Research article

Journal of RAMAN SPECTROSCOPY

Received: 30 March 2017 Revised: 18 June 2017 Accepted: 17 July 2017 Published online in Wiley Online Library: 15 September 2017

(wileyonlinelibrary.com) DOI 10.1002/jrs.5225

Applications of Raman spectroscopy in graphene-related materials and the development of parameterized PCA for large-scale data analysis

João Luiz Elias Campos,[a] Hudson Miranda,[b] Cassiano Rabelo,[b] Emil Sandoz-Rosado,[c] Sugandha Pandey,[a] Juha Riikonen,[d] Abraham G. Cano-Marquez[e] and Ado Jorio[a,b]*

A methodology for the structural analysis of graphene-related materials using parameterized principal component analysis (PCA), ideal for large-scale data treatment, is introduced. First, we review different aspects of Raman spectroscopy for structural and functional characterization of sp^2-bonded carbon materials, which are important for understanding the problem. The parameterized PCA is then introduced and applied to 2 different scenarios: to identify different sp^2 carbon structures and to identify graphene samples with different numbers of layers. Automating these Raman spectroscopy analysis techniques is desired for large-scale industrial applications. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: graphene; principal component analysis; applications; large-data analysis; carbon nanomaterial

Introduction

Graphene-related structures have generated great interest from the scientific community, in both basic and applied research.[1] From a perspective, the sp^2 carbon structures offered a number of broadly studied materials, such as graphite intercalated compounds (GICs),[2] amorphous carbons,[3] fullerenes,[4] nanotubes,[5–7] graphene,[8] and biochar.[9] From another perspective, graphene opened the field of 2-dimensional systems,[10] launching the research on x-enes[11] (phosphorene,[12] arsenene,[13] antimonene,[13] silicene,[14] germanene,[15] borophene[16]), transition metal dichalcogenides (TMDs)[17,18] and monochalcogenides,[19] and van der Walls heterostructures.[20]

Raman spectroscopy has been the leading technique to study and characterize graphene-related structures,[2,4,21,22] basically because it is a simple and non-invasive technique, and because the carbon atoms are light, the sp^2 σ bonds are strong, and the π-electron-related optical transitions range from the infrared up to the visible range. These structural characteristics are responsible for many properties of the sp^2-bonded carbons that contribute to the high degree of sensitivity observed in Raman spectra of these systems, which include high phonon frequencies, sensitivity to strain and temperature; strong electron-phonon coupling and many-body effects, sensitivity to doping, functionalization, environmental conditions and dimensionality, and optical resonances, allowing spectroscopy of single nanostructures and detailed analysis of their electronic structure.[22]

While our present knowledge of the applications of Raman spectroscopy to graphene-related systems has been sufficient for laboratory proof-of-concepts, it is desirable to generate automated procedures for big data analysis to turn Raman spectroscopy into a technique for industrial control. In this sense, principal component analysis (PCA)[23] is a technique for dimensionality reduction and feature extraction that consists of calculating the directions of greatest variance in a data set. The largest variance components are, naturally, linear combinations of the attributes of the original data set, and what PCA does essentially is to perform a transformation of coordinates by translating and rotating the original coordinate frame.[23]

However, while PCA is able to extract nontrivial features in the data set and display relationships between the samples, in most cases, the interpretation of the PCA results is nontrivial, and the analysis changes from one data set to another. To overcome this problem, we discuss the use of parameterized PCA for applications

* Correspondence to: Ado Jorio, Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte, MG 30270-970, Brazil.
E-mail: adojorio@fisica.ufmg.br

a Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 30270-970, Brazil

b Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 30270-970, Brazil

c US Army Research Laboratory, Aberdeen Proving Ground, Aberdeen, MD 21005, USA

d Department of Micro- and Nanosciences, Aalto University, 02150, Espoo, Finland

e Centro de Investigación en Materiales Avanzados, Unidad Monterrey, C.P. 66628, Apodaca, NL, Mexico

in sp^2 carbon systems. Initially, we present an overview of some Raman spectroscopy characteristics, and how they have been used to study and characterize graphene-related systems, addressing both structural and functional analysis. We do not intend to provide a complete overview of the literature, but just to exemplify the possibilities offered by the technique, which are then explored in the PCA. Subsequently, we introduce the concept of the PCA parameterization to generate a meaningful automated classification methodology in the field of Raman spectroscopy applied to sp^2 carbon structures, which can be used in large-scale analysis. The paper closes with summary and perspectives.

General features of Raman spectroscopy

The goal of this section is to introduce some spectral characteristics of Raman spectroscopy in graphene-related materials and to illustrate applications of this technique. The use of Raman spectroscopy includes structural characterization of sp^2 structures and the amorphization trajectory from pristine graphene to amorphous carbon[24]; structural and functional analysis of strain-stress relations in carbon nanocones[25] and graphene, including tribology aspects[26]; and the use of tip enhanced Raman spectroscopy (TERS) to elucidate the nanocrystallite structure of chemical vapor deposition samples.[27] Some application examples besides graphene quality control include classification of different types of biochar and their structural proximity[28–30]; Raman spectroscopy visualization of carbon nanotubes used as a carrier agent in biotechnology applications[31–33]; and the use of graphene/graphite as a standard for nanometrology.[34,35]

Graphene amorphization routes

From the dimensionality standpoint, the amorphization of a 2-dimensional crystalline system can follow 2 routes: increasing the number of zero-dimensional (point-like) defects, such as vacancies or 7-5 pairs in graphene, or increasing the number of 1-dimensional (line-like) defects, such as grain boundaries in graphene. Raman spectroscopy was shown to be able to quantify the amount of point-like and line-line defects in graphene using 2 pieces of spectral information together: (1) the integrated intensity (peak area) ratio between the D (~1350 cm^{-1}) and G (~1580 cm^{-1}) Raman peaks, A_D/A_G, and (2) the G band full width at half maximum, Γ_G.[24]

Figure 1(a) shows the carbon amorphization Raman diagram, with exemplary spectra from different locations in this diagram displayed in Figure 1(b,c). The A_D/A_G spectral information is normalized in the Raman diagram by multiplying the ratio by the excitation laser energy to the fourth power, E_L^4, to make the information independent of the excitation laser energy.[24,36] The solid and dotted lines in Figure 1(a) provide the amorphization routes for pure line-like and pure point-like defects, respectively. The spectra following the amorphization route of line-like defects come from diamond-like amorphous carbons heat treated at different temperatures.[36] The spectra following the amorphization route of point-like defects come from graphene subjected to different doses of Ar^+ ion bombardment.[37] Samples exhibiting both point and line defects exhibit the $[(A_D/A_G)*E_L^4,\Gamma_G]$ spectral information inside the area limited by these lines.[24]

The amorphization metrics are L_D, the average distance between points, and L_a, the average distance between lines. This analysis can be used to monitor and guide graphene-like sample growth and applications, as it has been done for characterizing the growth of graphene by chemical vapor deposition (CVD) from natural gas, and its application in conductive inks.[24]

Grain boundaries measured with tip-enhanced Raman spectroscopy

Graphene samples usually exhibit a polycrystalline structure. Figure 2(a) shows a scanning tunneling microscopy (STM) image of a CVD grown graphene sample.[27] The bright (yellow) lines indicate grain boundaries, i.e., a graphene sample full of line-like defects. Tip enhanced Raman spectroscopy (TERS) can be used to spectrally image defects, as shown in Figure 2(b), due to the local enhancement of the D band. Exemplary spectra are shown in Figure 2(c), from locations numbered 1, 2, and 3 in (a,b).

Point 3 in Figure 2(a) is a defective structure, and the respective spectra are shown in the bottom of Figure 2(c). The defect-induced Raman signature is only observed when the TERS tip is positioned exactly on top of the defect (tip down). Without the tip (tip retracted from the sample surface), the Raman signal comes from a large graphene area, mostly crystalline, which is being illuminated in a confocal configuration. When the tip is landed on top of the defect, the defect-induced signal is locally enhanced and becomes prominent in the Raman spectrum (see tip up vs tip down in the bottom spectra of Figure 2(c)).

