

# **UNIVERSIDAD DE SONORA**

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES  
DEPARTAMENTO DE FÍSICA

POSGRADO EN NANOTECNOLOGÍA

**Análisis Multivariable de la espectroscopía Raman  
amplificada por adición de nanopartículas en  
glóbulos rojos**

**TESIS**

**Que para obtener el título de:**

**MAESTRÍA EN NANOTECNOLOGÍA**

**PRESENTA:**

**LISSET FERRER GALINDO**

**DIRECTOR DE TESIS:**

**DR. RAÚL RIERA AROCHE**

**CODIRECTOR DE TESIS:**

**DR. MARTÍN R. PEDROZA MONTERO**

# Universidad de Sonora

Repositorio Institucional UNISON



**"El saber de mis hijos  
hará mi grandeza"**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

*A mis padres Maura y Lázaro, que gracias a su apoyo y amor incondicional me han  
convertido en la persona que soy.*

*A mi Alejandro, que ha estado desde hace muchos años a mi lado y siempre me ha  
apoyado en los momentos buenos y malos.*

## AGRADECIMIENTOS

A mis hermosos padres, quienes me inculcaron todos los valores para ser una persona honrada, que siempre me han apoyado y me brindaron seguridad, confianza. A ellos gracias por ser mi inspiración de superación y amor.

A mi Alejandro, por su gran ayuda durante todo el tiempo que llevamos juntos. Por ser aquel que me animaba a seguir en los tiempos de desconsuelo, por sacar lo mejor de mí, por su atención e interés prestados en los días de arduo trabajo, por mimarme, cuidarme. Pero sobre todo por amarme.

A toda mi familia por creer en mí y apoyarme siempre, a los Ferrer, a los Galindo, y a los Moreno.

A mis tutores el Dr. Raúl Riera Aroche y el Dr. Martín R. Pedroza Montero, por ser oportunos y brindarme su sabiduría cuando la necesitaba.

A mis compañeros de la maestría a los cuales puedo llamar amigos, con quienes discutí muchos aspectos del trabajo y tributaron ideas concretas para su desarrollo, muchas gracias.

Doy gracias por las vivencias y la experiencia adquirida como alumna del Programa de Posgrado en Nanotecnología, del Departamento de Física de la Universidad de Sonora, en Hermosillo, en donde crecí intelectual y espiritualmente e hice nuevos amigos.

Además de agradecer al Consejo Nacional de Ciencia y Tecnología (CONACYT) por confiar en mí y permitir mi ingreso en este programa y brindarme su apoyo, sin el cual no hubiese logrado mis metas.

# Índice

<b>RESUMEN.....</b>	<b>5</b>
<b>ABSTRACT.....</b>	<b>6</b>
<b>1. INTRODUCCIÓN.....</b>	<b>7</b>
1.1. Antecedentes.....	7
1.2. Objetivo general.....	8
1.3. Objetivos específicos.....	8
1.4. Hipótesis.....	9
1.5. Justificación.....	9
<b>2. DISPERSIÓN RAMAN DE SUPERFICIE MEJORADA (SERS).....</b>	<b>10</b>
2.1. Mecanismos de generación de SERS.....	11
2.1.1. Teoría Electromagnética.....	11
2.1.2. Teoría Química.....	13
2.2. Nanopartículas o Nanosuperficies.....	13
<b>3. EXPERIMENTO.....</b>	<b>15</b>
3.1. Espectros Raman de glóbulos rojos (RBC).....	15
3.2. Espectros Raman con nanopartículas de plata Ag (RBC+NPAg).....	16
<b>4. ANÁLISIS.....</b>	<b>18</b>
4.1. Procesamiento de la señal.....	20
4.1.1. Detección y eliminación de picos de rayos cósmicos.....	21
4.1.2. Suavizado (Smooth: FFT Filter).....	23
4.1.3. Normalización de espectros.....	25
4.2. Análisis estadístico.....	26
4.2.1. Análisis Multivariable (AMV).....	27
4.2.2. Análisis de Componentes Principales (ACP).....	29
4.2.2.1. Aplicación del ACP al tratamiento de los datos.....	31
4.2.2.2. ACP de espectros Raman de RBC y RBC+NPAg.....	38
<b>5. CONCLUSIONES.....</b>	<b>44</b>
<b>6. TRABAJO FUTURO.....</b>	<b>45</b>
<b>7. REFERENCIAS.....</b>	<b>46</b>

# RESUMEN

Con el uso de la espectroscopía Raman para el propósito biológico debido a su alta sensibilidad a los cambios moleculares sutiles, ha aumentado la necesidad de interpretar grandes conjuntos de datos resultantes de aplicar esta técnica espectroscópica. Al ser las señales Raman inherentemente débiles, existe una técnica que ofrece órdenes de incrementos de magnitud en la intensidad de estas señales, conocida como Dispersión Raman de Superficie Mejorada (SERS siglas en inglés).

Como los espectros resultantes del Raman y del SERS son engorrosos debido a los ruidos que presentan, es necesaria la aplicación de criterios para la detección y eliminación de estos, para una mejor interpretación de los datos. Eliminandose ruidos como el generado por fuentes externas (rayos cósmicos), el Shot (resultado de la naturaleza aleatoria de la luz), y el generado por la muestra (fluorescencia), se tienen mejores datos para pasar a un análisis multivariable, en específico el análisis de componentes principales, para disminuir las dimensiones del conjunto de datos tomando solo las componentes claves necesarias para una mejor interpretación de estos espectros.

Para la eliminación de los picos de rayos cósmicos y la fluorescencia se trabajó con un software desarrollado en el lenguaje Java en el software NetBeans IDE 7.0.1, mientras que para eliminar el ruido shot se utilizó el software OriginPro 9. Y el Análisis Multivariable se realizó en el Wolfram Mathematica 9, comparándose los resultados con los obtenidos en el OriginPro 9.

Este trabajo presenta el tratamiento de ruidos y la aplicación de un análisis multivariable en espectros Raman de glóbulos rojos amplificados por adición de nanopartículas, para una mejor interpretación del inmenso conjunto de datos obtenidos inicialmente de un experimento donde se vieron los espectros Raman de eritrocitos incubados con nanopartículas de plata y sin nanopartículas. Obteniéndose espectros sin ruidos y la reducción de estos, permitiendo a los investigadores obtener inferencias y predicciones de forma más rápida.

## ABSTRACT

With the use of Raman spectroscopy for biological purpose due to its high sensitivity to subtle molecular changes, it has increased the need to interpret large data sets resulting from applying this spectroscopic technique. As the inherently weak Raman signals, there is a technique that provides orders of magnitude increase in the intensity of these signals, known as Surface Enhanced Raman Scattering (SERS).

As the Raman and SERS spectra are cumbersome because of the noise present, the application of criteria for detection and removal of these, for better interpretation of the data is necessary. Eliminating noises like the generated by external sources (cosmic rays), Shot (result of the random nature of light), and the generated by the sample (fluorescence), we have better data to move to a multivariate analysis, specifically the principal component analysis, to reduce the size of the data set taking only the key components necessary for a better interpretation of these spectra.

For removing cosmic ray peaks and fluorescence we worked with a software developed in Java in NetBeans IDE 7.0.1, while to remove the shot noise we used OriginPro 9 and for the statistical analysis we used Wolfram Mathematica comparing the results with those obtained in the OriginPro 9.

This paper presents the treatment of noise and application of multivariate analysis in Raman spectra of red blood cells amplified by the addition of nanoparticles, for a better understanding of the enormous data set originally obtained from an experiment in which Raman spectra of erythrocytes were incubated with silver nanoparticles and without nanoparticles. Here we obtain spectra without noise and reducing these, allowing researchers to make inferences and predictions faster.

# 1. INTRODUCCIÓN

## 1.1. Antecedentes

En las últimas décadas, el uso de la espectroscopía Raman para el propósito biológico ha aumentado a causa de sus grandes beneficios, tales como su alta sensibilidad a los cambios moleculares sutiles. Además, la técnica Raman ofrece anchos de banda estrechos en general, la preparación mínima de la muestra y es fácilmente interconectada con fibra óptica para el análisis a distancia. La resolución espacial de la espectroscopía Raman en la escala del micrómetro y su capacidad para sondar muestras en condiciones in vitro, permitirá nuevos conocimientos sobre las células individuales.

Debido a que las señales Raman tienen la característica de ser débiles, se recurre a una técnica que ofrece órdenes de incrementos de magnitud en la intensidad de las señales Raman, conocida como Dispersión Raman de Superficie Mejorada (Surface Enhanced Raman Scattering SERS en inglés). Factores de mejora pueden ser tan altos como  $10^{14-15}$ , que son suficientes para permitir, incluso, detección de moléculas individuales utilizando Raman. SERS es de interés para el análisis de trazas, citometría de flujo y otras aplicaciones donde la sensibilidad de la velocidad de una medición Raman es insuficiente [1.1].

Como los espectros resultantes del Raman y el SERS son engorrosos por los ruidos que presentan, se han desarrollado criterios de detección y eliminación para ruidos como son, el ruido shot (resultado de la naturaleza aleatoria de la luz), el generado por la muestra (incluye emisiones ópticas no deseadas y generadas por la propia muestra, como es el caso de la fluorescencia), el generado por fuentes externas (picos de rayos cósmicos o spikes), entre otros.

Dependiendo de la finalidad del análisis de datos y el conocimiento previo de las muestras en estudio, los métodos quimiométricos son generalmente clasificados como: no supervisado y supervisado.

Los métodos no supervisados, como el análisis de componentes principales (PCA), el análisis factorial (FA), y el análisis de conglomerados (CA), como su nombre indica, no requieren de ningún conocimiento previo de las muestras en estudio, y pueden proporcionar patrones, agrupaciones, detección de valores atípicos. Por otro lado, los métodos supervisados, tales como regresión lineal múltiple (MLR), regresión de componentes principales (PCR), la regresión de mínimos cuadrados parciales (PLS) y el análisis discriminante lineal (LDA), entre otros, requieren un conjunto de muestras bien caracterizadas.

En el Análisis Univariante o Univariable, como su nombre lo dice, sólo se estudia una variable; mientras que en el Análisis Multivariable se estudian varias variables al mismo tiempo, además de que algoritmos múltiples están disponibles. El propósito de este Análisis Multivariable es medir, explicar y predecir el grado de relación que existe entre la variación (combinación lineal ponderada) de las variables.

## **1.2. Objetivo general**

Aplicar análisis multivariable, en particular el análisis de componentes principales, a espectros Raman amplificadas por adición de nanopartículas en glóbulos rojos luego de un tratamiento para eliminar algunos ruidos de estos espectros.

## **1.3. Objetivos específicos**

- ❖ Eliminar ruidos de los espectros Raman.
- ❖ Aplicar Análisis Multivariable a los datos sin ruidos, en específico el Análisis de Componentes Principales.

## **1.4. Hipótesis**

Con la eliminación de ruidos y la aplicación del análisis multivariable a espectros Raman amplificados por adición de nanopartículas (reportadores SERS) en glóbulos rojos, es posible obtener inferencias y predicciones sobre los datos resultantes, que ayudarán en el diagnóstico y tratamiento de enfermedades.

## **1.5. Justificación**

El uso de técnicas de análisis multivariable nos permite el desarrollo del procesamiento e interpretación de datos en múltiples espectros Raman amplificados por adición de nanopartículas (reportadores SERS) en glóbulos rojos, asociados a un punto o célula. Para que dicho análisis pueda ser llevado a cabo, se necesita que los datos (que se sabe que presentan ruidos) no tengan ruidos. Por lo que con la eliminación previa de los ruidos se obtienen mejores datos para el análisis.

