Research Article

Denoising and Principal Component Analysis of Amplified Raman Spectra from Red Blood Cells with Added Silver Nanoparticles

L. Ferrer-Galindo,¹ A. D. Sañu-Ginarte,¹ N. Fleitas-Salazar,² L. A. Ferrer-Moreno,³
R. A. Rosas,² M. Pedroza-Montero,² and R. Riera^{ID}²

¹Departamento de Física, P.O. Box 1626, Universidad de Sonora, 83000 Hermosillo, SON, Mexico

²Departamento de Investigación en Física, P.O. Box A-088, Universidad de Sonora, 83190 Hermosillo, SON, Mexico

³Departamento de Ciencias Básicas, Instituto Tecnológico de Hermosillo, 83170 Hermosillo, SON, Mexico

Correspondence should be addressed to R. Riera; rriera@cifus.uson.mx

Received 17 January 2018; Accepted 20 April 2018; Published 3 June 2018

Academic Editor: Bo Tan

Copyright © 2018 L. Ferrer-Galindo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Incubated erythrocytes with and without *silver nanoparticles* (AgNP) were analyzed by Raman spectroscopy, resulting in two Raman spectra datasets. AgNP were added to *red blood cells* (RBC) in order to enhance the Raman signals. This technique is known as *surface-enhanced Raman scattering* (SERS). A comparison was made between the Raman spectra with and without AgNP, to test if the SERS had taken place. Since Raman and SERS spectra are considered to be cumbersome due to the noises presented, we applied denoising criteria for detection and removal of noises like cosmic rays, shot, and fluorescence contribution. After this, the *principal component analysis* (PCA) was performed, in order to reduce the dimensions of the spectra being studied. Only the main key components necessary for a better interpretation of these spectra were considered. All of those noises had to be removed prior to the statistical analysis, to make sure the analysis was really based on the Raman measurements and not on other effects. As a result, RBC Raman spectra with and without AgNP got denoised, obtaining an improvement in its resolution for a better signal reading and data interpretation. Also, the first *principal components* (PC) were selected from each dataset under scrutiny, based on the weight of their information and their spectrum readability. In conclusion, we were able to represent the given reference system with a more affordable and smaller dimension in which information loss was minimal.

1. Introduction

The use of Raman spectroscopy for biological purposes has increased the need to interpret large datasets resulting from applying this spectroscopic technique especially due to its high sensitivity to subtle molecular changes. Although Raman spectroscopy is a powerful technique, its signals are inherently weak in biological systems. In order to record these weak signals, SERS [1–3] can be used. This is a technique that provides a signal increment of some orders of magnitude, where the enhancement factor can be as high as 10^{10} or 10^{11} .

As the Raman and SERS spectra are cumbersome by the noises presented, criteria application for detection and

removal of these noises for better data interpretation is necessary. Those noises that corrupt data are very common; we can see cosmic rays or “spikes” [4] that are produced by external sources, the “Shot” [4] is a result of random nature of light, and there is another one generated by the sample, which is known as fluorescence [4]. All these noises affect the real data from RBC Raman spectra with and without AgNP [5], masking bands that contain valuable information and must be removed.

After denoising the data, it is necessary to correct the variability and reject redundant variables. To do that, we need statistical/chemometrical methods [6]. The use of multivariate analysis techniques allows the process development and interpretation of multiple data spectra amplified by AgNP

(SERS reporters) [7] in RBC, associated with a point or cell. For a better data analysis, spectra should not show any signs of noise.

In this work, we apply denoising techniques and PCA [6, 8] as an alternative mathematical-statistical technique to improve relevant Raman signal reading (e.g., SERS), for a better interpretation of biological systems. Here, we propose an original theoretical/experimental method to enhance Raman spectra resolution in biological systems making it possible to obtain inferences and predictions about the resulting data that help with disease diagnostic and treatment.

2. Materials and Methods

2.1. Experiment. Certain noises were removed, and a multivariate analysis was applied to Raman spectra from an experiment that consisted in the incubation of erythrocytes with and without AgNP. Fresh cells obtained by blood extraction from a healthy volunteer were used. Two Raman spectra datasets were analyzed, the first one without AgNP, used as cell control, and a second one with the nanoparticles added to enhance the Raman signal.