Through similar reasoning, it is evident that point 1 is not actually a graphene grain boundary, because the TERS tip landing does not enhance a graphene defect-induced spectral signature, but rather a broad photoluminescence background (see inset to Figure 2(c)). Notice this point was measured in a location where TERS shows a bright signal (Figure 2(b)) without a corresponding signal in the STM image (Figure 2(a)), indicating the enhanced optical signal comes from underneath photoluminescent material.

Structural strain in carbon nanocone

Small variations in the chemical bonds influence material's vibrational properties and change their Raman spectra. Consequently, mechanical properties of materials can be monitored using Raman spectroscopy. Figure 3 (scanning electron microscopy, SEM image in Figure 3(a)[25] and Raman spectral image in Figure 3(b)) addresses the structural strain in a carbon nanocone, which is a natural strain related to the cone structure. The local strain in this structure is mapped in Figure 3(c). The colored points in Figure 3(c) indicate the local G band wavenumber (ω_G) measured at that position. The strain can be monitored locally with Raman spectroscopy because it causes the redshift in ω_G.

Functional strain in graphene under nanomanipulation

Tribology aspects related to the adhesion of nanostructures to substrates can also be monitored using Raman spectroscopy during nanomanipulation.[26] Figure 4(a) shows a schematic view of an atomic force microscope (AFM) tip inducing radial strain on a graphene flake suspended over a circular hole made on a SiN_x substrate. Figure 4(b,c) plots ω_G as a function of the AFM-induced strain in 2 physically distinct samples. The red bullets stand for strain loading, and the open circles for strain unloading.

The change in behavior for the G band loading procedure shown in Figure 4(b), taking place when strain is above 0.1% indicates partial slip of the graphene with respect to the substrate, which causes a relaxation of the structure, followed by a blueshift of ω_G.

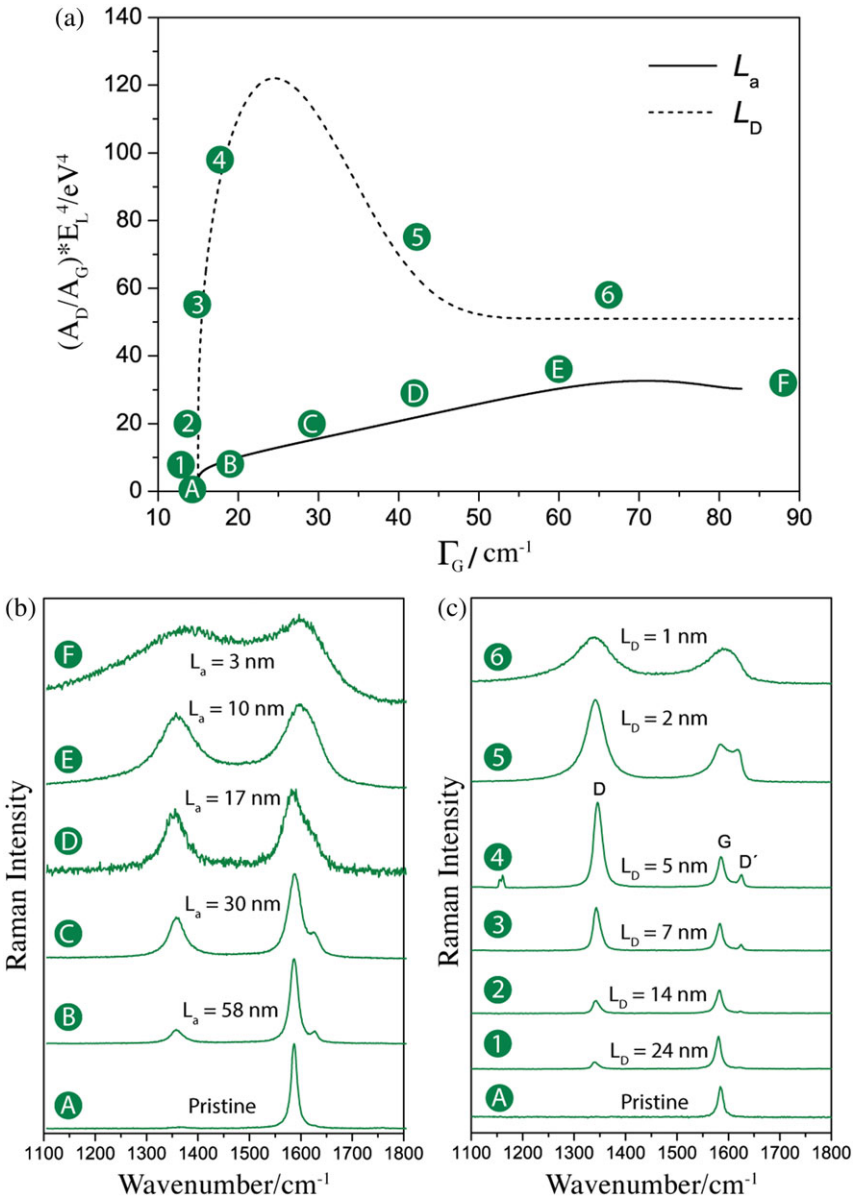

Figure 1. (a) sp^2 carbon amorphization Raman diagram. (b,c) Exemplary spectra from the reference standard samples with a well-controlled amount of line-like defects (b) and point-like defects (c). Average distance among line defects (L_a in b) and between point-like defects (L_D in c) are indicated above the respective spectra. The letters and number inside filled-circles in (a) are the diagram locations where the respective labeled spectra in (b,c) belong to. Reproduced from 2D Materials 4, 2017, 025039[24].

A second experiment is shown in Figure 4(c), on a similar sample, but here slip takes place at a larger strain value (above 0.5%), indicating stronger local adhesion of this graphene sample to the substrate. In this case, when slip takes place, the strain recoil is complete rather than partial, and the G band goes back to the unstrained value.

In similar experiments on different suspended graphene samples, an increase in the D band signal is observed. Sometimes, the usually strong G band signal disappears, giving rise to an amorphous carbon signal (Figure 4(d)) that actually comes from the AFM tip. These observations indicate graphene amorphization and rupture during the experiment, with the deposition of carbon on the tip.

The results shown in Figure 4(c) indicate that different samples exhibit different graphene-substrate adhesion forces. Connection of the observed G band shift with elasticity theory can provide quantitative values for the adhesion forces.[26]

Parameterized principal component analysis

When scaling up applications from laboratory proof-of-concept to industrial production, automated routines able to efficiently analyze hundreds or thousands of unlabeled spectra may be needed. One of the most well-established routines for such endeavor is PCA.[23]

The principal components (PCs) are used most frequently as means of constructing an informative graphical representation of the data, and as an exploratory technique for multivariate data. Often, its operation can be thought of as revealing the internal