## 2. DISPERSIÓN RAMAN DE SUPERFICIE MEJORADA (SERS)

La Dispersión Raman de Superficie Mejorada o Surface Enhanced Raman Scattering (SERS) es una técnica de superficie sensible, que mejora la dispersión Raman por moléculas adsorbidas sobre superficies metálicas rugosas. El factor de mejora puede ser tanto como  $10^{10}$  a  $10^{11}$ , lo que significa que la técnica puede detectar moléculas individuales.

SERS de piridina adsorbida en plata electroquímicamente rugosa, fue observada en un principio por los investigadores Martin Fleischman, Patric J. Hendra y A. James McQuillan en el Departamento de Química de la Universidad de Southampton, en Southampton, Reino Unido en 1973. En 1977, dos grupos de científicos independientemente observaron que la concentración de especies de dispersión no podría ser responsable de la señal mejorada, por lo que cada uno propone un mecanismo para la mejora observada.

Jean Maire y Van Duyne propusieron un efecto electromagnético, mientras que Albrecht y Creighton propusieron un efecto de transferencia de carga. Rufus Ritchie, de la División de Investigación de Ciencias de la Salud del Laboratorio Nacional de Oak Ridge, predijo la existencia del plasmón superficial [2.1].

SERS es un fenómeno asociado con el aumento del campo electromagnético que rodea pequeños objetos ópticamente excitados cerca de una intensa resonancia de plasmón. Consiste en utilizar los aumentos de campo localizado que pueden existir en superficies metálicas.

En la práctica, las ventajas de la SERS se pueden explorar en cualquier sistema Raman, y la medición real se realiza en la forma estándar. Normalmente es necesario utilizar una longitud de onda de láser que es compatible con el metal SERS elegido, pero más allá de esto no hay grandes dificultades.

## 2.1. Mecanismos de generación de SERS

El mecanismo exacto del efecto de mejora de SERS es todavía un tema de debate en la literatura. Existen dos teorías principales y si bien sus mecanismos difieren sustancialmente, distinguirlos experimentalmente no ha sido sencillo. La teoría electromagnética propone la excitación de plasmones de superficie, mientras que la teoría química propone la formación de complejos de transferencia de carga.

La teoría química se aplica sólo para las especies que han formado un enlace químico con la superficie, por lo que no puede explicar el aumento de la señal observada en todos los casos; mientras que la teoría electromagnética puede aplicarse incluso en aquellos casos en que la muestra es fisisorbida (cuando un elemento o compuesto químico se adhiere a una superficie, que puede estar formada por el mismo tipo de compuesto o por otro diferente, en el que la especie adsorbida conserva su naturaleza química). Se ha demostrado recientemente que la mejora SERS puede ocurrir incluso cuando una molécula excitada está relativamente lejos de la superficie que alberga nanopartículas metálicas, permitiendo fenómenos de plasmones de superficie [2.2]. Esta observación proporciona un fuerte apoyo a la teoría electromagnética de SERS.

### 2.1.1. Teoría Electromagnética

El aumento en la intensidad de la señal Raman para adsorbatos en superficies particulares se produce debido a una mejora en el campo eléctrico proporcionado por la superficie. Cuando la luz incidente en el experimento golpea la superficie, plasmones de superficie localizados son excitados. La mejora de campo es mayor cuando la frecuencia de plasmón  $\omega_p$ , está en resonancia con la radiación.

Para que ocurra la dispersión, las oscilaciones del plasmón deben ser perpendiculares a la superficie; si están en el plano con la superficie, no se producirá ninguna dispersión. Es debido a este requisito que las superficies rugosas o arreglos de las nanopartículas se emplean típicamente en los experimentos SERS ya que estas superficies proporcionan un área en la que pueden ocurrir estas oscilaciones colectivas localizadas [2.3].

La luz incidente sobre la superficie puede excitar a una variedad de fenómenos en la superficie, sin embargo, la complejidad de esta situación puede ser minimizada por las superficies con características mucho más pequeñas que la longitud de onda de la luz, ya que sólo la contribución dipolar será reconocida por el sistema. El término dipolar contribuye a las oscilaciones de plasmones, lo que conduce a la mejora. El efecto SERS es tan pronunciado debido a que la mejora de campo se produce dos veces.

En primer lugar, la mejora de campo aumenta la intensidad de la luz incidente, la cual excitará los modos Raman de la molécula en estudio, por lo tanto, el aumento de la señal de la dispersión Raman. La señal Raman está entonces reforzada por la superficie debido a el mismo mecanismo que excita la luz incidente, lo que resulta en un mayor aumento en la producción total.

La mejora no es igual para todas las frecuencias. Para aquellas frecuencias para las que la señal Raman es sólo ligeramente modificada de la luz incidente, tanto la luz láser incidente y la señal Raman pueden estar cerca de resonancia con la frecuencia de plasmones. Cuando el desplazamiento de frecuencia es grande, la luz incidente y la señal Raman no pueden estar ambos en resonancia con  $\omega_p$ , así la mejora en ambas etapas no puede ser máxima [2.4].

La elección de la superficie de metal también está dictada por la frecuencia de resonancia de plasmón. La radiación visible y el infrarrojo cercano (NIR) se utilizan para excitar modos Raman. La plata y el oro son metales típicos de experimentos SERS, debido a que sus frecuencias de resonancia de plasmones caen dentro de estos rangos de longitud de onda, proporcionando mejora máxima de la luz visible y NIR.

### **2.1.2. Teoría Química**

Mientras que la teoría electromagnética de mejora se puede aplicar independientemente de la molécula que se está estudiando, no explica completamente la magnitud de la mejora observada en muchos sistemas. Para muchas moléculas, a menudo las que tienen un par solitario de electrones, en el que las moléculas pueden unirse a la superficie, un mecanismo de mejora diferente que no implica plasmones de superficie ha sido descrito. Este mecanismo químico implica la transferencia de carga entre las especies quimisorbidas y la superficie de metal. El mecanismo químico sólo se aplica en casos específicos y probablemente se produce en conjunto con el mecanismo electromagnético [2.5].

## **2.2. Nanopartículas o Nanosuperficies**

Mientras que se puede realizar SERS en soluciones coloidales, hoy en día el método más común para realizar mediciones SERS es mediante el depósito de una muestra de líquido sobre una superficie de vidrio o de silicio con una superficie de metal noble nanoestructurado. Mientras que los primeros experimentos se realizaron en plata electroquímicamente rugosa, ahora las superficies se preparan a menudo usando una distribución de las nanopartículas de metal en la superficie [2.6], así como el uso de la litografía [2.7] o silicio poroso como soporte [2.8].

La aplicación de una película delgada de plata sobre obleas de silicio, a través de una inmersión de un día en una solución saturada de nitrato de plata en *n*-octanol, es un método ampliamente aceptado para preparar un sustrato de dispersión Raman de superficie mayor [2.9]. Los metales más comunes utilizados para superficies plasmónicas son plata y oro, sin embargo aluminio ha sido explorado recientemente como un material plasmónico alternativo, debido a que su banda de plasmón está en la región UV, al contrario de plata y oro. Por lo tanto, hay un gran interés para el uso de aluminio para la SERS UV.

En la presente década, se ha reconocido que el costo de sustratos SERS debe reducirse con el fin de convertirse en una técnica de medición química analítica utilizada. Para satisfacer esta necesidad, el papel plasmónico ha experimentado una gran atención en el campo, con sustratos SERS altamente sensibles que se forman a través de enfoques tales como remojo [2.10], síntesis *in situ* [2.11], serigrafía [2.12] y la impresión de inyección de tinta [2.13].

La forma y el tamaño de las nanopartículas de metal afectan fuertemente la intensidad de la mejora debido a que estos factores influyen en la relación de eventos de absorción y dispersión. Hay un tamaño ideal para estas partículas, y un espesor de superficie ideal para cada experimento [2.14]. Las partículas que son demasiado grandes permiten la excitación de multipolos. Como solamente la transición de dipolo conduce a la dispersión Raman, las transiciones de orden superior provocarán una disminución en la eficiencia global de la mejora.

Las partículas que son demasiado pequeñas pierden su conductancia eléctrica y no pueden mejorar el campo. Cuando el tamaño de partícula se aproxima a unos pocos átomos, la definición de un plasmón no se sostiene, ya que debe haber una gran colección de electrones a oscilar juntos [2.15]. Un sustrato SERS ideal debe poseer alta uniformidad y alto aumento de campo.

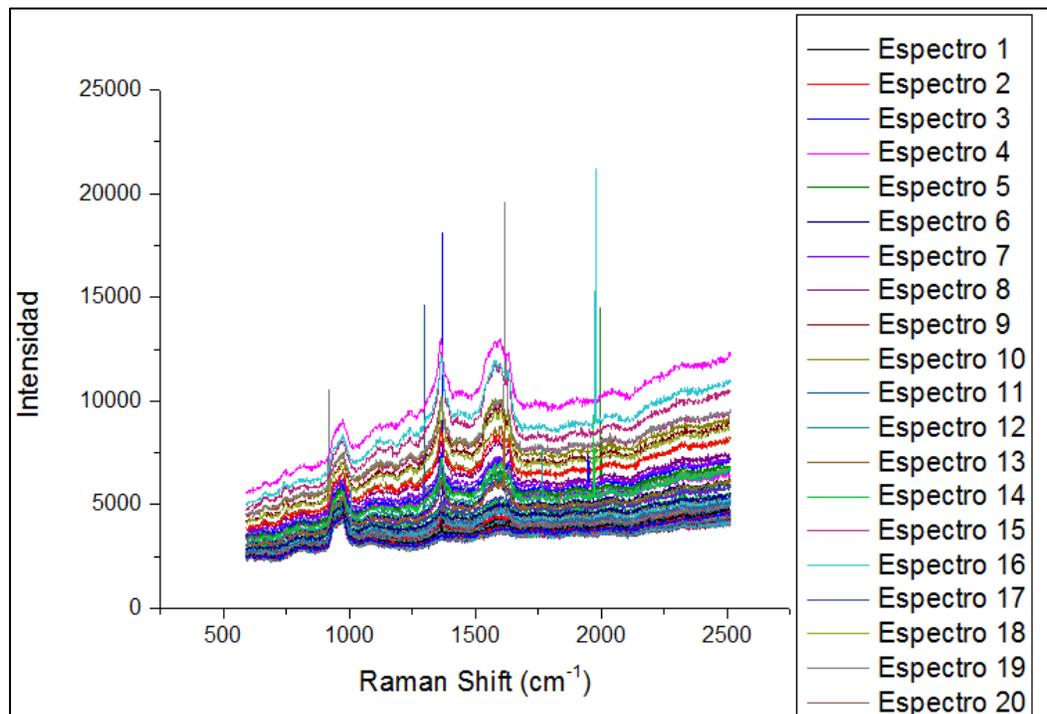
### 3. EXPERIMENTO

En este trabajo se eliminaron determinados ruidos y se aplicó un análisis multivariable a los espectros de un estudio, donde se vieron los espectros Raman de eritrocitos incubados con nanopartículas de plata y sin nanopartículas. Se emplearon células frescas obtenidas por extracción de sangre a un voluntario sano [3.1].

#### 3.1. Espectros Raman de glóbulos rojos (RBC)

Como control se emplearon células sin nanopartículas:

- Control de células: Una alícuota de 10  $\mu\text{L}$  de sangre se diluyó 1:100 (V:V) en solución BPS (siglas del inglés buffer phosphate saline) [3.1].