Cell control: a 10 μL aliquot of blood was diluted 1:100 (V:V) in buffer phosphate saline (BPS). In Figure 1, it can be seen as a spectrum of RBC without AgNP.

Cells with AgNP: a small aliquot of blood was diluted 1:100 (V:V) in BPS solution. The AgNP solution was added in a final dilution 1:10. Figure 2 shows a RBC spectrum with AgNP added.

The samples were incubated for 2 h at $25 \pm 4^\circ\text{C}$. Cell samples were then prepared for Raman spectrum measurement. A drop of the cell suspension was added to a silicon substrate previously coated with poly-L-Lysine. The cells in the substrate were allowed to settle for 15 minutes to promote the immobilization of the RBC for measurement.

2.2. Instrumentation and Data Acquisition. Measurements were performed with a Raman microspectrometer (LabRAM HR, Horiba), equipped with a refrigerated CCD camera. Multiple scans were conducted, and each cell was irradiated five times at different points. The excitation wavelength was 488 nm, with a 100x objective, in “nonconfocal” conditions (1000 μm aperture). Each measurement was performed with a single exposure of 20 seconds and spectra were collected in the 2500–550 cm^{-1} range.

The Raman spectra were acquired with LabSpec software; Figure 3 shows 94 spectra of RBC without AgNP, which were taken as control in the experiment, and Figure 4 shows 89 RBC spectra with AgNP, to be analyzed and compared with the control spectra. The spectra denoising was performed using a self-created software, named *My Raman Data Processing* (MyRDP). The multivariate analysis was implemented using the Wolfram Mathematica 9 software.

2.3. Preprocessing Methods. Data was then loaded to MyRDP to remove the principal noises: spikes (cosmic rays), shot (background), and fluorescence, as shown in Figure 5. Since this was manually performed, the program was

FIGURE 1: RBC spectrum without AgNP.

FIGURE 2: RBC spectrum with AgNP added.

FIGURE 3: RBC Raman spectra without AgNP.

FIGURE 4: RBC Raman spectra with AgNP.

FIGURE 5: Principal noises in Raman spectra (consider explaining the sample).

developed in NetBeans IDE 7.0.1, which is a free integrated development environment, made mainly for the Java programming language.

The programming of this data treatment was made considering the needs for this experiment, always trying to yield the greatest possible robustness and allowing a user-friendly interaction. From the spectra, only a specific region is needed, the program allows the user to select the region of interest, which ranges from 1800 cm^{-1} to 599 cm^{-1} (it may be different depending on what is being looked for). After the range of interest was selected, we moved on to the detection and elimination of spikes (cosmic ray beams).

2.3.1. Cosmic Rays (Spikes). An external nonoptical source of noise is generated by high-energy particles such as cosmic rays, which reach the equipment detector. Cosmic rays release many electrons that are indistinguishable from photoelectrons. Electrons generated by cosmic rays are concentrated in one or two of the elements of the detector. The

result is a very narrow peak of high intensity in the Raman scattering spectrum. These peaks occur infrequently, in random time and at random positions of the spectrum. It is difficult to confuse a cosmic peak with a Raman band, because of the disparity in its characteristics, unless several cosmic rays fall into adjacent pixels of the detector, which is unlikely [4].

Cosmic rays or spikes are characterized by being intense and narrow peaks, generally three or four points with intensities n times greater than the adjoining points. It is known that hemoglobin (Hb) spectrum shows intense peaks in the $1380\text{--}1330\text{ cm}^{-1}$ and $1630\text{--}1530\text{ cm}^{-1}$ regions. As the sample was measured on a silicon substrate, the peak was found around 960 cm^{-1} in the $990\text{--}930\text{ cm}^{-1}$ region. Knowing where the characteristic bands are is very useful, because if a spike is detected on one of these bands, the whole spectrum is discarded.

Once this is known, a spike can be detected by scanning the spectrum point by point. As each point in the spectrum is scanned, it is checked that the absolute value of the difference between the current point and the next point does not exceed the value of 500; if so, we are in the presence of a spike. The elimination process is described in the following way:

The current point, which is not more than the previous point where the spike begins, is subtracted from the following points when the difference is greater than 500 in order to determine the spectral width of the spike called " n " (it can be between 1 and 4 points). After determining the width of the spike, the current point becomes the point where the spike begins, then the average of the point that is n positions before the current point and n positions after the current point is calculated and replaced by the value that is in the current point. This procedure of replacing spike values is repeated n times.