Journal of RAMAN SPECTROSCOPY

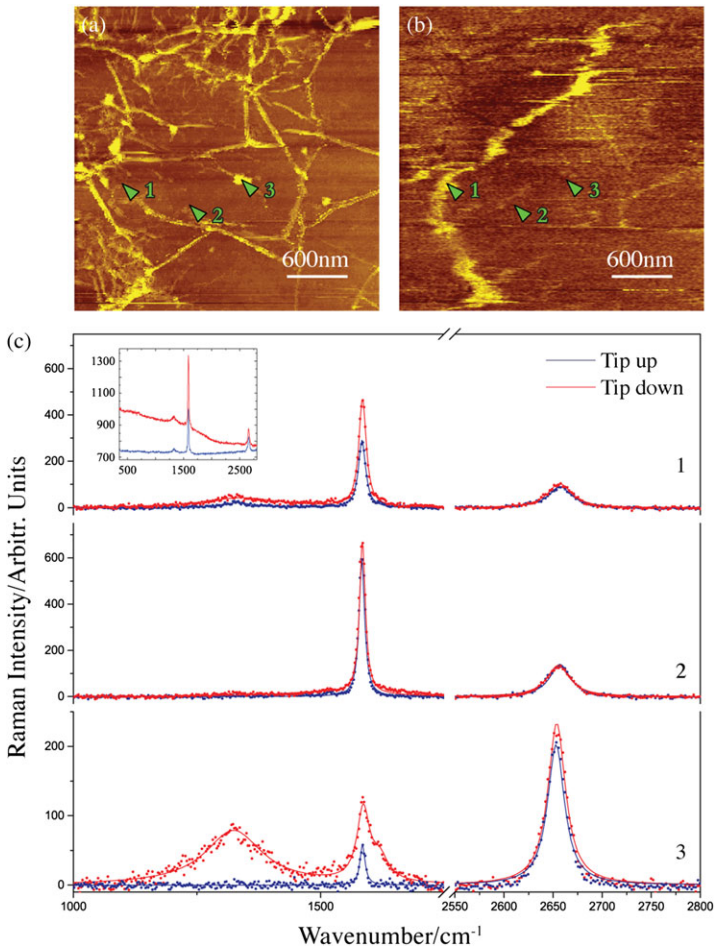

Figure 2. Simultaneous STM (a) and TERS (b) images of graphene. For the TERS imaging, the spectral energy between 1250 and 1780 cm^{-1} (D and G bands) is being recorded. (c) Raman spectra at 3 different locations in (a,b), as indicated by numbers 1, 2, and 3, after background subtraction. Red and black traces stand for spectra obtained with the TERS tip (tip down) and without the TERS tip (tip up, equivalent to a confocal configuration). The inset shows the spectra from point #1 over a larger spectral range, without background subtraction.

structure of the data in a way that best explains the variance in the data. As such, PCA has the drawback of depending on the particular batch of data and of carrying, in principle, no clear physical information. However, PCA can be parameterized to overcome these problems. Here, we discuss the development of parameterized PCA for Raman spectroscopy applied to sp^2 carbon structures.

Basic definitions for PCA

The mathematical background for PCA is well established in the literature,[23] and various open source programming tools are available for its implementation. However, for an understanding of the logics implemented here, it is interesting to qualitatively discuss some basic definitions.

Consider a data matrix with *n* columns (samples) and *m* rows (variables). PCA is mathematically defined[23] as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some projection of the data lies on the first coordinate (first principal component—PC1), the second greatest variance lies on the second coordinate (second principal component—PC2), and so on. The eigenvectors (principal components, PCs) determine the directions of the new feature space, and the eigenvalues determine their magnitude. In other words, the eigenvalues explain the variance of the data along the new feature axes.

The typical goal of a PCA is to reduce the dimensionality of the original feature space by projecting it onto a transformed space, where the eigenvectors will form the axes. In order to decide which eigenvector(s) can be dropped without losing too much information for the construction of a lower-dimensional subspace, we need to inspect the corresponding eigenvalues. The eigenvectors with the lowest corresponding eigenvalues bear the least information about the distribution of the data and can be dropped. In order to do so, the common approach is to rank the eigenvalues from highest to lowest in order.

After sorting, the next step is to determine how many principal components are going to be selected for the new feature subspace. A useful measure is the so-called "explained variance", which can be calculated from the eigenvalues. The explained variance tells us how much information (variance) can be attributed to each of the principal components.

Because PCA yields a feature subspace that maximizes the variance along the axes, it makes sense to standardize the data, especially if it was measured on different scales. Whether or not the data should be standardized prior to a PCA on the covariance matrix depends on the measurement scales of the original features. For Raman spectroscopy, removal of artifacts, such as random

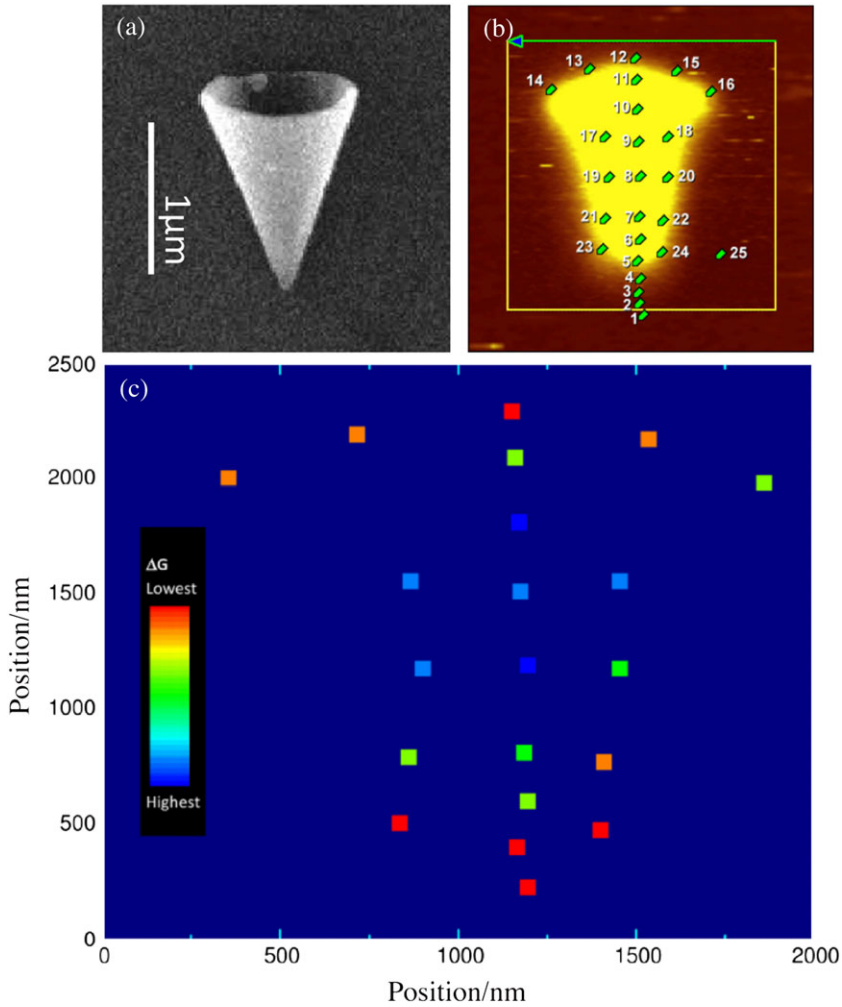

Figure 3. (a) SEM image of a carbon nanocone. (b) Raman spectral image of a carbon nanocone, determined from the intensity of the G band as a function of sample location (confocal resolution is ~300 nm). The green pointers indicate places where the G band wavenumber (ω_G) was measured and used to build the strain map drawn by the colored squared symbols in (c), which indicate the variations in ω_G. The lower the ω_G, the higher the local strain. Points numbered 1, 2, 3, and 25 in (b) are out of the cone and not shown in (c).

peaks in the spectrum coming from cosmic rays, may be important. If absolute intensity is not a meaningful information (usually it is related to system alignment), it may also be important to normalize the dataset.

Parameterized PCA for structural analysis of sp^2 carbon structures

Figure 1 in this paper describes how Raman spectroscopy can be analyzed to characterize the possible amorphization trajectories of sp^2 carbon structures, from pristine graphene to amorphous carbon. Figure 5 shows a PCA analysis of the same data utilized to develop Figure 1,[24] considering the 2 PCs with the most variances.

In short, PCA works here in the following manner: one starts by constructing an $n \times m$ matrix, where each of the n lines indicates a sample spectrum, and each of the m columns indicates a wavenumber. In the current case n = 25 and m = 1101 (ranging from 800 cm^{-1} to 1900 cm^{-1}, in steps of 1 cm^{-1}). The analysis consists of performing a mathematical transformation in this coordinate space so as to rewrite each Raman spectrum as a linear combination of principal components PCi, where i = 1, 2,…, m. Each PC is itself in the form of a Raman spectrum, and the method extracts the most important spectral characteristics that capture spectral variations from sample to sample.[23] For instance, the inset to Figure 5(a) shows the principal component variance for the 10 first PCs, which indicates the relative importance of each PC for describing the spectral variation within the data set. The 2 first components (PC1 and PC2) are responsible for more than 90% of the variance in the spectral data, from pristine to amorphous material.