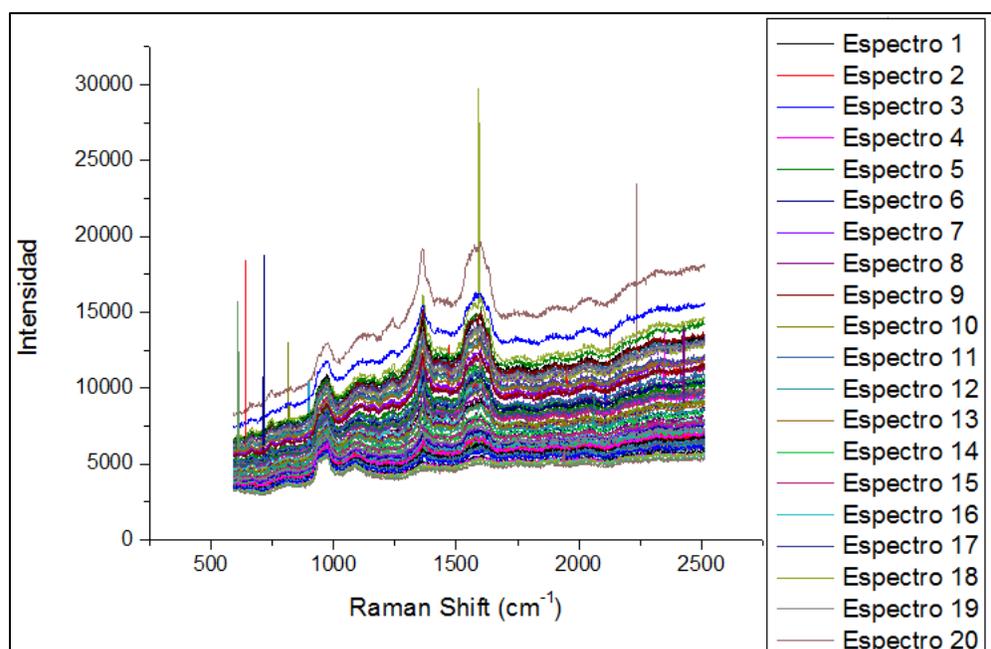


**Figura 3.1.** Espectros Raman de glóbulos rojos sin nanopartículas.

En la figura 3.1 se observan 95 espectros de glóbulos rojos sin nanopartículas, los cuales fueron tomados como Control en el experimento.

## 3.2. Espectros Raman con nanopartículas de plata Ag (RBC+NPAg)

- Células con NPAg: Una pequeña alícuota de sangre se diluyó 1:100 (V:V) en solución BPS. Se añadió la solución de NPAgs en una dilución final 1:10 [3.1].



**Figura 3.2.** Espectros Raman de glóbulos rojos con nanopartículas de plata (RBC+NPAg).

La figura 3.2 muestra 90 espectros de glóbulos rojos con NPAg, los cuales serán analizados y comparados con los espectros de Control.

Ambas muestras se incubaron durante 2 h a  $25\pm 4^{\circ}\text{C}$ . Luego, se prepararon las muestras de células para la medición del espectro Raman. Una gota de la suspensión de células se añadió a un sustrato de silicio previamente recubierto con poli-L-Lisina [3.1].

Las células en el sustrato se dejaron en reposo 15 minutos para propiciar la inmovilización de los glóbulos rojos para la medición [3.1].

Las mediciones se realizaron con un micro-espectrómetro Raman (LabRam HR Horiba), que se muestra en la figura 3.3, equipado con una cámara CCD refrigerada. Empleamos el objetivo 100X, en condiciones "no confocal" (apertura de 1000  $\mu\text{m}$ ) y el láser de 488 nm. Cada medición se realizó con 1 sola adquisición de 20 segundos.

En la figura 3.3 podemos ver el equipo con el que fueron realizadas dichas mediciones, el cual se encuentra en el edificio 3R del Departamento de Física de la Universidad de Sonora, Hermosillo, México.



**Figura 3.3.** Micro-espectrómetro Raman (LabRam HR Horiba) utilizado.



**Figura 3.4.** Captura de pantalla de medición de muestras en el LabRam HR Horiba.

---

## 4. ANÁLISIS

Los espectros Raman obtenidos del experimento, no son más que matrices de datos las cuales van a estar contenidas en archivos de extensión “.txt”. Los archivos pueden contener mediciones de una o más células, según la preferencia del investigador.

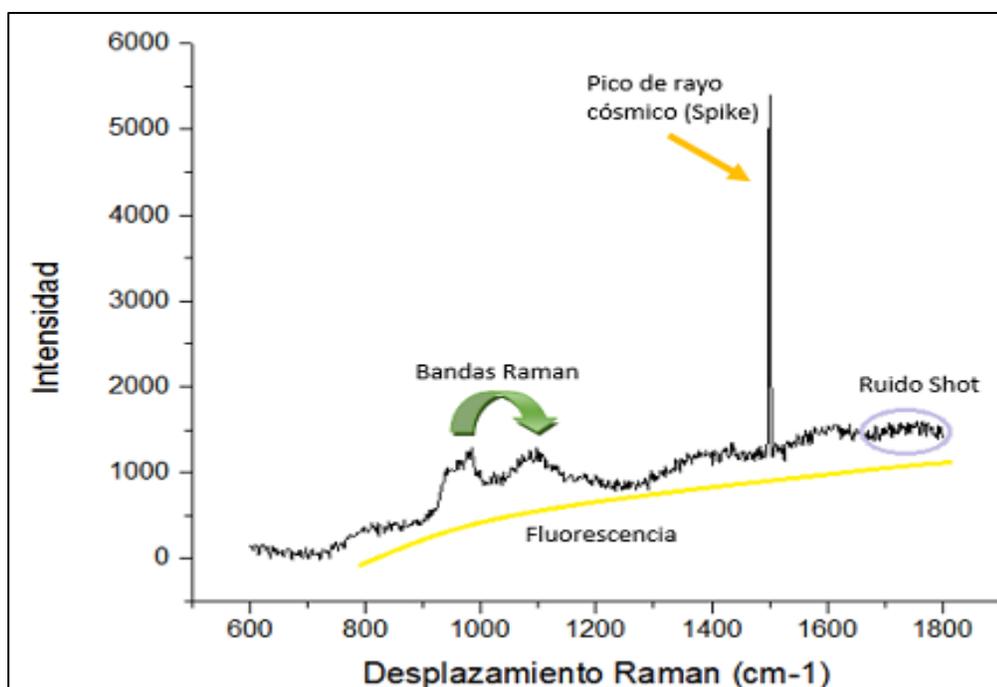
Aparte de que las señales Raman son inherentemente débiles y podemos amplificarlas aplicando SERS, los espectros como hemos visto, presentan ruidos que muchas veces no podemos eliminar durante la medición en el equipo. El ruido se considera como todo lo que no corresponde a la dispersión Raman, es decir, que no depende de la estructura molecular de la muestra. En otras palabras, el ruido básicamente se compone de fluctuaciones aleatorias en la intensidad.

Los ruidos más comunes se clasifican en cinco grupos diferentes [4.1]:

- ❖ **Ruido Shot:** es el resultado de la naturaleza aleatoria de la luz. Su intensidad es igual a la raíz cuadrada del número de fotones detectados. Es una fuente de ruido inevitable en la medida de espectros Raman.
- ❖ **Ruido generado por la muestra:** incluye emisiones ópticas no deseadas y generadas por la propia muestra, como es el caso de la **fluorescencia**, fenómeno que se produce si al incidir un fotón sobre la molécula, este es absorbido y la molécula pasa a un estado electrónico excitado donde permanece unas decenas de nanosegundos, para saltar a otro estado excitado pero de menor energía, liberando un fotón de frecuencia más baja que el incidente. En los espectros Raman, la fluorescencia suele presentarse como una suave curvatura de la línea de base y puede alcanzar una intensidad que llegue a enmascarar por completo la intensidad de las bandas Raman. El ruido generado por la muestra incluye también los cambios de intensidad Raman debidos a cambios en la muestra no relacionados con la concentración; tanto la intensidad de las bandas como la posición pueden variar en función de la temperatura de la muestra,

aunque estos cambios tienden a ser pequeños. La heterogeneidad de la muestra también puede crear ruido, ya que el análisis realizado en un punto de la muestra no tiene por qué ser representativo de la muestra completa.

- ❖ **Ruido generado por la instrumentación:** depende del diseño específico de la instrumentación empleada en el análisis. Este tipo de ruido incluye a los ruidos introducidos por el detector, como ruido térmico, el ruido de lectura o la dependencia de la eficiencia cuántica del detector con la longitud de onda.
- ❖ **Ruido computacional:** se refiere al introducido en el proceso de digitalización de la señal de salida del detector.
- ❖ **Ruido generado por fuentes externas:** es el generado externamente al equipo Raman o a la muestra que se esté analizando. Es causado mayormente por alguna fuente de luz externa que contamina la señal en algún punto del equipo de medida, aunque si el equipo y el contenedor de la muestra a medir, están cuidadosamente diseñados, deberían ser inmunes a las radiaciones externas. Puede ser fuentes externas de ruido, la luz de los fluorescentes y la luz solar. Una fuente de ruido externo de origen no óptico es el generado por las partículas de alta energía como los **rayos cósmicos**, que llegan al detector del equipo. Los rayos cósmicos liberan un gran número de electrones que son indistinguibles de los fotoelectrones. Los electrones generados por los rayos cósmicos se concentran en uno o máximo dos de los elementos del detector. El resultado es un pico muy estrecho y de gran intensidad en el espectro de dispersión Raman. Estos picos ocurren infrecuentemente, en tiempo aleatorio y posiciones también aleatorias del espectro. Es difícil confundir un pico cósmico con una banda Raman, por la disparidad en sus características, a no ser que el azar haga caer varios rayos cósmicos en pixeles adyacentes del detector, lo cual es poco probable.



**Figura 4.1.** Principales ruidos en espectroscopía Raman.

Es por lo que esos archivos “.txt” donde están contenidos los datos serán procesados para la eliminación de los ruidos de picos de rayos cósmicos (spikes), shot y fluorescencia. Dicho proceso lo vamos a definir como “procesamiento de la señal”, para que luego esos datos “limpios” sean tratados en un análisis multivariante para una mejor interpretación.

#### 4.1. Procesamiento de la señal

Para este procesamiento, se ha construido un programa de tratamiento de texto en el lenguaje de programación Java, para su optimización, denominado MyRDP (My Raman Data Processing), el cual será utilizado junto con el OriginPro 9, en el procesamiento de las señales. Dado que este se realiza manualmente por los investigadores, el programa fue realizado en el NetBeans IDE 7.0.1, que es un entorno de desarrollo integrado libre, hecho principalmente para el lenguaje de programación Java. Además es un producto libre y gratuito sin restricciones de uso.

La programación de este tratamiento para los datos, ha sido realizada teniendo en cuenta las necesidades para este experimento, tratando siempre de darle la mayor robustez posible y permitir una interacción amigable con el usuario. De los espectros solamente se necesita una región específica, el programa permite que el usuario seleccione la región de interés, la cual va desde  $599\text{ cm}^{-1}$  a  $1800\text{ cm}^{-1}$  (puede ser otra en dependencia de lo que se esté buscando). Luego de que se ha seleccionado el rango de interés pasamos a la detección y eliminación de spikes (picos de rayos cósmicos).

#### **4.1.1. Detección y eliminación de picos de rayos cósmicos**

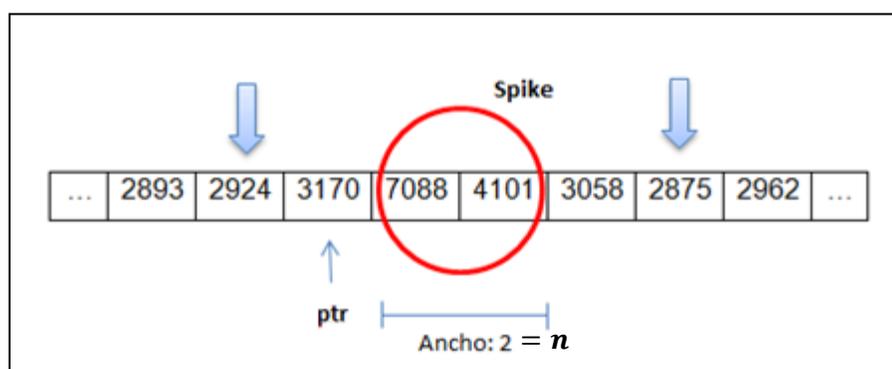
En los espectros aparecen a veces señales provenientes de rayos cósmicos (spikes), que se caracterizan por ser picos intensos y estrechos, generalmente tres o cuatro puntos con intensidad  $n$ -veces mayor que los puntos colindantes. Un punto de referencia puede ser el máximo del espectro de la hemoglobina, que se encuentra en la región  $1330\text{-}1380\text{ cm}^{-1}$ . Si se tiene una señal por encima de este máximo, estamos en presencia de un spike. También se debe considerar la aparición de spikes menos intensos, lo cual puede determinarse por variaciones súbitas del valor. En un pico normal, el valor de intensidad se incrementa gradualmente.

