteristic Group Frequencies* (3rd edn), Wiley, Chichester (West Sussex, England), **2004**.
[6] F. Tuinstra, J. L. Koenig, *The J. Chem. Phys.* **1970**, *53*.
[7] A. Cuesta, P. Dhamelincourt, J. Laureyns, A. Martinez-Alonso, J. M. D. Tascon, *Carbon* **1994**, *32*, 1523–1532.
[8] A. Sadezky, H. Muckenhuber, H. Grothe, R. Niessner, U. Poschl, *Carbon* **2005**, *43*, 1731–1742.
[9] S. Sze, N. Siddique, J. Sloan, R. Escribano, *Atmos. Environ.* **2001**, *35*, 561–568.
[10] Y. Batonneau, S. Sobanska, J. Laureyns, C. Bremard, *Environ. Sci. Technol.* **2006**, *40*, 1300–1306.
[11] A. Worobiec, S. Potgieter-Vermaak, A. Brooker, L. Darchuk, E. Stefaniak, R. V. Grieken, *Microchem. J.* **2010**, *94*, 65–72.
[12] R. Sparkes, N. Hovius, A. Galy, R. V. Kumar, J. T. Liu, *Appl. Spectrosc.* **2013**, *67*, 779–788.
[13] H.-J. Jung, H.-J. Eom, H.-W. Kang, M. Moreau, S. Sobanska, C.-U. Ro, *Analyst* **2014**, *139*, 3949–3960.
[14] T. Catelani, G. Pratesi, M. Zoppi, *Aerosol Sci. Technol.* **2014**, *48*, 13–21.
[15] I. M. Bell, R. J. Clark, P. J. Gibbs, *Spectrochim. Acta, Part A* **1997**, *53*, 2159–2179.
[16] O. Beyssac, B. Goffe, C. Chopin, J. N. Rouzaud, *J. Metamorph. Geol.* **2002**, *20*, 859–871.
[17] K. Niwase, T. Tanaka, Y. Kakimoto, K. N. Ishihara, P. H. Shingu, *Mater. Trans., JIM* **1995**, *36*, 282–288.
[18] J. Y. Huang, *Acta Mater.* **1999**, *47*, 1801–1808.
[19] J. Y. Huang, H. Yasuda, H. Mori, *Chem. Phys. Lett.* **1999**, *303*, 130–134.
[20] N. J. Welham, V. Berbenni, P. G. Chapman, *J. Alloy. Compd.* **2003**, *349*, 255–263.
[21] T. Makarova, M. Riccò, D. Pontiroli, M. Mazzani, M. Belli, A. Goffredi, *Phys. Status Solidi b* **2008**, *245*, 2082–2085.
[22] W. Zhao, M. Fang, F. Wu, H. Wu, L. Wang, G. Chen, *J. Mater. Chem.* **2010**, *20*, 5817–5819.
[23] M. Tommasini, C. Castiglioni, G. Zerbi, A. Barbon, M. Brustolon, *Chem. Phys. Lett.* **2011**, *516*, 220–224.
[24] T. Xing, L. H. Li, L. Hou, X. Hu, S. Zhou, R. Peter, M. Petravic, Y. Chen, *Carbon* **2013**, *57*, 515–519.
[25] C. Mapelli, C. Castiglioni, G. Zerbi, K. Mullen, *Phys. Rev. B* **1999**, *60*, 12710–12725.
[26] F. Negri, C. Castiglioni, M. Tommasini, G. Zerbi, *J. Phys. Chem. A* **2002**, *106*, 3306–3317.
[27] E. Di Donato, M. Tommasini, G. Fustella, L. Brambilla, C. Castiglioni, G. Zerbi, C. D. Simpson, K. Mullen, F. Negri, *Chem. Phys.* **2004**, *301*, 81–93.
[28] C. Castiglioni, M. Tommasini, G. Zerbi, *Philos. Trans. R. Soc. London Ser. A* **2004**, *362*, 2425–2459.
[29] M. Tommasini, C. Castiglioni, G. Zerbi, *Phys. Chem. Chem. Phys.* **2009**, *11*, 10185–10194.
[30] S. Reich, C. Thomsen, *Philos. Trans. R. Soc. London. Ser. A* **2004**, *362*, 2271–2288.