The similarity between Figures 1(a) and 5 (2 distinct amorphization routes, one for point-like, and another for line-like defective samples) shows that the parameters analyzed in Figure 1 [$(A_D/A_G{*}E_L^4), \Gamma_G$] reveal the main differences among the analyzed spectra, i.e., Γ_G and A_D/A_G are responsible for most of the data variance. The identification of a reduced number of relevant variables that describe variations among spectra is the advantage of using PCA, and the fact that we are disentangling 2 physical parameters (L_D, L_a) is consistent with the need of 2 spectral parameters or 2 PCs.

In more detail, Figure 5(a) shows a plot of the PC1 and PC2 weights for 25 spectra. Notice that this plot generates 2 different trajectories, indicated by the full and open symbols (samples with

Journal of RAMAN SPECTROSCOPY

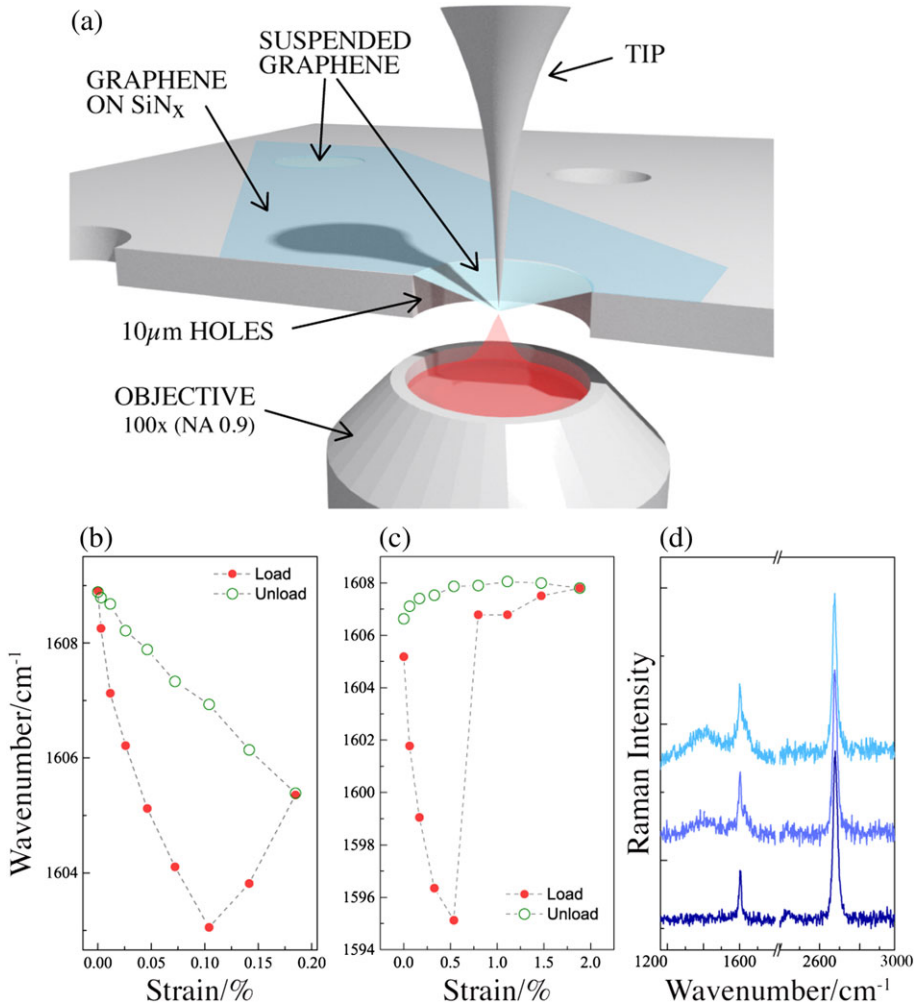

Figure 4. (a) Schematic view of the system used to perform graphene nanomanipulation to study strain effects and adhesion to the SiO_x substrate. (b,c) Plots of the G band wavenumber (ω_G) during 2 strain load (red bullets) and unload (open circles) experiments. (d) Raman spectra obtained at the AFM tip after a few cycles of load-unload experiments where: a reproducible process took place (bottom); graphene break is observed once (middle), graphene break is observed twice (top). Notice spectral evidence for amorphous carbon deposition at the AFM tip.

point-defects and line-defects, respectively), which are coincident at the 2 extremes, corresponding to fully ordered graphene/graphite (L_a,$L_D \rightarrow \infty$, right side of the plot) and highly disordered carbon (L_a,$L_D \rightarrow 0$, left side of the plot). Figure 5(b,c) shows the spectral composition of PC1 and PC2, respectively, depicting the information contained in these 2 principal components. The highly ordered/amorphous material appears in the right/left extremes of the plots because, after the PCA procedure, their spectral decomposition exhibits the largest positive/negative PC1 eigenvalues. The similarity with Figure 1 indicates that the variational diagram shown in Figure 5(a) can be roughly reconstructed based on 2 well-defined spectral features: the G-band linewidth Γ_G replacing PC1, and the integrated area ratio between the D and the G peaks A_D/A_G replacing PC2.

One aspect that is not addressed in the PCA depicted in Figure 5 is the well-known excitation laser energy (E_L) dependence of the A_D/A_G ratio,[36] considered in the treatment of Figure 1 by normalizing the *Y* axis in the amorphization diagram with the E_L^4 term. Figure 6(a) gives the PCA result (also the 2 first PCs) for the same set of samples analyzed in Figure 5, but measured with up to 4 different laser excitation energies (see figure legend). Notice that, for a given excitation laser energy (open and filled symbols of same color), the PC1 × PC2 trend is kept, but when changing E_L (open and filled symbols of different colors), the value of PC2 is predominately changed, in agreement with the major dependence of A_D/A_G on E_L.

The concept of "parameterization" of the PCA is now introduced: Figure 6(b) has the same PCA data set space of Figure 6(a), showing only the data for the sample with line defects measured with the 633-nm laser (red open circles), plus the new data from 14 Raman spectra measured in different locations of the CVD graphene sample shown in Figure 2 (black bullets), with a key difference: here, PCA does not calculate a new space that maximizes the variances among the spectra. Instead, it keeps the same orthogonal linear transformation procedure defined by the standard samples in Figure 6(a), and it simply projects the new data set (black bullets) onto this space, as shown in Figure 6(b).

Notice that the CVD-graphene data (black bullets) falls along the red-solid line in Figure 6(b), which is the one defined by the standard sample with line-like defects, measured with the excitation laser wavelength 633 nm. This result is exactly as expected, considering the new data introduced in Figure 6(b) were measured with a 633-nm excitation laser, and the sample is composed mostly of graphene nanocrystallites separated by