Se conoce que en el espectro de la hemoglobina (Hb) los picos más intensos están en las regiones de  $1330 - 1380\text{ cm}^{-1}$  y  $1530 - 1630\text{ cm}^{-1}$ . Como la muestra se mide en un sustrato de silicio, se tiene el pico de este alrededor de los  $960\text{ cm}^{-1}$ , en la región de  $930 - 990\text{ cm}^{-1}$ . Conocer dónde se encuentran las bandas características es de gran utilidad, debido a que si un spike es detectado sobre una de estas bandas, el espectro completo se desecha.

Conociendo esto, vamos a detectar el spike, escaneando el espectro punto por punto. A medida que se explora cada punto del

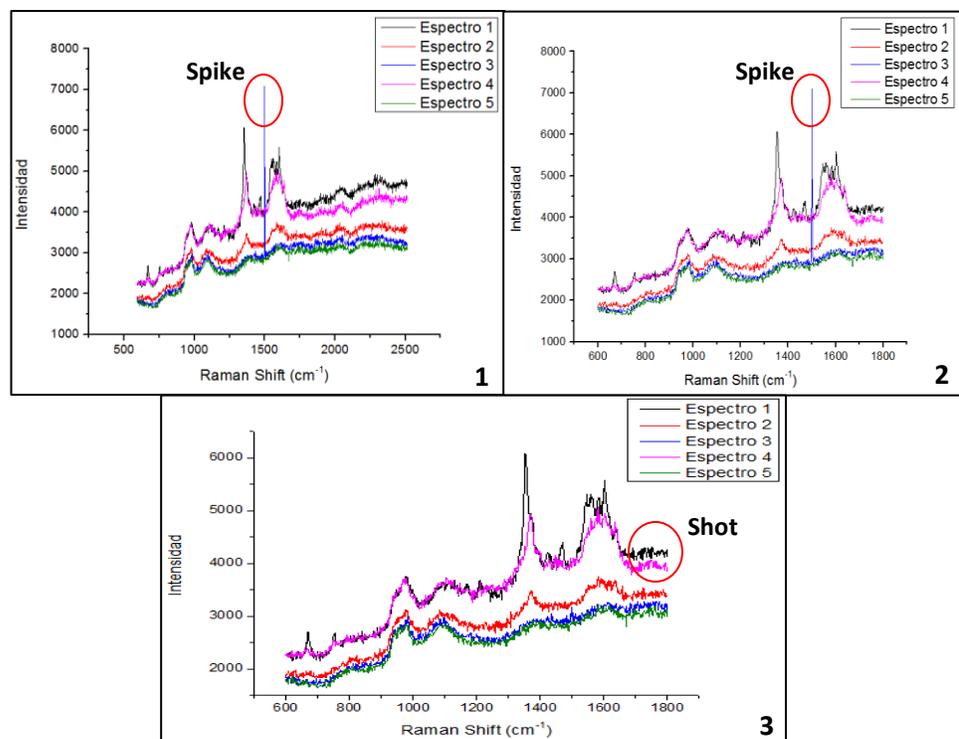
espectro, se comprueba que el valor absoluto de la diferencia entre el punto actual y el siguiente no sobrepase el valor de 500, en caso afirmativo estamos en presencia de un spike, para eliminarlo se procede de la siguiente manera:

Se le va restando el punto actual o puntero **ptr**, que no es más que el punto anterior donde comienza el spike, a los siguientes puntos mientras que la diferencia sea mayor que 500. Esto se hace para determinar el ancho espectral del spike denominado **n** (puede ser entre 1 y 4 esta anchura). Después de determinar el ancho del spike, el punto actual pasa a ser el punto donde comienza dicho spike, luego se calcula el promedio del punto que se encuentra **n** posiciones antes del punto actual y **n** posiciones después del punto actual, y ese valor se sustituye por el valor que se encuentra en el punto actual. Este procedimiento de sustituir los valores del spike se repite **n** veces.



**Figura 4.2.** Ejemplo algoritmo para la detección y eliminación de spikes.

A continuación se muestran los espectros de un conjunto de datos GR Célula 1, los originales, los originales con el rango de interés ( $599 - 1800 \text{ cm}^{-1}$ ) seleccionado y estos sin spikes.



**Figura 4.3.** 1) Espectros de GR Célula 1 originales. 2) Seleccionado el rango de interés (599 – 1800 nm). 3) Sin spikes.

#### 4.1.2. Suavizado (Smooth: FFT Filter)

Teniendo ya los datos libres de spikes, podemos eliminar el ruido shot. Por lo que suavizaremos los espectros en el OriginPro 9, el cual tiene implementado el método de filtro de la Transformada de Fourier Rápida (FFT Filter en inglés) [4.2]. La FFT es un algoritmo eficiente que permite calcular la transformada de Fourier discreta (DFT siglas en inglés) y su inversa. La FFT es de gran importancia en una amplia variedad de aplicaciones, desde el tratamiento digital de señales y filtrado digital, entre otros. El algoritmo pone algunas limitaciones en la señal y en el espectro resultante, por ejemplo: la señal de la que se tomaron muestras y que se va a transformar debe consistir de un número de muestras igual a una potencia de dos [4.3].

La idea que permite esta optimización es la descomposición de la transformada a tratar en otras más simples y éstas a su vez hasta llegar a transformadas de 2 elementos donde  $k$  puede tomar los valores 0 y 1. Una vez resueltas las transformadas más simples hay que agruparlas en otras de nivel superior que deben resolverse de nuevo y así sucesivamente hasta llegar al nivel más alto. Al final de este proceso, los resultados obtenidos deben reordenarse. El algoritmo de la Transformada de Fourier Rápida fue popularizado por Cooley y Tukey en 1965 [4.4].

Este suavizado (smooth) consiste de seis pasos:

- Calcular la media del primer 1% de los datos y del último 1% de los datos del espectro.
- Construir una línea recta a través de esos dos puntos y restar a los datos esta línea.
- Desarrollar la FFT en el conjunto de datos adquiridos en el paso anterior.
- Aplicar el filtrado con el filtro parabólico de paso bajo (low-pass parabolic filter).
- Desarrollar la IFFT en el espectro filtrado.
- Añadir una línea base al conjunto de datos adquiridos en el paso anterior.

FFT:

$$F_n = \sum_{i=0}^{N-1} x_i e^{-\frac{2\pi j}{N} ni} \quad \mathbf{4.1.2 (1)}$$

Donde  $x_i$  es una secuencia de longitud  $N$ .

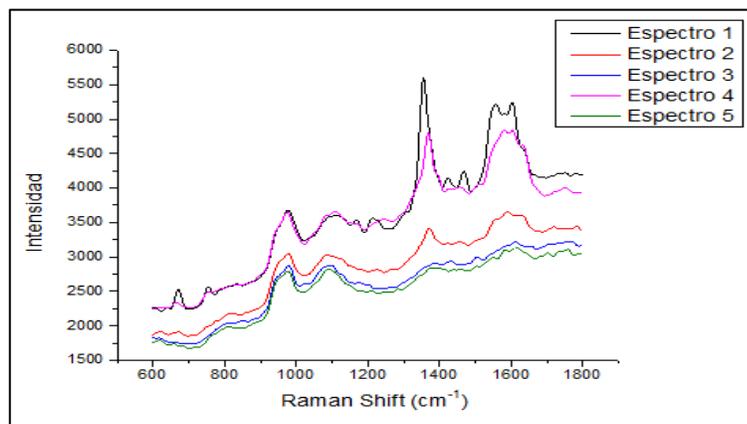
IFFT:

$$x_i = \frac{1}{N} \sum_{n=0}^{N-1} F_n e^{-\frac{2\pi j}{N} ni} \quad \mathbf{4.1.2 (2)}$$

Low-pass parabolic filter:

Dejando que  $f_{c_1}$  sea la frecuencia de paso (pass frequency) y  $f_{c_2}$  sea la frecuencia de parada (stop frequency). La función de ventana puede ser expresada por:

$$w(f) = \begin{cases} 1, & \text{si } f \leq f_{c_1} \\ 1 - \frac{(f-f_{c_1})^2}{(f_{c_2}-f_{c_1})^2}, & \text{si } f_{c_1} < f < f_{c_2} \\ 0, & \text{si } f \geq f_{c_2} \end{cases} \quad 4.1.2 (3)$$



**Figura 4.4.** Espectros sin ruido shot, suavizados mediante el Smooth: FFT Filter en el OriginPro 9.

### 4.1.3. Normalización de espectros

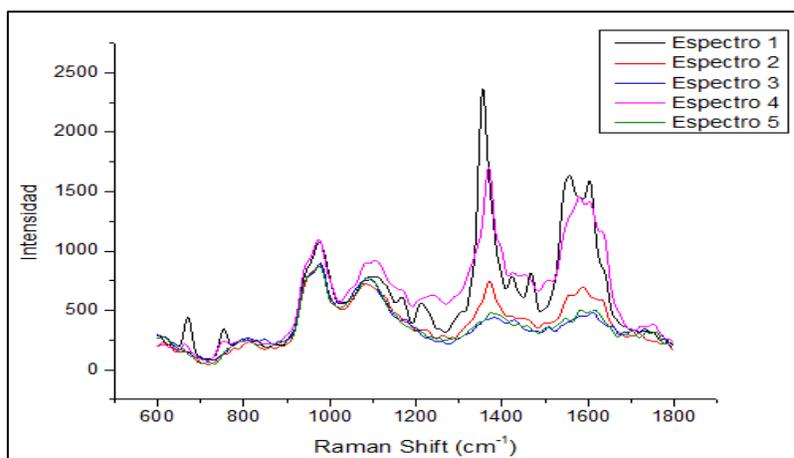
Debido a que la señal de los distintos espectros no parte del mismo valor de intensidad, se necesita reducir fondos. Para homogeneizar los espectros, primeramente se comprueba que el fondo no sea lineal. Esto es posible determinarlo comparando dos regiones determinadas que definiremos como: región llana inferior y región llana superior. Donde la región llana inferior va de  $599 - 640 \text{ cm}^{-1}$  y la región llana superior va de  $1750 - 1800 \text{ cm}^{-1}$ .

- Se compara la media de la región llana inferior ( $599-640 \text{ cm}^{-1}$ ) con la media de la región llana superior ( $1750-1800 \text{ cm}^{-1}$ ).

- Si el promedio de la región llana superior (1750 – 1800  $\text{cm}^{-1}$ ) es mayor que: la resta de los promedios de la región llana superior y la inferior + el promedio de la región llana inferior, entonces el fondo no es lineal.
- La corrección consiste en normalizar cada espectro por la ecuación de la recta dados dos puntos  $(x_1, y_1), (x_2, y_2)$ .

$$\frac{y-y_1}{y_2-y_1} = \frac{x-x_1}{x_2-x_1} \quad \mathbf{4.1.3 (1)}$$

Primero se selecciona para cada espectro el mínimo de la región llana inferior y el mínimo de la superior. Con esos dos puntos se construye una recta, la cual se le resta a cada espectro para su normalización (ecuación 4.1.3 (1)).