2.3.2. Shot (Smooth: FFT Filter). Shot noise is the result of the random nature of light. Its intensity is equal to the square root of the number of photons detected. It is an inevitable source of noise in the measurement of Raman spectra [4].

Shot noise can be eliminated from free-spike data. So, we smoothed the spectra implementing the filter method of the fast Fourier transform (FFT) [9]. The algorithm places some limitations on the signal and on the resulting spectrum, for example, the signal from which samples were taken and to be transformed must consist of a number of samples equal to a power of two [10].

The idea that allows this optimization is the transformed decomposition to get to transformations of 2 elements where " k " can take the values 0 and 1. Once the simplest transformations have been resolved, they must be grouped into higher-level transformations, which must be solved again and so on until the highest level is reached. At the end of this process, the results obtained must be rearranged. The algorithm of the fast Fourier transform was popularized by Cooley and Tukey in 1965 [11].

This smoothness consists of six steps:

- (i) Calculation of the mean value of the first 1% of the data and the last 1% of the spectrum data

- (ii) Construction of a straight line through these two points and subtraction of the data from this line
- (iii) Development of the FFT in the set of data acquired in the previous step
- (iv) Application of the filtrate with the low-pass parabolic filter
- (v) Development of the IFFT in the filtered spectrum
- (vi) Addition of a baseline to the set of data purchased in the previous step

FFT:

$$F_n = \sum_{i=0}^{N-1} x_i e^{-(2\pi j/N)(ni)}, \quad (1)$$

where x_i is a sequence of length N .

IFFT:

$$x_i = \frac{1}{N} \sum_{n=0}^{N-1} F_n e^{-(2\pi j/N)(ni)}. \quad (2)$$

Low-pass parabolic filter:

Let f_{c_1} be the pass frequency and f_{c_2} be the stop frequency. The window function can be expressed by $w(f)$:

$$w(f) = \begin{cases} 1, & \text{if } f \leq f_{c_1}, \\ 1 - \frac{(f - f_{c_1})^2}{(f_{c_2} - f_{c_1})^2}, & \text{if } f_{c_1} < f < f_{c_2}, \\ 0, & \text{if } f \geq f_{c_2}. \end{cases} \quad (3)$$

2.3.3. Fluorescence (Spectra Normalization). The noise generated by the sample includes unwanted optical and sample-generated emissions, as is the case of fluorescence, a phenomenon that occurs if a photon hits the molecule, it is absorbed and the molecule goes to an electronic excited state where it remains tenths of nanoseconds, to jump to another excited state but of lower energy, releasing a photon of frequency lower than the incident one. In Raman spectra, fluorescence usually presents as a gentle curvature of the baseline and it can reach an intensity that completely masks the intensity of the Raman bands. The noise generated by the sample also includes Raman intensity changes due to nonconcentration-related changes in the sample. Both the intensity of the bands and the position may vary depending on the temperature of the sample, although these changes tend to be small. The heterogeneity of the sample can also create noise, since the analysis performed at one point in the sample need not be representative of the whole sample [4].

Because the signal from different spectra does not share the same intensity value, we need to reduce funds. To homogenize the spectra, it is first checked that the background is not linear. This is possible to determine by comparing two specific regions that we defined as a lower flat region and upper flat region. Where the lower flat region

was $640\text{--}599\text{ cm}^{-1}$ and the upper flat region ranges from $1800\text{--}1750\text{ cm}^{-1}$.

- (1) The mean of the lower flat region ($640\text{--}599\text{ cm}^{-1}$) is compared with the mean of the upper flat region ($1800\text{--}1750\text{ cm}^{-1}$).
- (i) If the average of the upper flat region ($1800\text{--}1750\text{ cm}^{-1}$) is greater than the subtraction of the averages of the upper and lower flat region plus the average of the lower flat region, then the background is nonlinear.
- (a) The correction consists of normalizing each spectrum by the line equation given by two points $(x_1, y_1), (x_2, y_2)$.