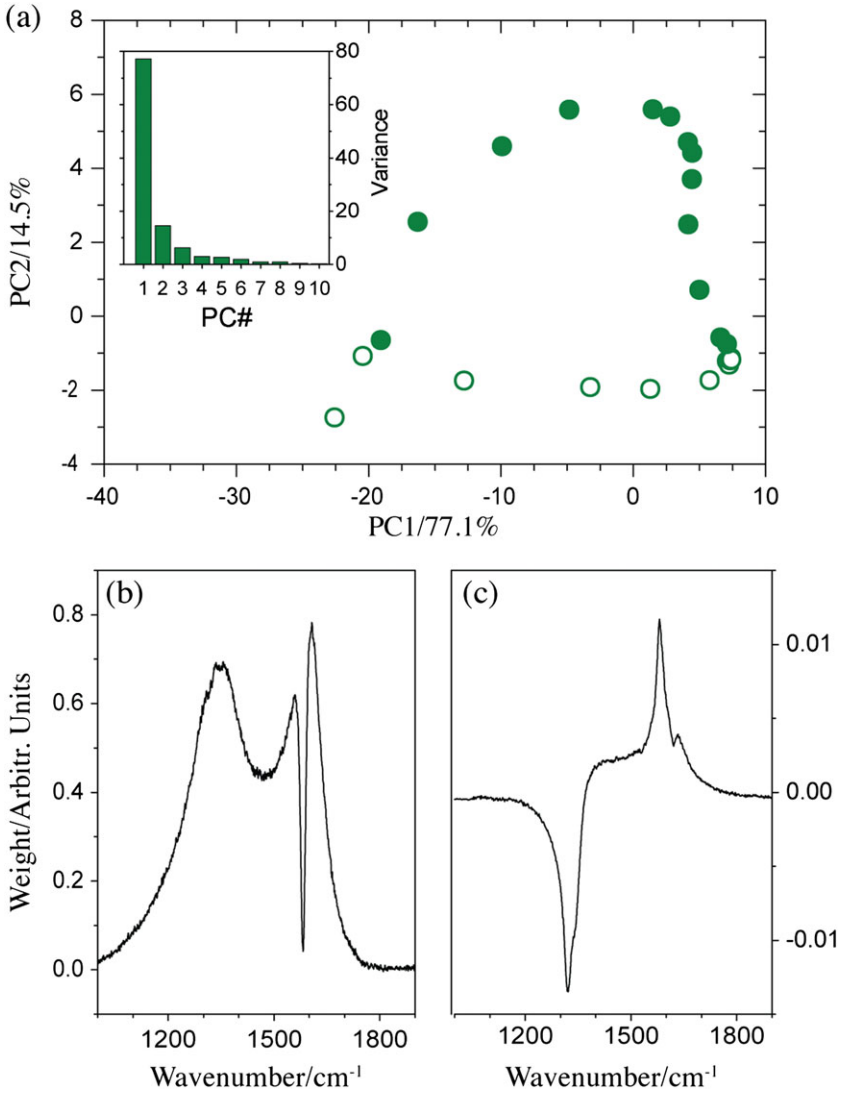

Figure 5. (a) Principal component analysis (PCA) of Raman spectra of graphene samples varying from perfectly crystalline to fully disordered, following 2 routes: increasing the number of point defects (filled symbols—ion bombardment of mechanically exfoliated graphene)[19,20,24] and increasing the number of line defects (open symbols—amorphous carbon heat treated at different temperatures).[14,30] Spectra were obtained with excitation laser energies E_L = 2.41 eV (514 nm, sample with point-defects, filled symbols) and E_L = 2.33 eV (532 nm, sample with line-defects, open symbols). The spectral contribution (weight) from the first and second principal components (PC1 and PC2, respectively) to the 25 spectra analyzed is shown. The inset shows the spectral variance for the 10 first principal components. (b) and (c) show the spectral information within PC1 (b) and PC2 (c).

grain boundaries, i.e., graphene with line-like defects (see Figure 2(a)). The different spectra are measurements in different locations in the sample, with more or fewer line defects, as indicated by the major variation in the PC1 weight for each black bullet (see exemplary spectra in the insets to Figure 6(b)). Therefore, the PCA with a parameterized set projects a sample in a space with well-defined characteristics, generating a comprehensible result.

The parameterized PCA is now used to analyze different types of sp^2 carbon materials, namely CVD graphene ("CVD" in the legend, same sample as in Figure 2),[27] graphite, plasma enhanced-CVD graphene from methane ("peCVD" in the legend),[38] and 2 types of biochar—a natural deposit ("Rio Negro") and the anthropogenic material TPI.[28–30] Figure 7(a) shows a free (non-parameterized) PCA analysis. Notice the samples are split according to the different samples, but there is no understanding of what the PC1 and PC2 values mean. There is no clear distinction between TPI and Rio Negro. For an unknown reason, the data from sample "peCVD" splits in 2 groups.

Figure 7(b) shows a PCA of the same set of samples shown in Figure 7(a), but now using the parameterization imposed by the standard material used to develop Figures 5 and 6. By correlating the results of Figure 7(b) with Figures 5(a) and 6, the information obtained from PCA is now clear. Graphite is the perfectly ordered structure, appearing at the right side of the PC1 × PC2 plot. The CVD sample has the addition of line defects (grain boundaries); therefore, it departs towards the left side of the plot changing mostly in PC1, as already discussed in Figure 6. The sample "peCVD" has more point defects probably due to the use of natural gas as precursor, consequently departing along the PC2 axis. Notice the splitting observed in the free PCA shown in Figure 7(a) for the data of sample "peCVD" is diminished in the parameterized PCA shown in Figure 7(b), indicating a unique structural characteristic for this type of material. Samples TPI

Journal of RAMAN SPECTROSCOPY

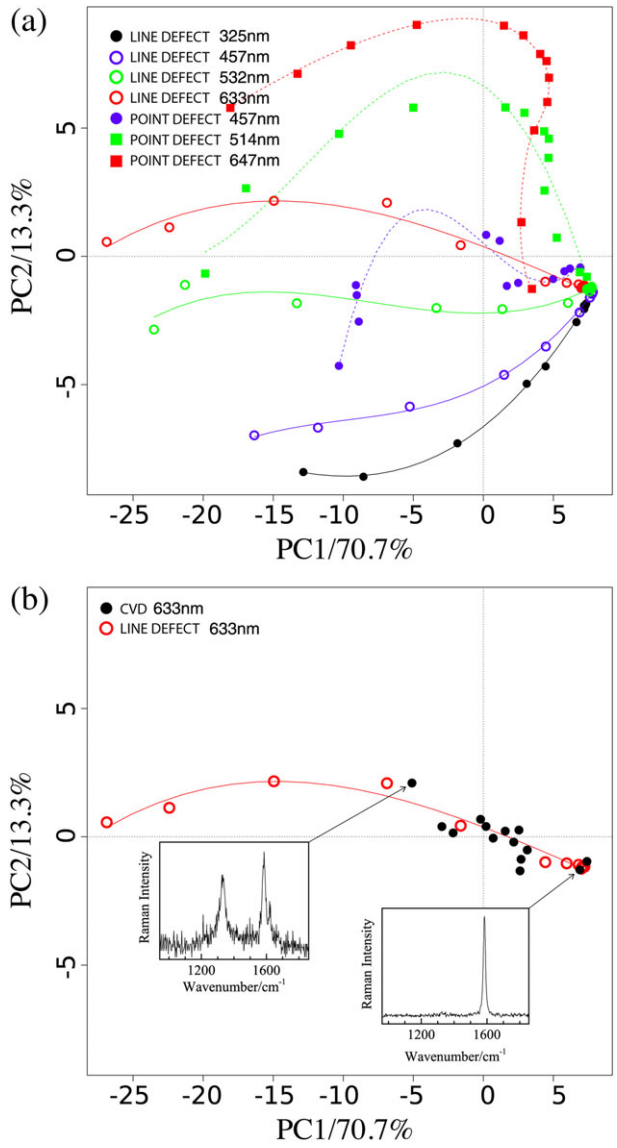

Figure 6. (a) Similar PCA analysis as in Figure 5, but now for Raman spectra acquired with different excitation laser energies (red, 633 and 647 nm; green, 514 and 532 nm; blue, 457 nm and UV, 325 nm—see legends). (b) Similar procedure as in (a), showing only the data for line defects with 633 nm and adding the data (black bullets) from the CVD graphene sample shown in Figure 2. The insets are representative spectra.