**Figura 4.5.** Espectros normalizados por la ecuación de la recta dados dos puntos.

## 4.2. Análisis estadístico

La evolución de las técnicas de análisis estadístico se remonta desde los años 60 y han estado sometidas a la complejidad de su cálculo, debido a la gran masa de datos a analizar. Al emerger los sistemas informáticos esta gran masa de datos se vuelve manejable en su conjunto y las técnicas de análisis multivariable contribuyeron como una revolución en la investigación.

En la actualidad se dispone de una gran cantidad de información que impulsa el estudio de fenómenos como calidad de vida, nivel socio-económico, grado de bienestar, entre otros, que no pueden medirse directamente, sino que son el resultado del estudio y análisis de un conjunto de variables relacionadas.

El análisis estadístico de datos incluye un conjunto de métodos y técnicas univariadas y multivariadas que permiten estudiar y tratar en bloque una o varias variables medidas u observadas en una colección de individuos [4.5]. El Análisis Multivariable es considerado como un conjunto de técnicas o métodos científicos que permiten tratar matrices de grandes dimensiones. Dentro de este análisis se consideran diversas técnicas, entre las que podemos destacar al Análisis de Componentes Principales (ACP).

Para la implementación de este análisis se utilizó el software Wolfram Mathematica 9. En el cual se leen los ficheros resultantes del software para el procesamiento de las señales (espectros Raman) MyRDP.

#### **4.2.1. Análisis Multivariable (AMV)**

El Análisis Multivariable es un método estadístico, utilizado en la determinación de la contribución de varios factores en un evento o resultado. Es una de las principales técnicas de análisis de datos cuando se tiene un gran número de variables que van a ser consideradas al mismo tiempo para su interpretación. Básicamente es una colección de métodos que permiten tratar problemas muy diversos en los que diferentes propiedades se miden en un conjunto específico de objetos.

Su existencia se fundamenta en un mejor entendimiento del fenómeno bajo estudio, obteniendo información que los métodos estadísticos univariantes y bivariantes no son capaces de conseguir.

---

Teniendo como objetivo proporcionar métodos o técnicas capaces de ayudar al analista o investigador a tomar decisiones óptimas en el contexto en el que se encuentre teniendo información disponible por el conjunto de datos analizado.

Las técnicas multivariantes se pueden clasificar en tres grandes grupos:

- ❖ **Métodos de dependencia:** Supone que las variables analizadas están divididas en dos grupos – las variables dependientes y las independientes – y su objetivo consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y en qué forma.
- ❖ **Métodos de interdependencia:** Estos no distinguen entre variables dependientes e independientes, su objetivo es identificar que variables están relacionadas, como es esa relación y por qué.
- ❖ **Métodos estructurales:** Incluye la función de los métodos de dependencia y además analizan cómo es la relación de las variables de los dos grupos entre sí.

(El Análisis de Componentes Principales, que veremos con más detalle a continuación, se clasifica como un método de interdependencia.)

El primer paso en el análisis de datos multivariantes es tener la matriz de datos  $\mathbf{X}$ , de la cual calculamos su vector de medias  $\bar{x}$  [4.7]. De ahí podríamos pasar a obtener la matriz de covarianzas (de dispersión) y la de correlación y de ahí comenzar a obtener las componentes principales; o primeramente tipificar los datos para de una vez tener los datos centrados y reducidos, lo que tiene como efecto que la varianza de todas las variables sea la misma e igual a la unidad.

	<b>Variable # 1</b>	<b>Variable # 2</b>	<b>...</b>	<b>Variable # p</b>
<b>Objeto # 1</b>	$x_{11}$	$x_{12}$	...	$x_{1p}$
<b>Objeto # 2</b>	$x_{21}$	$x_{22}$	...	$x_{2p}$
...	...	...	...	...
<b>Objeto # n</b>	$x_{n1}$	$x_{n2}$	...	$x_{np}$

**Tabla 4.1.** Forma de la matriz de datos X.

Veremos más adelante los pasos detallados de este análisis, al aplicarlos a nuestros datos.

#### 4.2.2. Análisis de Componentes Principales (ACP)

El Análisis de Componentes Principales (ACP) es una técnica descriptiva de reducción de dimensiones, que persigue el estudio de las relaciones de interdependencia entre grupos de variables e individuos. Este análisis tiene sus orígenes en los trabajos de Pearson publicados en 1901, en la *Philosophical Magazine* con el título “On lines and planes of closest fit to systems of points in space” [4.6].

Resulta de mucha utilidad este análisis cuando se tiene un gran número de variables e individuos y se desea representar el sistema de referencia dado por las variables (de gran dimensión) por otro más asequible, de menor dimensión. En este último los elementos serán las combinaciones lineales de las variables originales, denominadas componentes principales; de forma tal que la pérdida de información sea mínima. Donde las combinaciones lineales óptimas de las variables originales, las nuevas componentes principales, serán variables incorrelacionadas.

El mayor número posible de componentes coincide con el número original de variables. Si nos quedamos con todas ellas, no simplificaríamos el problema, por lo que el investigador debe seleccionar entre distintas

alternativas, aquellas que siendo pocas e interpretables, expliquen una proporción aceptable de la información.

El ACP permite la descomposición de los datos originales en un modelo que consta, de una parte de la señal y una parte de ruido. El análisis está orientado a modelar una estructura de varianza-covarianza de una matriz de datos a partir del cual los valores propios, correspondientes a los componentes principales, se extraen. Cada componente principal en nuestro caso, es una combinación lineal de las variables originales.

El primer componente principal representa la mayor varianza, y así corresponde al mayor valor propio. El segundo componente principal es ortogonal al primero, con cada sucesivo componente principal siendo tanto ortogonal a todos los anteriores, y que constituyan una proporción decreciente de la varianza [4.6].

Los componentes principales más pequeños suelen interpretarse como componentes de ruido. Uno puede utilizar herramientas como scores (contienen la información de cómo las muestras están relacionadas unas con otras, además tienen la propiedad de ser ortogonales) y loadings o factores (informan de la relación existente entre las variables y tienen la cualidad de ser ortonormales) de presentar los datos descompuestos por ACP.

Los scores indican cuales objetos son responsables de la mayor parte de la varianza en los datos. Cada score representa todo un espectro de trazado en una, dos o tres dimensiones espaciales. Un factor se utiliza para asignar los coeficientes a los componentes principales relevantes de un conjunto de datos de x, y-variables (es decir, valores de número de onda), y permiten decir qué variables están dominando o influenciando el modelo y como se asocian entre sí.

En resumen, en el análisis de componentes principales se dispone de una muestra de tamaño  $n$  sobre  $p$  variables  $X_1, X_2, \dots, X_p$  tipificadas, inicialmente correlacionadas, para luego a partir de ellas obtener un número  $k \leq p$  de variables incorrelacionadas  $Z_1, Z_2, \dots, Z_p$  las cuales sean combinación lineal de las variables originales y expliquen la mayor parte de su variabilidad.

Las componentes principales quedarán de la siguiente forma:

$$\mathbf{Z}_1 = a_1 X_1 ; \mathbf{Z}_2 = a_2 X_2 ; \dots ; \mathbf{Z}_p = a_p X_p$$

Donde  $\mathbf{a}_i$  son los vectores propios asociados a la matriz de covarianzas.

Sólo se retienen las  $p$  componentes principales que explican un porcentaje alto de la variabilidad de las variables originales [4.6].

#### 4.2.2.1. Aplicación del ACP al tratamiento de los datos

Los espectros registrados a partir del experimento realizado, se procesaron usando el antes mencionado Análisis de Componentes Principales (ACP), para modelar los datos y determinar los cambios espectrales asociados con los glóbulos rojos, a los cuales se les añadieron nanopartículas de plata NPAg, para intensificar su señal en el espectroscopio Raman.

Del equipo se obtiene una matriz de datos que consta de  $n$  espectros con 1201 valores. Los factores expresan las relaciones entre las variables de número de onda ( $1800-599 \text{ cm}^{-1}$ ). Por lo que los espectros tendrán  $n = 615$  valores. Los datos fueron tipificados, y una validación cruzada completa de análisis de componentes principales (ACP), fue realizada utilizando de una, dos o seis componentes principales, dependiendo de la cantidad de variables analizadas en cada caso.

Los espectros Raman son vectores de  $n$  coordenadas definidos en lo que se conoce como “espacio de señal de los números de onda normalizados”:

$$E_i = [e_{i1}, \dots, e_{ij}, \dots, e_{in}] \quad 4.2.2.1 (1)$$

Donde  $e_{ij}$  son las intensidades Raman del espectro  $E_i$  para un número de onda (desplazamiento Raman).

La rutina para realizar el ACP sobre espectros Raman es realizada con el software Wolfram Mathematica 9 y es la siguiente:

1. Crear la matriz de datos  $E$  de coordenadas  $e_{ij}$ .

$$E = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1p} \\ e_{21} & e_{22} & \dots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \dots & e_{np} \end{bmatrix} \quad 4.2.2.1 (2)$$

Las matrices de datos que se van a analizar como resultado del experimento, son de grandes dimensiones, por lo que daremos una corta representación de un ejemplo de estas:

291.475	155.455	310.351	216.53	306.147
274.966	218.726	298.851	240.657	252.82
325.457	241.996	307.35	253.782	273.492
277.95	207.268	248.851	193.91	250.166
285.442	273.539	309.351	177.036	197.839
288.937	272.812	301.854	234.165	281.514
327.431	196.084	297.355	191.293	300.188
220.927	162.358	296.858	161.423	299.864
197.422	242.631	182.361	195.552	265.539
217.919	223.906	275.865	277.683	287.216
240.418	258.182	291.37	273.815	338.893
142.915	167.457	309.874	142.946	189.569
196.415	179.734	286.379	208.079	293.247
233.916	258.012	216.886	207.213	241.926
257.415	236.289	270.391	154.345	234.604
240.917	272.567	259.899	145.48	191.284
152.42	165.847	220.407	148.616	189.964
167.921	117.125	122.914	243.75	236.643
254.426	169.405	191.423	159.887	240.325
214.931	132.686	262.932	201.024	175.006
290.435	131.967	223.441	170.16	136.687

**Figura 4.6.** Muestra de la matriz de datos de GR Célula 1.

## 2. Tipificar las variables:

Se necesita obtener el vector de medias de la matriz de datos:

$$\bar{\mathbf{e}} = (\bar{e}_1, \dots, \bar{e}_j, \dots, \bar{e}_p) \quad 4.2.2.1 (3)$$

Donde

$$\bar{e}_j = \frac{1}{n} \sum_{i=1}^n e_{ij}, \quad 4.2.2.1 (4)$$

es la media de cada variable. Obteniendo en el análisis de ejemplo lo siguiente:

$$\bar{e}_j = (498.514 \quad 356.324 \quad 321.315 \quad 583.509 \quad 332.304)$$

Luego es necesario obtener de cada variable su desviación estándar:

$$\sigma_k = \sqrt{\frac{\sum_{i=1}^n (e_{ij} - \bar{e}_j)^2}{n}} \quad 4.2.2.1 (5)$$

Para los espectros analizados tenemos:

$$\sigma_1 = 420.561, \sigma_2 = 203.176, \sigma_3 = 178.793, \sigma_4 = 381.721, \sigma_5 = 179.545$$

Debido a que la tipificación consiste en restar a cada variable su media y dividir el resultado entre la desviación estándar. Obteniéndose de este modo, nuevos valores para la matriz de datos.

$$x_{ij} = \frac{e_{ij} - \bar{e}_j}{\sigma_i} \quad 4.2.2.1 (6)$$

Donde la matriz de datos tipificada será:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad 4.2.2.1 (7)$$

-0.687938	-0.973524	-0.582103	-1.19505	-0.616782
-0.686963	-0.960398	-0.605529	-1.20585	-0.635787
-0.690853	-0.94155	-0.630554	-1.21092	-0.646775
-0.70068	-0.922068	-0.653994	-1.21065	-0.648963
-0.716639	-0.905898	-0.673412	-1.20648	-0.643121
-0.73807	-0.895225	-0.687815	-1.20057	-0.631591
-0.763598	-0.89045	-0.697928	-1.19533	-0.618
-0.791334	-0.890667	-0.706006	-1.19287	-0.606702
-0.819119	-0.894478	-0.715226	-1.19453	-0.602057
-0.844765	-0.900828	-0.728858	-1.20063	-0.607699
-0.866286	-0.909588	-0.749431	-1.21044	-0.625948
-0.882115	-0.921709	-0.778133	-1.22244	-0.657493

**Figura 4.6.** Muestra de la matriz de datos tipificada.

Calculamos la media y la varianza de los datos para cerciorarnos que la tipificación fue exitosa. Estas deben ser 0 y 1 respectivamente.

$$\bar{x}_j = (0 \quad 0 \quad 0 \quad 0 \quad 0)$$

$$\sigma_k^2 = (1.00163 \quad 1.00163 \quad 1.00163 \quad 1.00163 \quad 1.00163)$$

3. Luego de tipificar los datos y formar una nueva matriz de datos  $X$  con estos, pasamos a obtener la matriz de covarianzas  $S$ , que vendrá dada por:

$$S = \frac{1}{n} X^T X \quad 4.2.2.1 (8)$$

Donde los componentes de  $S$  van a ser proporcionales a los coeficientes de correlación, debido a que el proceso de tipificación no afecta a la correlación entre las variables.