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}. \quad (4)$$

2.4. Multivariate Analysis. After denoising the spectra, we are ready to do the PCA statistical analysis. This multivariate analysis can be done in a program that was also self-made in the software Wolfram Mathematica 9, following the algorithm of the multivariate analysis, PCA [6]. Multivariate analysis is one of the main techniques of data analysis when there are many variables to be considered at the same time for its interpretation.

The first step in multivariate data analysis is to have the data matrix $\mathbf{X}$, from which the vector of mean value $\bar{\mathbf{x}}$ [8] can be computed. From there, we could proceed to obtain the matrix of covariance (of dispersion) and the correlation matrix and from there we can begin to obtain the main components in order to typify the data. This typification process affects the variance of all variables to be the same and equal to 1.

2.4.1. Principal Component Analysis (PCA). PCA is a descriptive technique of dimension reduction, which seeks to study the relationships of interdependence between groups of variables and individual variables. This analysis has its origins in the works of Pearson published in 1901, in the *Philosophical Magazine* with the title "On Lines and Planes of Closest Fit to Systems of Points in Space" [12].

It allows the decomposition of the original data in a model consisting of a part of the signal and a part of noise. The analysis is oriented to model a variance-covariance structure of a data matrix from which the eigenvalues corresponding to the principal components are extracted. Each main component is a linear combination of the original variables. The first major component represents the largest variance and thus corresponds to the higher eigenvalue. The second main component is orthogonal to the first one, with each successive main component being either orthogonal to all of the above and constituting a decreasing proportion of the variance [12].

In summary, in the main component analysis, a sample of size n on p variables $X_1, X_2, \dots, X_p$, initially correlated, is used to obtain a number $k \leq p$ of uncorrelated variables $Z_1,$

$Z_2, \dots, Z_p$ which are a linear combination of the original variables and which explain most of their variability. The main components will be as follows: $Z_1 = \mathbf{a}_1 X_1$, $Z_2 = \mathbf{a}_2 X_2, \dots$, $Z_p = \mathbf{a}_p X_p$, where $\mathbf{a}_p$ are the eigenvectors associated with the covariance matrix. Only the main p components that explain a high percentage of the variability of the original variables are being preserved [12].

Raman spectra are n coordinate vectors defined in what is known as the signal space of normalized wave numbers:

$$E_i = [e_{i1}, \dots, e_{ij}, \dots, e_{in}], \quad (5)$$

where e_{ij} are the Raman intensities of the spectrum E_i for a wave number (Raman shift).

The routine for performing the Raman spectra is performed with Wolfram Mathematica 9 software and is as follows:

- (1) Creation of the data matrix $\mathbf{E}$ with e_{ij} coordinates:

$$\mathbf{E} = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1p} \\ e_{21} & e_{22} & \dots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \dots & e_{np} \end{bmatrix}. \quad (6)$$

- (2) Typification of the variables:

It is necessary to obtain the mean vector from $\mathbf{E}$:

$$\bar{\mathbf{e}} = (\bar{e}_1, \dots, \bar{e}_j, \dots, \bar{e}_p), \quad (7)$$

where

$$\bar{e}_j = \frac{1}{n} \sum_{i=1}^n e_{ij} \quad (8)$$

is the mean of each variable.

Then, it is necessary to obtain the standard deviation from each variable:

$$\sigma_k = \sqrt{\frac{\sum_{i=1}^n (e_{ik} - \bar{e}_k)^2}{n}}, \quad (9)$$

since typification consists of subtracting each variable from its mean value and dividing the result by the standard deviation, thus obtaining new values for the data matrix.

$$x_{ij} = \frac{e_{ij} - \bar{e}_j}{\sigma_i}, \quad (10)$$

where the typified data matrix will be

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}. \quad (11)$$

We calculated the mean and the variance of the data to make sure that the typing was successful. These must be 0 and 1, respectively.

- (3) After typing the data and forming a new matrix of data $\mathbf{X}$ with these, we proceed to obtain the covariance matrix $\mathbf{S}$, which will be given by

$$\mathbf{S} = \frac{1}{n} \mathbf{X}^T \mathbf{X}, \quad (12)$$

where the components of $\mathbf{S}$ are going to be proportional to the correlation coefficients, because the typing process does not affect the correlation between the variables.