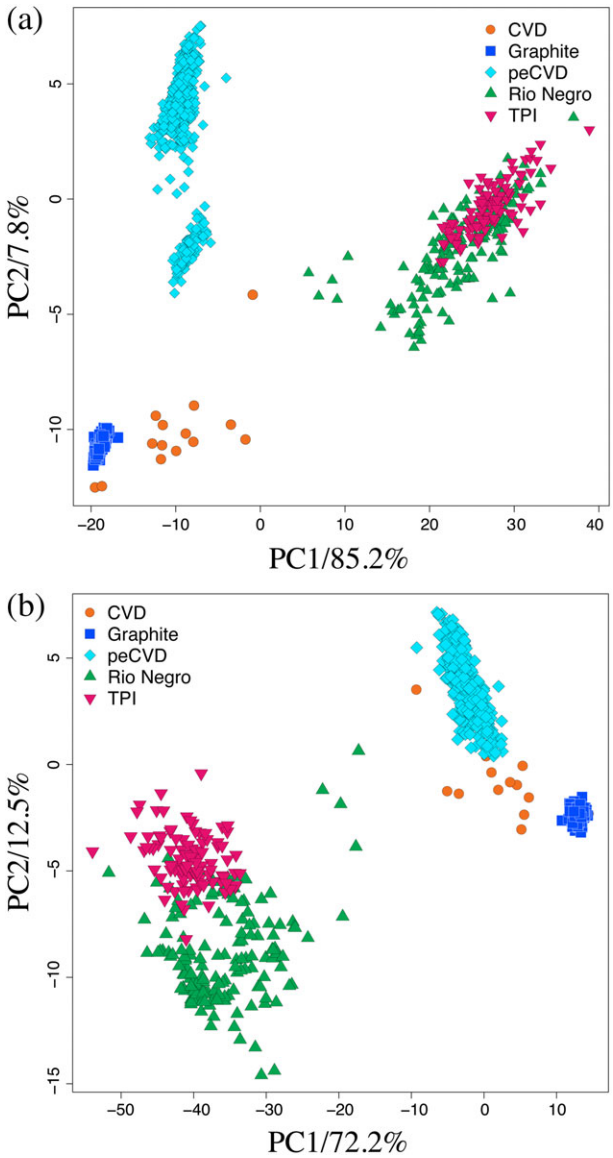

Figure 7. (a) PC1 × PC2 plot after PCA of 5 different types of samples: CVD, graphite, peCVD, Rio Negro, and Terra Preta de Índio (TPI). (b) Same procedure as in (a), but now with the orthogonal linear transformation parameterized by the standard graphene samples used to develop the Raman amorphization diagram in the PCA of Figure 6(a).

and Rio Negro appear in the extreme left of the plot, as expected, due to the highly amorphous structure of biochar. Furthermore, there is a more clear separation among these 2 groups of biochar. The TPI samples appear shifted towards larger values of PC2 when compared with Rio Negro sample, which indicates TPI has a larger number of point defects. This is consistent considering TPI biochar is composed by a highly reactive carbon material, ideal for soil enrichment.

Parameterized versus non-parameterized PCA

Here, we provide another example of the difference between parameterized and non-parameterized PCA. Figures 3 and 4 show examples of structural and functional analysis of strain in graphene-related systems, respectively. Figure 8 shows PC1 × PC2 plots resulting from a non-parameterized PCA of the spectral measurements from Figure 3 (panel (a)), Figure 4 (panel (b)), and from both data together (panel (c)). Figure 8(d) shows the result of the **parameterized** PCA of both data (from Figures 3 and 4) together.

As already discussed, the non-parameterized PCA generates a plot in a space that maximizes the differences among the samples, whatever these differences are. The data from Figure 3 comes from a well-defined structure, where different strains in the sp^2 carbon network are found in different locations in the carbon nanocone. The data from Figure 4 come from well-controlled deformation-induced experiments, where the changes are also related to strain in the graphene structure. Therefore, we can understand the routes formed in the PC1 × PC2 plots of Figure 8(a,b) as related to strain, which is spectrally evidenced by changes in ω_G. However, if we

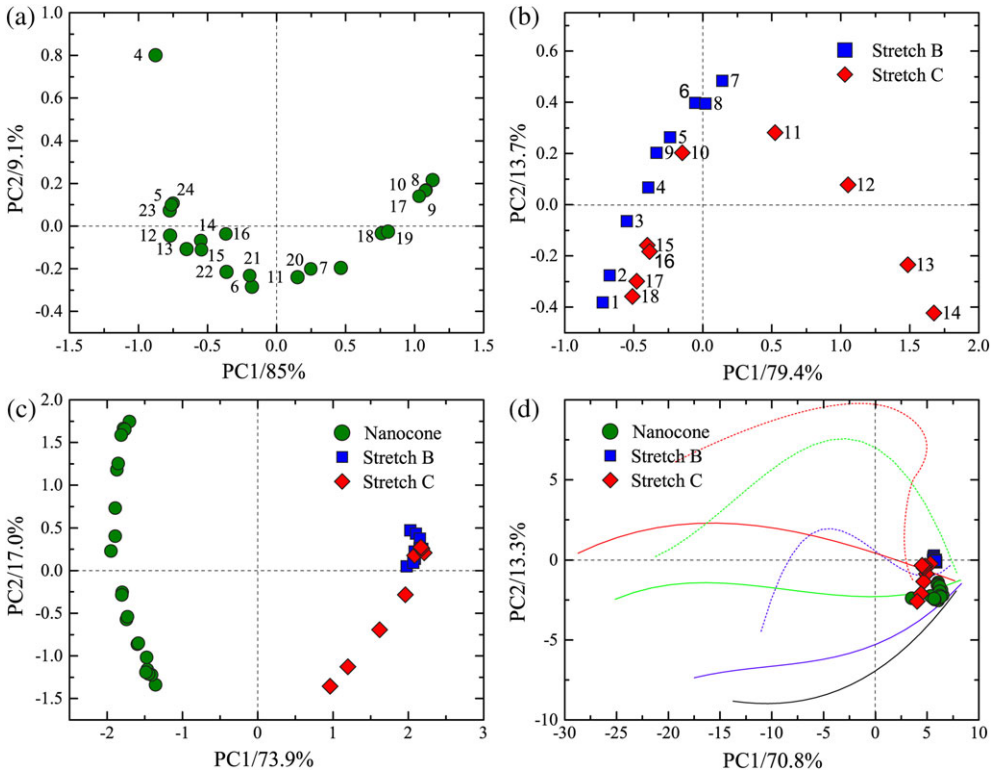

Figure 8. PC1 × PC2 plots of: (a) strain load data from Figure 3; (b) strain load data from Figure 4; (c,d) strain load data from both Figures 3 and 4 together. (a-c) are the result from non-parameterized PCA; (d) same as in (c) but using the PCA parameterized for structural analysis of sp^2 carbon structures. Lines are the amorphization routes from Figure 6(a). Numbers in (a) indicate the location in the nanocone, as displayed in Figure 3(b). Points 1, 2, 3, and 25 are absent because they were acquired outside the cone. Numbers in (b) label the data according to the loading procedures in Figure 4b,c. Data 1 to 9 correspond to the loading data in Figure 4(b); Data 10 to 18 correspond to the loading data in Figure 4(c).

had no information about what were the differences among those spectra, the only information we could extract from the PCA plots would be that the spectra differ somehow, and that they form 2 distinct groups (clearly shown in Figure 8(c)). The reason why there are 2 distinct groups in Figure 8(c) is because of the unusually high values observed for ω_G in the spectra from Figure 4 (~1600 cm^{-1} rather than ~1580 cm^{-1}), probably due to graphene doping within this device.

Differently, if you use the parameterized PCA developed in the previous subsection, the data collapse on the right extreme of the sp^2-amorphization space, as shown in Figure 8(d). There is, therefore, clear and well-defined information being displayed: the spectra come from highly ordered graphitic materials, with a small amount of defects.

In conclusion, the parameterized PCA discussed here generates meaningful results and can be utilized to classify the type of sp^2 carbon structure according to the amount of point-like and line-like defects, from pristine graphene to amorphous carbon structures. When using the sp^2 structure parameterization procedure, other information, such as strain, may be lost, because the chosen parameterization space just does not emphasizes it. One could, however, use the parameterization procedure, and the data from Figure 3 or Figure 4 to generate a parameterization space for strain.

Parameterized PCA for number of layers in graphene

To show the versatility of the PCA parameterization procedure, a different problem is now addressed, where a 3-dimensional space is needed: number of layers rather than sp^2 structural ordering.

Figure 9(a,b) shows 2 Raman spectral images of a graphene sample deposited in a SiO_x substrate using the exfoliation method from HOPG. The images are built by raster scanning the graphene sample to acquire the Raman spectra as a function of sample position, and then plotting the local intensity of (a) G (1580 cm^{-1}) and (b) G′ (2700 cm^{-1}) bands. The G′ band is also commonly named 2D in the literature. The relative intensities between the G and G′ bands and the G′ lineshape are known to depend on the number of graphene layers in the sample,[37] and these features have been broadly used for the determination of the number of layers.