Por lo que podemos decir que la matriz de covarianzas  $\mathbf{S}$  es proporcional a la matriz de correlación  $\mathbf{R}$  de las variables originales o de las tipificadas. Ahora estamos listos para el análisis de componentes principales.

$$\mathbf{S} = \frac{1}{615} X^T X$$

$$\mathbf{S} = \begin{bmatrix} 1 & 0.782514 & 0.543203 & 0.910249 & 0.582536 \\ 0.782514 & 1 & 0.925223 & 0.87458 & 0.945054 \\ 0.543203 & 0.925223 & 1 & 0.650663 & 0.990081 \\ 0.910249 & 0.87458 & 0.650663 & 1 & 0.690609 \\ 0.582536 & 0.945054 & 0.990081 & 0.690609 & 1 \end{bmatrix}$$

4. Se procede a la obtención de los valores y vectores propios.

Los valores propios son las raíces de la ecuación:

$$|\mathbf{S} - \lambda \mathbf{I}| = 0 \quad \mathbf{4.2.2.1 (9)}$$

$$\lambda_i = (4.17099 \quad 0.732139 \quad 0.0806134 \quad 0.00930916 \quad 0.00694674)$$

Los vectores propios asociados a los valores propios  $\lambda_i$  son resultado de resolver la ecuación:

$$\mathbf{S} \mathbf{a}_i = \lambda_i \mathbf{a}_i \quad \mathbf{4.2.2.1 (10)}$$

Que nos lleva a los vectores propios:

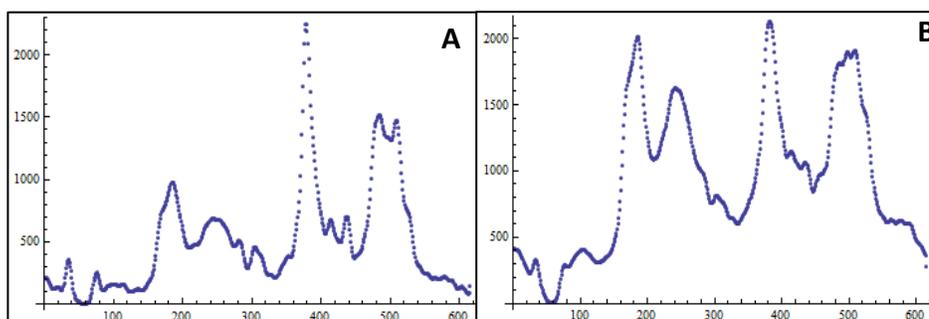
$$\mathbf{a}_1 = \begin{pmatrix} 0.406231 \\ 0.486816 \\ 0.443975 \\ 0.440934 \\ 0.454368 \end{pmatrix}; \mathbf{a}_2 = \begin{pmatrix} 0.612984 \\ -0.0781829 \\ -0.483552 \\ 0.448971 \\ -0.427482 \end{pmatrix}; \mathbf{a}_3 = \begin{pmatrix} -0.673277 \\ 0.124114 \\ -0.142528 \\ 0.710225 \\ -0.0809894 \end{pmatrix}; \mathbf{a}_4 = \begin{pmatrix} 0.0456168 \\ -0.523588 \\ 0.713104 \\ 0.232033 \\ -0.401769 \end{pmatrix}; \mathbf{a}_5 = \begin{pmatrix} 0.0620035 \\ -0.683629 \\ -0.200581 \\ 0.213876 \\ 0.665456 \end{pmatrix}$$

5. Una vez obtenidos los vectores propios en la matriz  $\mathbf{a}$ , pasamos a sacar las componentes principales:

$$\mathbf{Z} = \mathbf{a} \cdot \mathbf{X} \quad \mathbf{4.2.2.1 (11)}$$

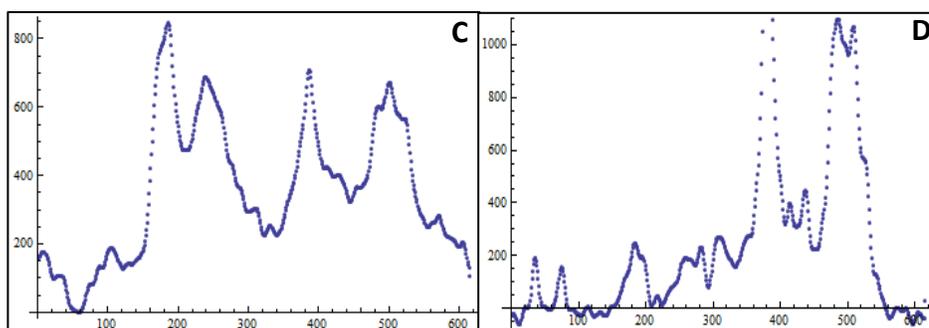
Las componentes principales de este análisis son de dimensiones tan grandes que no podemos mostrarlas, pero podemos ver la dispersión de los datos mediante estas componentes graficadas y sus porcentos de variabilidad respectivos, obtenidos de la siguiente manera:

$$\text{variabilidad}_{Z_i} = \frac{\lambda_i}{n} \times 100 \quad 4.2.2.1 (12)$$



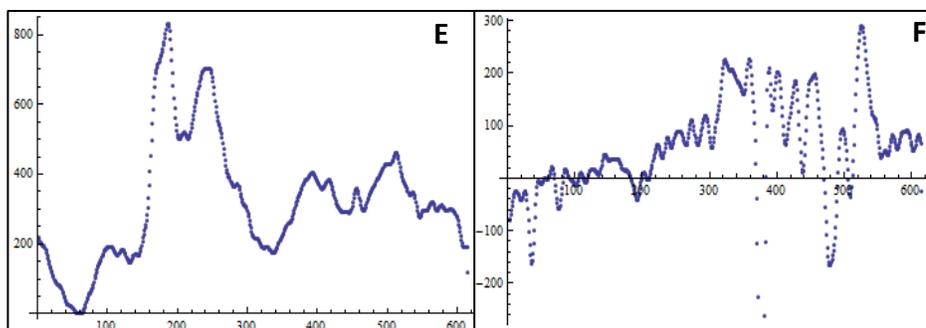
**Figura 4.7.** A) Variable original 1. B) Primera componente principal.

La primera componente principal tiene un porcentaje de variabilidad de 83.42 %.



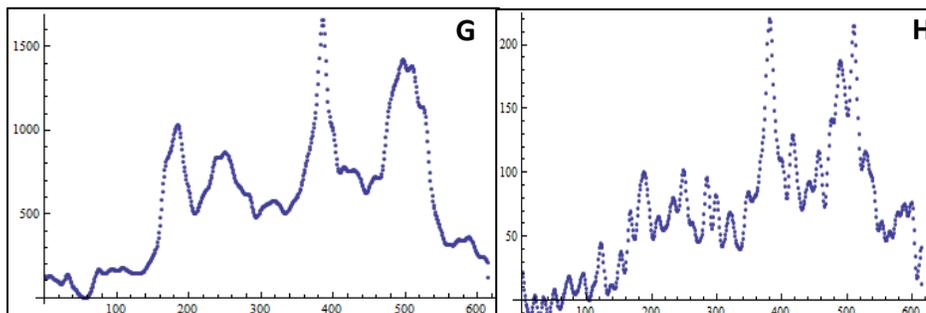
**Figura 4.8.** C) Variable original 2. D) Segunda componente principal.

Donde la segunda componente principal tiene un porcentaje de variabilidad de 14.64 %.



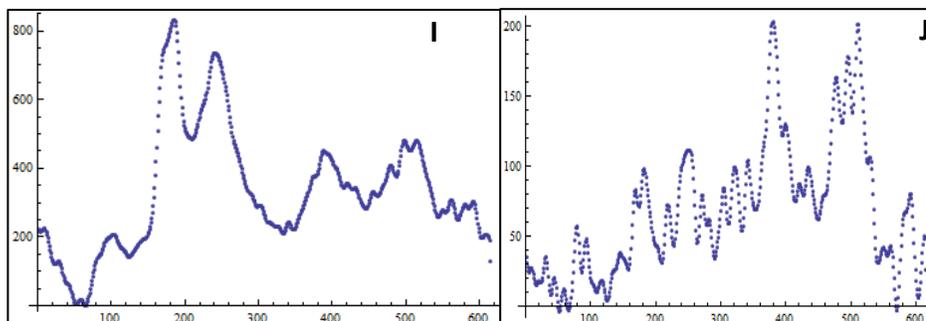
**Figura 4.9.** E) Variable original 3. F) Tercera componente principal.

Donde la tercera componente principal tiene un porcentaje de variabilidad de 1.61 %.



**Figura 4.10.** G) Variable original 4. H) Cuarta componente principal.

Donde la cuarta componente principal tiene un porcentaje de variabilidad de 0.19 %.



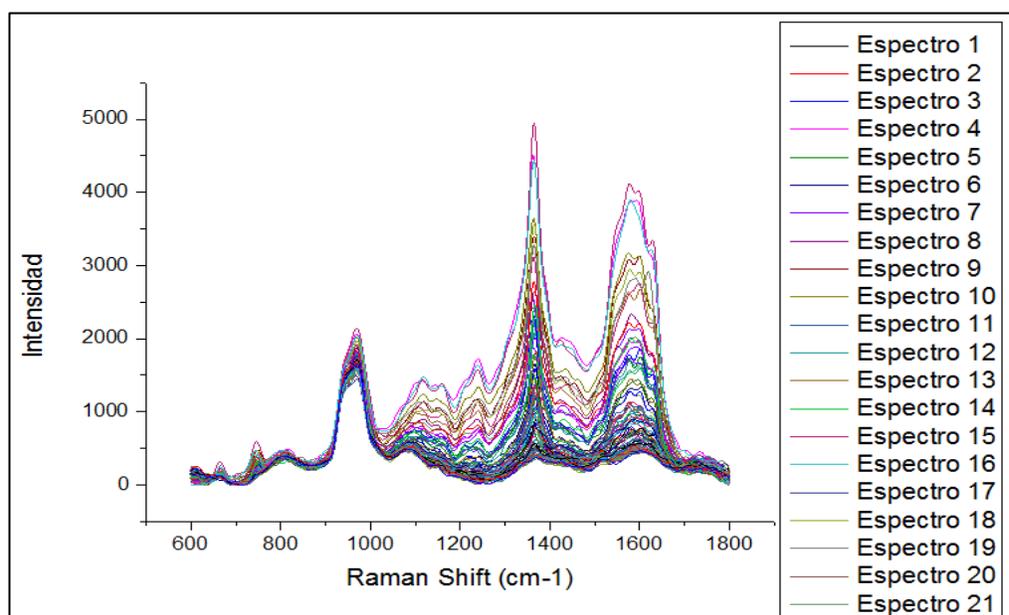
**Figura 4.11.** I) Variable original 5. J) Quinta componente principal.