So, we can say that the covariance matrix $\mathbf{S}$ is proportional to the correlation matrix $\mathbf{R}$ of the original or typified variables. We are now ready for the main component analysis.

- (4) We proceed to obtain the values and eigenvectors. The eigenvalues are the roots of the equation:

$$|\mathbf{S} - \lambda \mathbf{I}| = 0. \quad (13)$$

The eigenvectors associated with the eigenvalues λ_i are the result of solving the equation:

$$\mathbf{S} \mathbf{a}_i = \lambda_i \mathbf{a}_i. \quad (14)$$

- (5) Once the eigenvectors have been obtained in matrix $\mathbf{a}$, we proceed to extract the main components:

$$\mathbf{Z} = \mathbf{a} \cdot \mathbf{X}. \quad (15)$$

For the preservation or retention of the main components, the most logical criterion is to retain the least number of components that explain a certain percentage of variability, which is generally between 90% and 80% of the total. Another criterion would be to preserve the components whose eigenvalues are greater than or equal to 1 [6]. Finally, it is up to the researcher to decide what the best choice is.

The variability percentages were obtained in the following way:

$$\text{Var} Z_i = \frac{\lambda_i}{n} \times 100. \quad (16)$$

3. Results and Discussion

An intensity difference is visible in Figures 3 and 4 spectra. The Raman signal increment due to the presence of the AgNP favors a better resolution of all peaks, especially those with low intensity in RBC spectra of control cells. This

FIGURE 6: RBC Raman spectra with AgNP, with selected interest region 1800–599 cm^{-1} (spikes can be observed).

increments the signal/noise relationship favor the use of this technique in the obtainment of bioimages.

3.1. Denoising Raman Spectra. Then, we analyze the RBC Raman spectra incubated with and without AgNP in order to remove the specified noises, as it is described in Section 2.3. In Figures 6–9, we can observe the denoising process gradually. For this process, we take into consideration the order to remove noises, because if we remove shot noise before removing spikes, these get masked as Raman bands [13]. Figure 6 shows the scientists desired region between 1800–599 cm^{-1} , which is full of noises; we can clearly see the huge spikes, the shot noise, and the fluorescence contribution.

Figure 7 shows the Raman spectra after removing spikes. Those spikes were removed by implementing a criteria sequence that allowed us to identify and remove them, best described in Section 2.3.1. In Figure 8, we can observe spectra after smoothing. The smooth is done applying the FFT filter, which is an efficient algorithm that allows to calculate the discrete Fourier transform (DFT) and its inverse. It is of great importance in a wide variety of applications, from digital signal processing and digital filtering, among others.

In Figure 9, we see the spectra in the denoising process final state, where the fluorescence has been removed by creating a grade one polynomial function, a straight line, from two specific regions. We selected the lowest region from the lower section and the lowest region from the upper section for each spectrum. At the end, a spectral correction was made using the line equation given by two points (x_1, y_1) , (x_2, y_2) . With these two points, we construct a line, which we subtract from each spectrum for its normalization (4). Then, the spectra can be processed using multivariate analysis.

3.2. Statistical Analysis (PCA). After denoising all RBC Raman spectra with and without AgNP, we moved on to perform the PCA. From the equipment (LabRAM HR, Horiba), we obtained a data matrix consisting of n spectra with 1201

FIGURE 7: RBC Raman spectra with AgNP after removing the spikes.

FIGURE 8: RBC Raman spectra with AgNP after softening.

FIGURE 9: RBC Raman spectra with AgNP normalized (without fluorescence).

TABLE 1: Principal component analysis results from RBC Raman spectra with and without AgNP.