Each pixel in Figure 9(a,b) represents a full Raman spectrum. We performed a PCA of this set of 4096 spectra, each spectrum composed of 1024 data points, each data point comprising the Raman shift and Raman intensity. Figure 9(c,d) shows a plot of the weights of the 3 first principal components (PC1, PC2, and PC3), which are found to be representative in the variance analysis (more than 56% of dataset variance). PCA thus generates 4 distinct groups in this 3-dimensional PC space, indicating that the sampled material has 4 distinct categories of materials. The 4 groups are connected by a continuous evolution of points, indicating a smooth transition from a category of material to the other along the PCs, which is, in principle, unexpected.

To understand the meaning of the PCA weight plot presented in Figure 9(c,d), we first look into the spectral weight that each PC assigns to the data, as shown in Figure 9(e,f,g), for PC1, PC2, and PC3, respectively. Although the 3 PCs are clearly focused on the main Raman features, it is not yet trivial to understand what kind of information is hidden into PC1, PC2, and PC3. We then look into

Journal of RAMAN SPECTROSCOPY

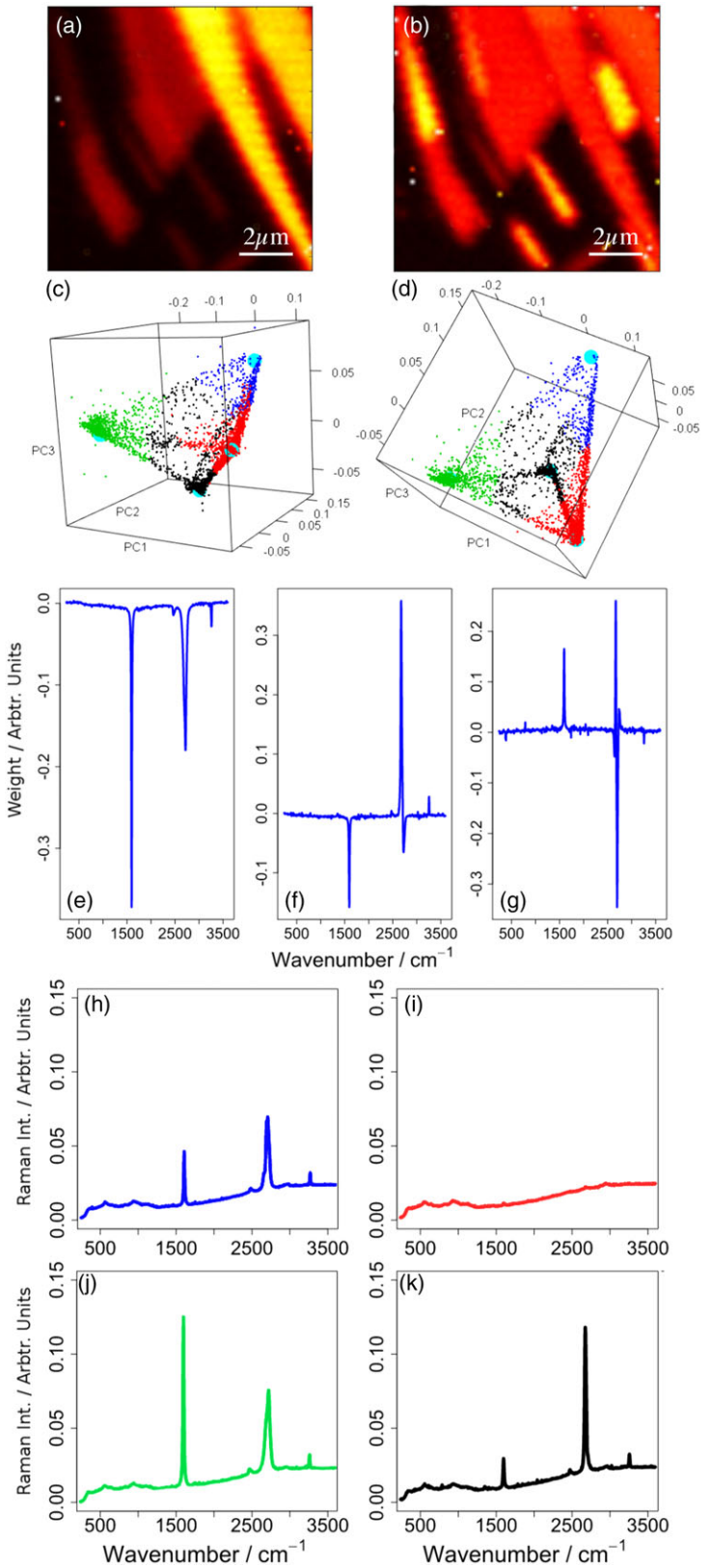

Figure 9. (a) G band hyperspectral image of a graphene deposited on a SiO_x substrate, using the mechanical exfoliation method from an HOPG starting material. (b) Corresponding G′ (or 2D) hyperspectral image. Every pixel in (a,b) contains a full Raman spectrum at a given sample location, and the PCA of all these spectra generates the 2 weight plots in (c,d), where the first 3 principal components (PC1, PC2, and PC3) were considered. (d) is the same as (c), only rotating the coordinates to change the view. (e,f,g) The spectral representation of the 3 first principal components, PC1, PC2, and PC3, respectively. (h,i,j,k) Representative spectra from the vertices of the PCA tetrahedral in (c,d), obeying the same color code.

the reconstructed spectra at the vertices of the tetrahedral, representative of the 4 categories that have been separated in the PCA (see Figure 9h-k). Figure 9(h) has the typical spectral shape of the Raman spectrum of multilayer graphene; Figure 9(i) is typical of the substrate; Figure 9(j) represents the typical spectral shape of the Raman spectrum of bilayer graphene; Figure 9(k) represents the typical Raman spectrum of monolayer graphene. After this inspection of the 4 different categories of materials' response, it is now possible to elucidate the information contained in the PCs.

PC1 carries the largest variance among the PCs, and it represents the difference between the substrate (just background) and the samples (graphene spectra with G and G′ peaks). PC2 carries weight relative to the intensity relation between the G and G′ peaks, while PC3 carries weight relative to the internal structure of the G′ peaks. Therefore, PC2 and PC3 carry the main attributes known to differentiate single-layer, bi-layer, and many-layer graphene samples.[39]

With respect to the points connecting the tetrahedral vertices in the PCA, inspection of the data reveals that the linear transition between categories does not reflect a property of the material itself, but it actually reflects a limitation of the measurement system. Specifically, the spatial resolution of the confocal Raman spectroscopy used to raster scan the sample is limited, and when moving from 1 category of material to the other, the spectral information from both materials is acquired.

The data from this sample can then be used to parameterize PCA as a method for differentiating the number of layers in any graphene sample. The number of data points in the transitions between groups is an indicator of the measurement system's spatial resolution.

Summary and perspectives

The parameterization of PCA, as presented here, provides a methodology for large-scale analysis and characterization of a wide variety of sp^2 carbon materials. In short, with the Raman data from a set of standard samples that spans the possible variations of specific well-known properties, it is possible to find the orthogonal linear transformation of the data following a PCA algorithm, which maximizes the ability of classification of an unknown Raman spectrum within the space spanned by the desired properties.

Such an automated procedure is required if one wants to apply Raman spectroscopy to industrial procedures, including growth, functionalization, composites synthesis, or large-scale device fabrication using graphene-related materials. Here, we demonstrated 2 case studies: (1) the amorphization structure of sp^2 carbons, parameterized with the amount of point-like and line-like defects in the 2-dimensional structure, and (2) the number of layers in graphene samples. These 2 examples of parameterized PCA are ready to be used for the quality control of graphene growth via CVD or chemical/physical exfoliation methods or to monitor the production of biochar. However, to implement the procedure utilizing any available programming tool, it is necessary to have in hands either the set of standard spectra that will be utilized to parameterize the PCA analysis, or the transformation matrices generated by the standard data.