Donde la quinta componente principal tiene un porcentaje de variabilidad de 0.14 %.

Para la conservación o retención de las componentes principales, el criterio más lógico es retener la menor cantidad de componentes que expliquen un porcentaje de variabilidad determinado, que en general está entre el 90% y el 80% del total. Otro criterio sería el de conservar las componentes cuyos valores propios (eigenvalues) sean mayor o igual a 1. Por último queda en manos del investigador decidir lo que mejor le conviene, debido a que depende en gran medida de la pregunta de investigación, así como del tipo de dato analizado.

#### 4.2.2.2. ACP de Espectros Raman de RBC y RBC+NPAg

- ❖ Luego de procesar los **espectros Raman de glóbulos rojos sin NPAg** con los programas MyRDP y OriginPro 9, dichos espectros quedan listos para el ACP de la siguiente forma:



**Figura 4.12.** Espectros de glóbulos rojos “Control”, 95 espectros visualizados con el OriginPro 9.

El ACP nos da los siguientes resultados (es una muestra debido a lo extenso de los datos):

Media	Desviación Estándar	Valores Propios	Coef. de CP1	Coef. de CP2	% de Variabilidad	Cumulativo	CP 1	CP 2
382.55	309.46441	<b>79.30504</b>	0.11001	-0.0489	<b>83.48%</b>	<b>83.48%</b>	-7.32278	-3.29714
466.07	353.79243	<b>15.02153</b>	0.10966	0.0386	<b>15.81%</b>	<b>99.29%</b>	-7.2976	-3.31389
399.67	327.95371	0.22762	0.11155	-0.0078	0.24%	99.53%	-7.24787	-3.32934
452.79	329.10518	0.10944	0.11076	0.0232	0.12%	99.65%	-7.19368	-3.33979
458.73	337.52877	0.06091	0.11145	0.0159	0.06%	99.71%	-7.15831	-3.34125
482.11	336.62874	0.0227	0.1097	0.0464	0.02%	99.73%	-7.16416	-3.33007
777.42	604.33305	0.01746	0.08863	0.1576	0.02%	99.75%	-7.22854	-3.30364
414.97	294.52911	0.01391	0.10824	-0.0599	0.01%	99.77%	-7.3602	-3.26085
341.12	291.54023	0.01268	0.10599	-0.0837	0.01%	99.78%	-7.55754	-3.20248
330.91	298.06853	0.01132	0.10619	-0.0823	0.01%	99.79%	-7.80874	-3.13113
473.77	353.36418	0.01052	0.11062	0.0349	0.01%	99.80%	-8.09401	-3.05092
625.51	460.54236	0.00948	0.09995	0.1147	0.01%	99.81%	-8.38936	-2.96685
...	...	...	...	...	...	...	...	...

**Tabla 4.2.** Resultados del AM → ACP para 95 espectros de Control de glóbulos rojos sin NPAg.

En la tabla se puede ver por qué se escogen las primeras dos componentes principales. Aplicando los criterios de retención de componentes, tomamos la menor cantidad de componentes que expliquen un porcentaje de variabilidad entre el 80 - 90% del total. O podría ser el de conservar las componentes cuyos valores propios (eigenvalues) sean mayor o igual a 1.

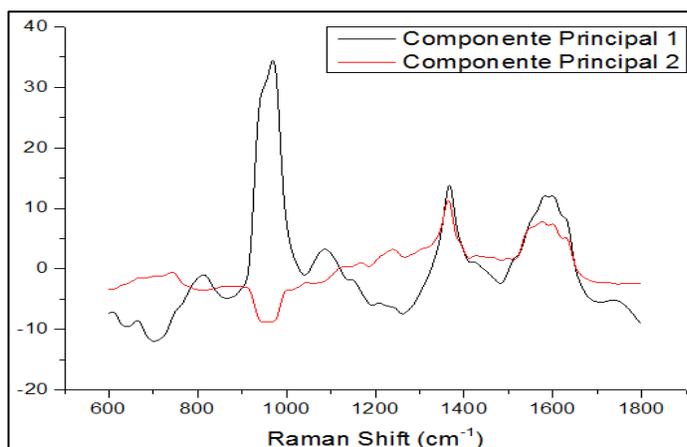
Las dos primeras componentes principales son capaces de proporcionar el 99.29% de la información que ofrecen los 95 espectros analizados. Pero al hacer una correlación entre las variables originales y las componentes retenidas, nos damos cuenta que lo mejor para el investigador es tomar sólo la primera componente. Debido a que para ver las componentes en términos de las variables originales se deben observar los coeficientes de correlación más importantes, ya sean estos

positivos o negativos próximos a 1 o -1. Puede verse en la tabla 4.3 a continuación, la cual es una pequeña muestra de los datos.

Espectros	Componente Principal 1	Componente Principal 2
1	-0.979711	-0.426388
2	-0.976525	-0.425001
3	-0.99338	-0.432337
4	-0.986372	-0.429287
5	-0.99254	-0.431971
6	-0.976899	-0.425164
7	-0.789322	-0.343527
8	-0.9639	-0.419507
9	-0.943904	-0.410804
10	-0.945682	-0.411578
11	-0.985128	-0.428745
12	-0.890115	-0.387394
13	-0.718549	-0.312725
14	-0.901906	-0.392525
...	...	...

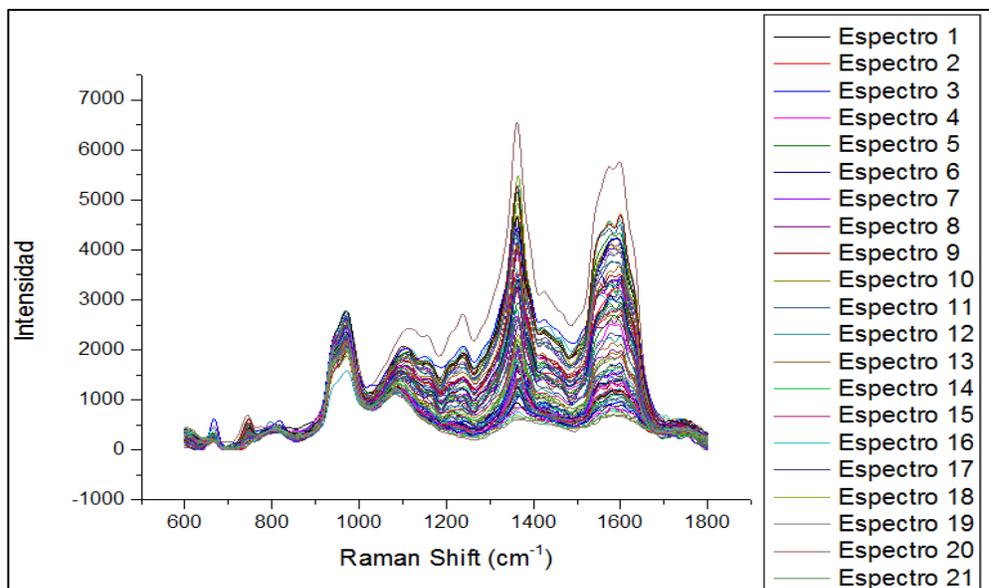
**Tabla 4.3.** Matriz de correlación entre variables originales y componentes retenidos de Control de glóbulos rojos sin NPAg.

En la segunda componente ningún valor se acerca a 1 o -1, mientras que en la primera casi todos se acercan. Podemos ver estas componentes retenidas en la figura 4.13.



**Figura 4.13.** Espectros de las 2 componentes principales con mayor porcentaje de variabilidad de Control de glóbulos rojos sin NPAg.

- ❖ Al procesar los **espectros Raman de glóbulos rojos con NPAg**, con los programas antes mencionados, los espectros quedan listos para el ACP de la siguiente forma:



**Figura 4.14.** Espectros de GR NPAg Cit 1 a 17 MP visualizados con el OriginPro 9.

El ACP nos da los siguientes resultados:

Media	Desviación Estándar	Valores Propios	Coef. de PC 1	Coef. de PC 2	% de Variabilidad	% de Variabilidad Cumulativo	CP 1	CP 2
822.52	463.97688	<b>78.07306</b>	0.10796	0.0767	<b>86.75%</b>	<b>86.75%</b>	-9.0389	1.76986
567.01	386.55598	<b>11.23389</b>	0.09223	0.1709	<b>12.48%</b>	<b>99.23%</b>	-9.0546	1.77705
1153.99	982.36779	0.28583	0.10315	-0.102	0.32%	99.55%	-9.0549	1.78984
698.30	453.76455	0.09403	0.1116	0.0472	0.10%	99.65%	-9.0530	1.80231
1055.71	744.07332	0.07872	0.10916	-0.0762	0.09%	99.74%	-9.0644	1.80754
791.03	514.74528	0.0552	0.11248	0.0167	0.06%	99.80%	-9.1033	1.79878
1449.24	1151.15063	0.03518	0.10314	-0.1222	0.04%	99.84%	-9.1809	1.77065
1197.31	859.94616	0.01207	0.10681	-0.0959	0.01%	99.85%	-9.3023	1.72013
704.64	471.56506	0.00903	0.11229	0.0261	0.01%	99.86%	-9.4661	1.64714
683.23	433.804	0.00769	0.11164	0.0424	0.01%	99.87%	-9.6644	1.55455
...	...	...	...	...	...	...	...	...

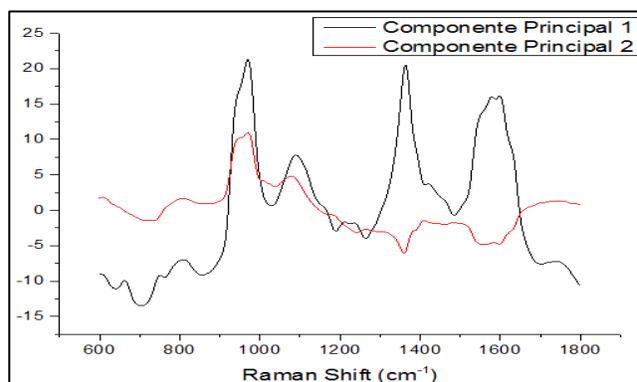
**Tabla 4.4.** Resultados del AM → ACP para 90 espectros de GR NPAg Cit 1 a 17 MP.

Las dos primeras componentes principales son capaces de proporcionar el 99.23% de la información que ofrecen los 90 espectros analizados.

Haciendo una correlación entre las variables originales y las componentes retenidas, también podemos desechar la segunda componente.

