



# ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

**DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA**

**CARRERA DE INGENIERÍA EN ELECTRÓNICA, Y  
TELECOMUNICACIONES**

**TRABAJO DE TITULACIÓN, PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE INGENIERO EN ELECTRÓNICA, Y  
TELECOMUNICACIONES**

**TEMA: CREACIÓN DE UN MODELO DE CALIBRACIÓN  
MULTIVARIANTE PARA DETERMINAR EL LÍMITE DE DETECCIÓN  
DE UN PROTOTIPO NARIZ ELECTRÓNICA PARA MEDICIÓN DE  
SUSTANCIAS EXPLOSIVAS.**

**AUTOR: STA. SALAZAR MARTÍNEZ, JESSICA NATHALY**

**DIRECTOR: ING. GUAMAN NOVILLO, ANA VERÓNICA PHD.**

**SANGOLQUÍ**

**2018**

# CERTIFICACIÓN



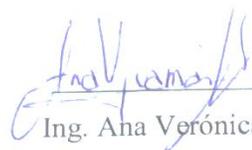
DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA EN ELECTRÓNICA,  
Y TELECOMUNICACIONES

## CERTIFICACIÓN

Certifico que el trabajo de titulación: “**CREACIÓN DE UN MODELO DE CALIBRACIÓN MULTIVARIANTE PARA DETERMINAR EL LÍMITE DE DETECCIÓN DE UN PROTOTIPO NARIZ ELECTRÓNICA PARA MEDICIÓN DE SUSTANCIAS EXPLOSIVAS**”, realizado por la señorita **SALAZAR MARTÍNEZ JESSICA NATHALY**, ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar a la señorita **SALAZAR MARTÍNEZ JESSICA NATHALY** para que lo sustenten públicamente.

Sangolquí, 06 de marzo de 2018



---

Ing. Ana Verónica Guamán Novillo PhD.  
**DIRECTOR**

## AUTORÍA DE RESPONSABILIDAD



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA EN ELECTRÓNICA,  
Y TELECOMUNICACIONES

### AUTORÍA DE RESPONSABILIDAD

Yo, SALAZAR MARTÍNEZ JESSICA NATHALY, con cédula de identidad N°: 1713973707, declaro que este trabajo de titulación “**CREACIÓN DE UN MODELO DE CALIBRACIÓN MULTIVARIANTE PARA DETERMINAR EL LÍMITE DE DETECCIÓN DE UN PROTOTIPO NARIZ ELECTRÓNICA PARA MEDICIÓN DE SUSTANCIAS EXPLOSIVAS**” ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaro que este trabajo es de mi autoría, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada.

Sangolquí, 06 de marzo de 2018

Sta. Salazar Martínez Jessica Nathaly

C.C. 1713973707

## AUTORIZACIÓN



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA EN ELECTRÓNICA,  
Y TELECOMUNICACIONES

### AUTORIZACIÓN

Yo, **SALAZAR MARTÍNEZ JESSICA NATHALY**, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca virtual de la institución el presente trabajo de titulación “**CREACIÓN DE UN MODELO DE CALIBRACIÓN MULTIVARIANTE PARA DETERMINAR EL LÍMITE DE DETECCIÓN DE UN PROTOTIPO NARIZ ELECTRÓNICA PARA MEDICIÓN DE SUSTANCIAS EXPLOSIVAS**”, cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Sangolquí, 06 de marzo de 2018

---

Sta. Salazar Martínez Jessica Nathaly

C.C. 1713973707

## DEDICATORIA

En primer lugar quiero dedicar este logro a Dios que ha sido el pilar fundamental en mi vida todos estos años, quien ha sido mis fuerzas y mi fe.

A mis padres, Ruth y Juan Carlos, quienes me han dado su apoyo incondicional siempre, y han sido partícipes de todos los triunfos y fracasos a lo largo de mi carrera universitaria, por convertirme en la persona que soy hoy en día, por transmitirme sus valores y toda su calidad humana, por no permitir que me rinda nunca ante las adversidades de la vida.

A mis hermanos, David, Kevin y Juan Pablo, por darme la motivación para ser mejor cada día y sacarme mil sonrisas cuando lo necesitaba, por invertirse los papeles y ser mi ejemplo muchas de las veces.

A mi mejor amiga Vane, que ya no se encuentra junto a mí, pero que en su corta instancia por mi vida dejó una huella imborrable, fue como una hermana para mí y siempre estará presente en mi corazón.

A mi abuelito Cesar y a mi abuelita Olguita, por haber sido los segundos padres que Dios y la vida me dieron, por haberme dado tanto cariño y tantas muestras de amor, por haber creído siempre en mí.

Jessica Nathaly Salazar Martínez

## AGRADECIMIENTO

Quiero agradecer a Dios por todas sus bendiciones, por darme a los mejores padres y hermanos, y por haber puesto en mi camino a personas que han sido un apoyo a lo largo de mi carrera.