Mean	Standard deviation	Eigenvalues	PC1 coefficient	CP2 coefficient	Variability (%)	Cumulative	PC 1	PC 2
RBC Raman spectra without AgNP								
435.95	313.05875	79.03374	0.11001	−0.04796	83.19%	83.19%	−7.50495	−3.14694
481.68	358.32162	14.88045	0.10944	0.03865	15.66%	98.86%	−7.56427	−3.15832
417.55	331.27881	0.25699	0.11141	−0.00945	0.27%	99.13%	−7.35318	−3.22056
473.72	332.63329	0.10526	0.11058	0.02211	0.11%	99.24%	−7.00289	−3.28952
490.56	341.39515	0.07569	0.11131	0.01602	0.08%	99.32%	−6.71021	−3.31462
500.20	339.27534	0.0405	0.1097	0.04543	0.04%	99.36%	−6.58359	−3.31285
819.22	605.57175	0.02471	0.08893	0.15731	0.03%	99.39%	−6.62827	−3.30024
421.03	297.20846	0.02142	0.10817	−0.06154	0.02%	99.41%	−6.82522	−3.26039
371.24	294.55975	0.01972	0.10645	−0.08063	0.02%	99.43%	−7.16513	−3.17401
356.21	301.66804	0.01873	0.10604	−0.08307	0.02%	99.45%	−7.60888	−3.05894
RBC Raman spectra with AgNP								
838.76	466.44873	77.80994	0.10796	0.0767	86.46%	86.46%	−9.0323	1.7079
589.22	390.11247	11.23373	0.09223	0.1709	12.48%	98.94%	−9.0674	1.74138
1237.61	1002.19047	0.29913	0.10315	−0.102	0.33%	99.27%	−9.0429	1.78999
727.51	460.3899	0.1151	0.1116	0.0472	0.13%	99.40%	−8.9426	1.85733
1116.57	756.02231	0.08343	0.10916	−0.0762	0.09%	99.49%	−8.8108	1.92021
801.73	519.32077	0.0572	0.11248	0.0167	0.06%	99.55%	−8.7346	1.93386
1486.09	1157.09662	0.04709	0.10314	−0.1222	0.05%	99.61%	−8.7759	1.87638
1171.28	850.75433	0.02201	0.10681	−0.0959	0.02%	99.63%	−8.9351	1.77626
740.12	474.83072	0.01593	0.11229	0.0261	0.02%	99.65%	−9.1801	1.67912
700.17	435.21653	0.01309	0.11164	0.0424	0.01%	99.66%	−9.4829	1.59353

values each. These factors express the relationships between the wave number variables ($1800\text{--}599\text{ cm}^{-1}$). So, the spectra had $n = 615$ values. All the data was typified, and a complete PCA cross validation was performed using one or two principal components.

After applying PCA over the denoised Raman spectra, we obtained statistical results shown in Table 1, where the PC variability percent shows that the first two PCs explain 98.86% of the information from RBC Raman spectra without AgNP and 98.94% of the information from RBC Raman spectra with AgNP. PC selection was made by applying component retention criteria; as we explained in Section 2.4.1, we took the fewest components that explain a percentage of variability between 80–90% of the total to preserve components whose eigenvalues are greater than or equal to 1.

In this case, we obtain two strong principal components, but we can see in Figures 10 and 11, where we show the two PC spectra from each dataset, that only the first PC in both cases gives readable information.

By making a correlation between the original variables and the retained components, we can also discard the second component, as shown in Table 2.

In Figure 12, we have the PC retained from both datasets, from which we can say that the characteristic peaks are distinguished and the main components in it match with the ones reported in [14], but they need more treatment to, for example, remove the silicon band and see the intensities' difference better between the spectra.

FIGURE 10: Two principal component spectra with the highest variability percentage from RBC Raman spectra without AgNP.

4. Conclusions

In this work, RBC Raman spectra noises from two representative independent tests were eliminated, one incubated without AgNP and the second one with AgNP. During the Raman

FIGURE 11: Two principal component spectra with the highest variability percentage from RBC Raman spectra with AgNP.

TABLE 2: Correlation between original variables and components retained from RBC Raman spectra with and without AgNP.