The concept of parameterized PCA is not limited to the examples given here, and it is not limited to graphene-related material. It is, in general, an automated model to allow Raman spectroscopy large-scale analysis with rich interpretability.

Acknowledgments

This article is devoted to Professor Mildred S. Dresselhaus, who gave enormous contribution to the field of Raman Spectroscopy in carbon-related materials. We strongly acknowledge technical assistance from Prof. Palash Bharadwaj on the TERS experiments. We also acknowledge fruitful discussion with Profs. Luiz Gustavo Cançado and Rodrigo Barbosa Capaz. This work was financed by CNPq, INMETRO, and the US Army Research Development and Engineering Command (RDECOM) International Technology Center.

References

[1] J. Ribeiro-Soares, M. S. Dresselhaus, *Brazilian Journal of Physics.* **2014**, *44*, 278.
[2] M. S. Dresselhaus, G. Dresselhaus, *Advances In Physics.* **1981**, *30*, 139.
[3] J. Robertson, *Materials Science and Engineering R: Reports.* **2002**, *37*, 129.
[4] M. S. Dresselhaus, G. Dresselhaus, P. C. Eklund, *Science of Fullerenes and Carbon Nanotubes: Their Properties and Applications*, Academic Press, **1996**.
[5] R. Saito, G. Dresselhaus, M. S. Dresselhaus, *Physical Properties of Carbon Nanotubes*, Imperial College Press, **1998**.
[6] S. Reich, C. Thomsen, J. Maultzsch, *Carbon Nanotubes: Basic Concepts and Physical Properties*, John Wiley & Sons, **2008**.
[7] A. Jorio, G. Dresselhaus, M. S. Dresselhaus, *Topics in Applied Physics.* **2008**, 111.
[8] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, I. V. Grigorieva, S. V. Dubonos, A. A. Firsov, *Nature* **2005**, *438*, 197.
[9] J. Lehmann, S. Joseph, *Biochar for Environmental Management: Science, Technology and Implementation*, Routledge, **2015**.
[10] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, A. A. Firsov, *Science* **2004**, *306*, 666.
[11] J. Ribeiro-Soares, R. M. Almeida, L. G. Cançado, M. S. Dresselhaus, A. Jorio, *Physical Review. B, Condensed Matter and Materials Physics.* **2015**, *91*, 205421.
[12] J. Guan, Z. Zhu, D. Tománek, *Phys. Rev. Lett.* **2014**, *113*, 046804.
[13] S. Zhang, Z. Yan, Y. Li, Z. Chen, H. Zeng, *Angew. Chem. Int. Ed.* **2015**, *54*, 3112.
[14] P. Vogt, P. Padova, C. Quaresima, J. Avila, E. Frantzeskakis, M. C. Asensio, G. Le Lay, *Phys. Rev. Lett.* **2012**, *108*, 155501.
[15] M. E. Dávila, L. Xian, S. Cahangirov, A. Rubio, G. Le Lay, *New J. Phys.* **2014**, *16*, 095002.
[16] B. Feng, O. Sugino, R. Liu, J. Zhang, R. Yukawa, M. Kawamura, T. Iimori, H. Kim, Y. Hasegawa, H. Li, L. Chen, K. Wu, H. Kumigashira, F. Komori, T. Chiang, S. Meng, I. Matsuda, *Phys. Rev. Lett.* **2017**, *118*, 096401.
[17] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, M. S. Strano, *Nat. Nanotechnol.* **2012**, *7*, 699.
[18] J. Ribeiro-Soares, R. M. Almeida, E. B. Barros, P. T. Araujo, M. S. Dresselhaus, L. G. Cançado, A. Jorio, *Physical Review B.* **2014**, *90*, 115438.
[19] A. Jayaraman, V. Narayanamurti, E. Bucher, R. G. Maines, *Phys. Rev. Lett.* **1970**, *25*, 1430.
[20] A. K. Geim, I. V. Grigorieva, *Nature* **2013**, *499*, 419.
[21] A. C. Ferrari, J. Robertson, *Philos. Trans. R. Soc. A.* **2004**, *362*, 2477.
[22] A. Jorio, M. S. Dresselhaus, R. Saito, G. Dresselhaus, *Raman Spectroscopy in Graphene Related Systems*, Wiley-VCH, Germany, **2011**.
[23] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, **2002**.
[24] L. G. Cançado, M. G. Silva, E. H. M. Ferreira, F. Hof, K. Kampioti, K. Huang, A. Pénicaud, C. A. Achete, B. R. Capaz, A. Jorio, *2D Materials.* **2017**, *4*, 025039.
[25] A. G. C. Marques, W. G. Schimidt, J. Ribeiro-Soares, L. G. Cançado, W. N. Rodrigues, A. P. Santos, C. A. Furtado, P. A. S. Autreto, R. Paupitz, D. S. Galvao, A. Jorio, *Sci. Rep.* **2015**, *5*, 10408.
[26] U. C. P. A. M. Mussnich, H. Chacham, J. S. Soares, N. M. B. Neto, N. Shadmi, E. Joselevich, L. G. Cançado, A. Jorio, *Nano Lett.* **2015**, *15*, 5899.
[27] J. Riikonena, W. Kima, C. Lia, O. Svenska, S. Arpiainenb, M. Kainlaurib, H. Lipsanena, *Carbon.* **2013**, *62*, 43.
[28] A. Jorio, J. Ribeiro-Soares, L. G. Cançado, N. P. S. Falcão, H. F. Dos Santos, D. L. Baptista, E. H. Martins Ferreira, B. S. Archanjo, C. A. Achete, *Soil Tillage Res.* **2012**, *122*, 61.

[29] M. C. Pagano, J. Ribeiro-Soares, L. G. Cançado, N. P. S. Falcão, V. N. Goncalves, L. H. Rosa, J. A. Takahashi, C. A. Achete, A. Jorio, *Soil Tillage Res.* **2016**, *155*, 298.
[30] J. Ribeiro-Soares, L. G. Cançado, N. P. S. Falcão, E. H. Martins Ferreira, C. A. Achete, A. Jorio, *Journal of Raman Spectroscopy.* **2013**, *44*, 283.
[31] M. S. Ladeira, V. A. Andrade, E. R. Gomes, C. J. Aguiar, E. R. Moraes, J. S. Soares, E. E. Silva, R. G. Lacerda, L. O. Ladeira, A. Jorio, P. Lima, M. F. Leite, R. R. Resende, S. Guatimosim, *Nanotechnology* **2010**, *21*, 385101.
[32] M. Munk, L. O. Ladeira, B. C. Carvalho, L. S. A. Camargo, N. R. B. Raposo, R. V. Serapiao, C. C. R. Quintao, S. R. Silva, S. Soares Jaqueline, A. Jorio, H. M. Brandão, *Sci. Rep*. **2016**, *6*, 33588.
[33] A. Jorio (Ed), *Bioengineering Applications of Carbon Nanostructures*, Springer, **2016**.
[34] A. Jorio, *ISRN Nanotechnology*. **2012**, 2012.
[35] B. S. Archanjo, I. O. Maciel, E. H. Martins Ferreira, S. B. Peripolli, J. C. Damasceno, C. A. Achete, A. Jorio, *Ultramicroscopy* **2011**, *111*, 1338.
[36] L. G. Cançado, K. Takai, T. Enoki, M. Endo, Y. A. Kim, H. Mizusaki, A. Jorio, L. N. Coelho, R. Magalhães-Paniago, M. A. Pimenta, *Appl. Phys. Letters.* **2006**, *88*, 163106.
[37] M. M. Lucchese, F. Stavale, E. M. Ferreira, C. Vilani, M. V. O. Moutinho, R. B. Capaz, C. A. Achete, A. Jorio, *Carbon* **2010**, *48*, 1592.
[38] F. Hof, K. Kampioti, K. Huang, C. Jaillet, A. Derré, P. Poulin, H. Yusof, T. White, K. Koziol, C. Paukner, A. Pénicaud, *Carbon.* **2017**, *111*, 142.
[39] A. C. Ferrari et al., *Phys. Rev. Lett.* **2006**, *97*, 187401.