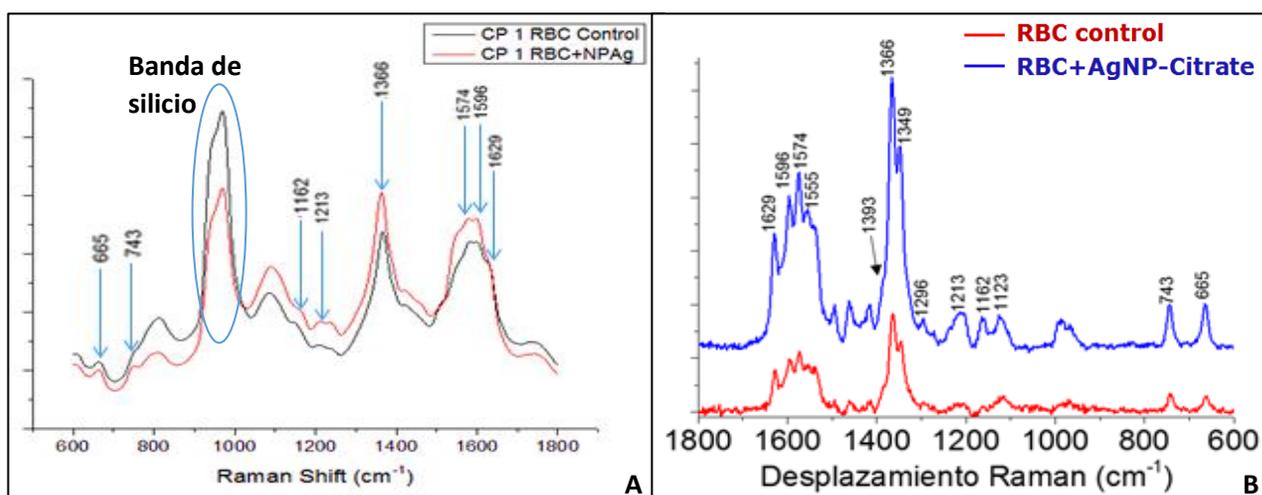
Espectros	Componente Principal 1	Componente Principal 2
1	<b>-0.953908</b>	-0.361844
2	<b>-0.814927</b>	-0.309124
3	<b>-0.911403</b>	-0.345721
4	<b>-0.986061</b>	-0.37404
5	<b>-0.964508</b>	-0.365864
6	<b>-0.993825</b>	-0.376985
7	<b>-0.911348</b>	-0.3457
8	<b>-0.943803</b>	-0.35801
9	<b>-0.992194</b>	-0.376367
10	<b>-0.986451</b>	-0.374188
11	<b>-0.925572</b>	-0.351095
12	-0.794028	-0.301197
13	<b>-0.868291</b>	-0.329367
14	<b>-0.948218</b>	-0.359685
...	...	...

**Tabla 4.5.** Matriz de correlación entre variables originales y componentes retenidos de GR NPAg Cit 1 a 17 MP.



**Figura 4.15.** Espectros de las 2 componentes principales con mayor porcentaje de variabilidad de GR NPAg Cit 1 a 17 MP.

En la figura 4.16 tenemos a las Componentes Principales retenidas de ambos análisis (Control y glóbulos rojos con las NPAGs), de las que podemos decir que se distinguen los picos característicos, pero necesitan más tratamiento para por ejemplo eliminar la banda del silicio y que se vea mejor la diferencia de intensidades entre los espectros. Para así lograr resultados semejantes a los obtenidos por N. Fleitas Salazar en ref. [3.1].



**Figura 4.16.** A) Espectros de las CP 1 de cada conjunto de datos analizados. B) Espectros obtenidos por [3.1].

## 5. CONCLUSIONES

En este trabajo se eliminaron los ruidos de los espectros obtenidos de dos ensayos independientes representativos, donde se vieron los espectros Raman de eritrocitos incubados con nanopartículas de plata y sin nanopartículas. A los cuales se les aplicó luego una técnica de análisis multivariable, el Análisis de Componentes Principales (ACP), lográndose representar el sistema referenciado dado, por otro más asequible de menor dimensión, de forma tal que la pérdida de información fue mínima.

En el procesamiento de la señal Raman para eliminar los ruidos se logró la eliminación de los picos de rayos cósmicos (spikes), mediante criterios de detección y eliminación; el suavizado de los datos mediante el método de filtro de la Transformada de Fourier Rápida; y la eliminación de la fluorescencia mediante la normalización de los datos por la ecuación de la recta dados dos puntos. Obteniéndose resultados más visibles y manejables para una mejor interpretación a la hora de realizar el análisis estadístico.

Con lo que se logró reducir el tiempo de análisis de los datos, el cual se hacía a mano y era muy engorroso. De hecho la limpieza de los espectros tenía una duración de un mes, para luego realizar el análisis estadístico.

## 6. Trabajo Futuro

Debido a que la espectroscopía Raman puede ofrecer información de diagnóstico de manera rápida y mínimamente invasiva. El desarrollo de algoritmos basados en la composición bioquímica y/o estructura del tejido permitirán un diagnóstico clínico más objetivo y que dependa menos de la experiencia del médico. La espectroscopía Raman tiene mucho potencial para proveer a la comunidad médica con una nueva herramienta de diagnóstico rápida, no invasiva y que proporcione información útil sobre el estado del tejido.

El procesamiento de los espectros Raman es un elemento importante que sería de gran ayuda que esté disponible para los médicos en tiempo real, para que fuera más rápida y eficiente la interpretación de la información clínica derivada de los espectros. Como continuación de este trabajo se podría hacer más amplio el análisis de espectros, no solo los de Raman, sino también los de UV, IR, RMN. Para esto será necesario desarrollar los algoritmos necesarios para el análisis de cada uno de ellos, que lleve a cabo todo el pretratamiento y análisis de la señal inmediatamente después de la medición, para que este resultado pueda ser usado para un diagnóstico rápido o como guía durante una intervención quirúrgica.

La espectroscopía Raman como método en el diagnóstico médico, es muy útil y necesaria, como se ha demostrado que es posible utilizarla para la diferenciación de tejido sano y enfermo. Ha permitido la detección temprana de dermatitis atópica, estimar el daño cronológico y fotoinducido en piel, la detección de fibrosis en tejidos del hígado [6.1]. Debido a lo complejo de los espectros Raman, entran en el análisis las técnicas de estadística multivariable, en particular el análisis de componentes principales, que permiten interpretar los resultados observando los porcentajes de variabilidad explicados por las componentes principales.

---

## 7. REFERENCIAS

- [1.1] Horiba Scientific <http://www.horiba.com/us/en/scientific/products/raman-spectroscopy/raman-academy/raman-faqs/what-is-surface-enhanced-raman-scattering-or-sers/> 05/02/2015.
- [2.1] “Technical Highlights. New Probe Detects Trace Pollutants in Groundwater”. Oak Ridge National Laboratory Review 26 (2).
- [2.2] Kukushkin, V. I.; Van'kov, A. B.; Kukushkin, I. V. (2013). “Long-range manifestation of surface-enhanced Raman scattering”. JETP Letters 98 (2): 64–69. doi:10.1134/S0021364013150113. ISSN 0021-3640.
- [2.3] Smith, E.; Dent, G., Modern Raman Spectroscopy: A Practical Approach. John Wiley and Sons: 2005 ISBN 0-471-49794-0
- [2.4] Campion, Alan; Kambhampati, Patanjali (1998). “Surface-enhanced Raman scattering”. Chemical Society Reviews 27 (4): 241. doi:10.1039/A827241Z.
- [2.5] Lombardi, J.R.; Birke, R.L. (2008). “A Unified Approach to Surface-Enhanced Raman Spectroscopy”. Journal of Physical Chemistry C 112 (14): 5605–5617. doi:10.1021/jp800167v.
- [2.6] Mock, J. J.; Barbic, M.; Smith, D. R.; Schultz, D. A.; Schultz, S. (2002). “Shape effects in plasmon resonance of individual colloidal silver nanoparticles”. The Journal of Chemical Physics 116 (15): 6755. Bibcode:2002JChPh.116.6755M. doi:10.1063/1.1462610.
- [2.7] Witlicki, Edward H. et al. (2011). “Molecular Logic Gates Using Surface-Enhanced Raman-Scattered Light”. J. Am. Chem. Soc. 133 (19): 7288–7291. doi:10.1021/ja200992x.
- [2.8] Talian, Ivan; Mogensen, Klaus Bo; Oriňák, Andrej; Kaniansky, Dušan; Hübner, Jörg (August 2009). “Surface enhanced Raman spectroscopy on

- novel black silicon based nanostructured surfaces”. *Journal of Raman Spectroscopy* 40 (8): 982–986. doi:10.1002/jrs.2213.
- [2.9] Shrestha, LK; Wi JS; Williams J; Akada M; Ariga K (March 2014). “Facile fabrication of silver nanoclusters as promising surface-enhanced Raman scattering substrates”. *Journal of Nanoscience and Nanotechnology* 14 (3): 2245–51. doi:10.1166/jnn.2014.8538. PMID 24745219.
- [2.10] Ngo, Ying Hui; Li, Dan; Simon, George P.; Garnier, Gil (2013). “Effect of cationic polyacrylamides on the aggregation and SERS”. *Journal of Colloid and Interface Science (Elsevier)* 392: 237–246. doi:10.1016/j.jcis.2012.09.080.
- [2.11] Chang, Yung; Yandi, Wetra; Chen, Wen-Yih; Shih, YuJu; Yang, Chang-Chung; Chang, Yu; Ling, Qing-Dong; Higuchi, Akon (2010). “Tunable Bioadhesive Copolymer Hydrogels of Thermoresponsive Poly (N – isopropyl acrylamide) Containing Zwitterionic Polysulfobetaine”. *Biomacromolecules (American Chemical Society)* 11 (4): 1101–1110. doi:10.1021/bm100093g.
- [2.12] Qu, Lu-Lu; Li, Da-Wei; Xue, Jin-Qun; Zhai, Wen-Lei; Fossey, John S.; Long, Yi-Tao (2012-02-07). “Batch fabrication of disposable screen printed SERS arrays”. *Lab Chip* 12 (5): 876–881. doi:10.1039/C2LC20926H. ISSN 1473-0189.
- [2.13] Yu, Wei W.; White, Ian M. (2013). “Inkjet-printed paperbased SERS”. *Analyst (Royal Society of Chemistry)* 138 (4): 1020. doi:10.1039/c2an36116g.
- [2.14] Bao, Li-Li; Mahurin, Shannon M.; Liang, ChengDu; Dai, Sheng (2003). “Study of silver films over silica beads as a surface-enhanced Raman scattering (SERS) substrate for detection of benzoic acid”. *Journal of Raman Spectroscopy* 34 (5): 394–398. Bibcode:2003JRSp...34..394B. doi:10.1002/jrs.993.

- 
- [2.15] Moskovits, M., (2006). Surface-Enhanced Raman Spectroscopy: a Brief Perspective. In Surface-Enhanced Raman Scattering – Physics and Applications; pp. 1–18 ISBN 3-540-33566-8.
- [3.1] Fleitas, N., (2015). Síntesis y caracterización de nanopartículas metálicas con propiedades ópticas favorables para la obtención de bioimágenes. Tesis DCF Universidad de Sonora, Departamento de Física.
- [4.1] Capítulo 1: Fundamentos de Espectroscopía Raman. Recuperado 14/11/2013, de <http://www.tdx.cat/bitstream/handle/10803/6887/03Rpp03de11.pdf>
- [4.2] William H. Press, Brian P. Flannery, Saul A. Teukolsky, William T. Vetterling., (1992). Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press. pp. 650-655.
- [4.3] Kraniuskas P., (1992). "Transforms in signals and systems". Suffolk, Great Britain. Addison-Wesley. 558 p.
- [4.4] Cooley, James W.; Tukey, John W. (1965). "An algorithm for the machine calculation of complex Fourier series". Math. Comput. 19: 297–301. doi:10.2307/2003354.
- [4.5] Pérez López, C., (2004). Técnicas de Análisis Multivariante de Datos. Madrid, España. Pearson Prentice Hall. ISBN 978-84-205-4104-4.
- [4.6] Philosophical Magazine 2:559-572. <http://pbil.univ-lyon1.fr/R/pearson1901.pdf>
- [4.7] Cuadras C. M., (2014). Nuevos Métodos de Análisis Multivariante. CMC Editions. Barcelona, España.
- [6.1] Ramírez M. G., (2011). Uso de la espectroscopía Raman como método no-invasivo en el diagnóstico médico: Tejido de piel e hígado. Tesis DCF Universidad Autónoma de San Luis Potosí, Facultad de Ciencias.