Spectra	RBC Raman spectra without AgNP		RBC Raman spectra with AgNP	
	Principal component 1	Principal component 2	Principal component 1	Principal component 2
1	-0.978001	-0.424367	-0.94774	-0.360109
2	-0.972961	-0.42218	-0.81615	-0.310112
3	-0.990444	-0.429766	-0.90712	-0.344673
4	-0.983093	-0.426576	-0.98619	-0.374718
5	-0.989591	-0.429396	-0.9614	-0.365299
6	-0.975209	-0.423155	-0.99189	-0.376884
7	-0.790572	-0.343039	-0.910368	-0.345909
8	-0.961614	-0.417256	-0.945826	-0.359381
9	-0.946343	-0.41063	-0.989432	-0.37595
10	-0.942681	-0.409041	-0.982628	-0.373365
11	-0.982285	-0.426225	-0.921893	-0.350288
12	-0.893712	-0.387793	-0.792295	-0.301045
13	-0.720864	-0.312792	-0.859795	-0.326693
14	-0.902113	-0.391438	-0.947023	-0.359836

signal process, cosmic rays (spikes) were eliminated by detection and elimination criteria (mathematical method); the data smoothing was performed using the FFT filter method and fluorescence elimination by normalizing the data by the line equation given two points. More visible and manageable results were obtained for a better handling in the statistical analysis, the PCA, being able to represent the given referenced system, by a more affordable and smaller dimension, in such a way that information loss was minimal. This reduced the data analysis time, which

FIGURE 12: PC spectra from dataset analyzed.

was done by hand and was very cumbersome. In fact, the spectra denoising lasted for a month, before the statistical analysis was carried out.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

An earlier version of this work was presented at the 1st International Congress of Engineering and Technology Science, 2017.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by CONACYT (Mexico) under Contract 564940/297799.

References

- [1] A. Campion and P. Kambhampati, "Surface-enhanced Raman scattering," *Chemical Society Reviews*, vol. 27, no. 4, pp. 241–250, 1998.
- [2] V. I. Kukushkin, A. B. Van'kov, and I. V. Kukushkin, "Relationship between the giant enhancement of the Raman scattering and luminescence on nanostructured metallic surfaces," *JETP Letters*, vol. 98, no. 6, pp. 342–347, 2013.
- [3] J. R. Lombardi and R. L. Birke, "A unified approach to surface-enhanced Raman spectroscopy," *Journal of Physical Chemistry C*, vol. 112, no. 14, pp. 5605–5617, 2008.

- [4] O. Sala, “Capítulo 1: fundamentos de espectroscopía Raman,” in *Fundamentos Da Espectroscopia Raman e No Infravermelho*, pp. 10–34, UNESP, Sao Paulo, Brazil, 1996.
- [5] N. Fleitas-Salazar, E. Silva-Campa, S. Pedroso-Santana, J. Tanori, M. R. Pedroza-Montero, and R. Riera, “Effect of temperature on the synthesis of silver nanoparticles with polyethylene glycol: new insights into the reduction mechanism,” *Journal of Nanoparticle Research*, vol. 19, no. 3, 2017.
- [6] C. Pérez, *Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS*, Pearson Educación S.A., Madrid, España, 2004.
- [7] J. M. Lerner and A. Thevenon, *Optics of Spectroscopy Tutorial*, HORIBA Scientific, Edison, NJ, USA, 2013, <http://www.horiba.com/us/en/scientific/products/optics-tutorial/>.
- [8] C. M. Cuadras, *Nuevos Métodos De Análisis Multivariante*, CMC Editions, Barcelona, Spain, 2014.
- [9] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C (2nd ed.): The Art of Scientific Computing*, Cambridge University Press, New York, NY, USA, 1992.
- [10] P. Kraniuskauskas, “A plain man’s guide to the FFT,” *IEEE Signal Processing Magazine*, vol. 11, no. 2, pp. 24–35, 1994.
- [11] J. W. Cooley and J. W. Tukey, “An algorithm for the machine calculation of complex Fourier series,” *Mathematics of Computation*, vol. 19, no. 90, p. 297, 1965.
- [12] K. Pearson, “Principal components analysis,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 6, no. 2, p. 559, 1901.
- [13] J. M. Yúfera, M. J. Soneira, S. Ruiz-Moreno, and M. J. Sedó, “Técnicas de procesamiento de señal aplicadas al tratamiento de espectros Raman,” *Buran*, vol. 8, pp. 45–48, 1996.
- [14] B. R. Wood and D. McNaughton, “Raman excitation wavelength investigation of single red blood cells in vivo,” *Journal of Raman Spectroscopy*, vol. 33, no. 7, pp. 517–523, 2002.