[31] A. Jorio, R. Saito, G. Dresselhaus, M. S. Dresselhaus, in *Raman Spectroscopy in Graphene Related Systems*, Wiley-VCH Verlag, Weinheim, **2011**, pp. I–XIV.
[32] A. C. Ferrari, D. M. Basko, *Nat. Nano.* **2013**, *8*, 235–246.
[33] A. Centrone, L. Brambilla, T. Renouard, L. Gherghel, C. Mathis, K. Mullen, G. Zerbi, *Carbon* **2005**, *43*, 1593–1609.
[34] J. Maultzsch, S. Reich, C. Thomsen, H. Requardt, P. Ordejon, *Phys. Rev. Lett.* **2004**, *92*, 075501.
[35] M. Tommasini, E. Di Donato, C. Castiglioni, G. Zerbi, N. Severin, T. Böhme, J. P. Rabe, *AIP Conf. Proc.* **2004**, *723*, 334–338.
[36] L. Radovic, B. Bockrath, *J. Am. Chem. Soc.* **2005**, *127*, 5917–5927.
[37] F. Negri, E. di Donato, M. Tommasini, C. Castiglioni, G. Zerbi, K. Mullen, *J. Chem. Phys.* **2004**, *120*, 11889–11900.

[38] D. Graf, F. Molitor, K. Ensslin, C. Stampfer, A. Jungen, C. Hierold, L. Wirtz, *Nano Lett.* **2007**, *7*, 238–242.
[39] A. K. Gupta, T. J. Russin, H. R. Gutierrez, P. C. Eklund, *ACS Nano* **2009**, *3*, 45–52.
[40] C. Casiraghi, A. Hartschuh, H. Qian, S. Piscanec, C. Georgi, A. Fasoli, K. S. Novoselov, D. M. Basko, A. C. Ferrari, *Nano Lett.* **2009**, *9*, 1433–1441.
[41] C. Castiglioni, F. Negri, M. Rigolio, G. Zerbi, *J. Chem. Phys.* **2001**, *115*, 3769–3778.
[42] M. Tommasini, E. Di Donato, C. Castiglioni, G. Zerbi, *Chem. Phys. Lett.* **2005**, *414*, 166–173.
[43] T. P. Mernagh, R. P. Cooney, R. A. Johnson, *Carbon* **1984**, *22*, 39–42.
[44] M. Matthews, M. Pimenta, G. Dresselhaus, M. Dresselhaus, M. Endo, *Phys. Rev. B* **1999**, *59*, R6585–R6588.
[45] G. A. Zickler, B. Smarsly, N. Gierlinger, H. Peterlik, O. Paris, *Carbon* **2006**, *44*, 3239–3246.
[46] M. M. Lucchese, F. Stavale, E. H. M. Ferreira, C. Vilani, M. V. O. Moutinho, R. B. Capaz, C. A. Achete, A. Jorio, *Carbon* **2010**, *48*, 1592–1597.
[47] E. H. Martins Ferreira, M. V. O. Moutinho, F. Stavale, M. M. Lucchese, R. B. Capaz, C. A. Achete, A. Jorio, *Phys. Rev. B* **2010**, *82*, 125429.
[48] L. G. Cancado, A. Jorio, E. H. M. Ferreira, F. Stavale, C. A. Achete, R. B. Capaz, M. V. . O. Moutinho, A. Lombardo, T. S. Kulmala, A. C. Ferrari, *Nano Lett.* **2011**, *11*, 3190–3196.
[49] P. Mallet-Ladeira, P. Puech, C. Toulouse, M. Cazayous, N. Ratel-Ramond, P. Weisbecker, G. L. Vignoles, M. Monthioux, *Carbon* **2014**, *80*, 629–639.
[50] A. Ferrugiari, *Studio degli aspetti molecolari della frazione carboniosa del particolato atmosferico: la ricaduta sui beni culturali*. Ph.D. Thesis, Politecnico di Milano, Milano, **2008**.
[51] I. Pócsik, M. Hundhausen, M. Koós, L. Ley, *J. Non-Cryst. Solids* **1998**, *227 – 230, Part 2*, 1083–1086.
[52] J. H. Ward, *J. Am. Stat. Assoc.* **1963**, *58*, 236–244.
[53] R Core Team, R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, **2013**.
[54] C. Russo, A. Ciajolo, *Combust. Flame* **2015**, *162*, 2431–2441.