A mis padres, Ruth y Juan Carlos por ser mi mayor ejemplo de perseverancia, constancia y responsabilidad guiándome a lo largo de mi vida, por ser mis mejores amigos y sufrir conmigo cuando me he equivocado, así mismo por alegrarse en con cada uno de mis triunfos, por todo su amor, comprensión y motivación cuando más lo necesite.

A mis hermanos Kevin, David, Juan Pablo y a mi familia por todo el cariño que me tienen, por alegrar mi vida en muchos momentos de angustia, por escuchar mis anécdotas de la universidad y aconsejarme, por estar siempre pendientes cada semestre que pasaba, por sus buenos deseos y sus palabras de aliento, muchísimas gracias.

A mi novio Cristian, por haber compartido conmigo esta linda experiencia en la universidad, por motivarme a ser mejor cada día retándome intelectualmente, por ser uno de mis mejores amigos y mi ejemplo muchas de las veces, por cuidarme y brindarme su cariño, además de ser un apoyo incondicional en los momentos difíciles dándome sus puntos de vista y sus consejos acertados, demostrarme que aunque la vida no siempre es justa hay que dar lo mejor de uno para cumplir con las metas propuestas y muchas gracias por ser esa inspiración para alcanzar mis sueños y anhelos.

A todos mis amigos, por ser la familia que escogí, por haber compartido conmigo buenos y malos momentos en esta etapa brindándome su amistad y su cariño, por acompañarme todas esas noches en vela cuando hacíamos trabajos y tareas, por compartirme sus conocimientos, por tantas risas y vivencias que han llenado mi corazón de alegría.

Agradecer de manera especial a la ing. Anita por todo el apoyo que me ha dado no solo al realizar este trabajo de titulación, sino a lo largo de la carrera, ya que como docente ha sabido impartirme sus conocimientos y ha sido una excelente guía y apoyo para mí, además de un ejemplo como persona y como profesional, que Dios la bendiga y a toda su familia.

Por ultimo quiero agradecer a todos y cada uno de los docentes que he tenido la dicha de conocer, que han sido mis profesores y amigos, ayudando en mi formación académica como ingeniera, gracias por brindarme sus conocimientos, experiencias, y consejos, por la excelente educación que he recibido, gracias a ustedes podre desenvolverme bien en el mundo laboral.

Jessica Nathaly Salazar Martínez

## ÍNDICE DE CONTENIDOS

<b>CERTIFICACIÓN .....</b>	<b>i</b>
<b>AUTORÍA DE RESPONSABILIDAD .....</b>	<b>ii</b>
<b>AUTORIZACIÓN.....</b>	<b>iii</b>
<b>DEDICATORIA .....</b>	<b>iv</b>
<b>AGRADECIMIENTO .....</b>	<b>v</b>
<b>ÍNDICE DE CONTENIDOS .....</b>	<b>vii</b>
<b>ÍNDICE DE FIGURAS.....</b>	<b>xi</b>
<b>RESUMEN.....</b>	<b>xiii</b>
<b>ABSTRACT .....</b>	<b>xiv</b>
<b>CAPÍTULO I.....</b>	<b>1</b>
<b>INTRODUCCIÓN.....</b>	<b>1</b>
1.1. Antecedentes.....	1
1.2. Justificación e importancia .....	5
1.3. Objetivos.....	7
1.3.1. Objetivo general .....	7
1.3.2. Objetivos específicos.....	7
1.4. Descripción del proyecto .....	8
<b>CAPÍTULO II.....</b>	<b>9</b>
<b>MARCO TEÓRICO .....</b>	<b>9</b>
2.1. Prototipo Nariz Electrónica .....	9
2.1.1. Funcionamiento .....	12

2.2. Modelos Multivariantes.....	13
2.2.1. Análisis de Componentes Principales (PCA).....	15
2.2.2. Modelos de Calibración .....	17
2.2.3. Modelos de regresión lineal univariante .....	18
2.2.4. Modelos de regresión lineal multivariante .....	22
2.2.5. Regresión Componentes Principales (PCR).....	24
2.2.6. Regresión por Mínimos Cuadrados Parciales (PLS).....	25
2.2.7. Validación Cruzada .....	26
2.2.8. K-fold Cross-Validation .....	27
2.2.9. Figuras de mérito.....	28
2.3. Límite de Detección .....	31
<b>CAPÍTULO III .....</b>	<b>35</b>
<b>METODOLOGÍA EXPERIMENTAL.....</b>	<b>35</b>
3.1 Experimentos.....	35
3.1.1. Toma de datos.....	37
3.2. Preprocesamiento de la señal .....	39
3.2.1. Corrección de líneas base de cada uno de los experimentos.....	40
3.2.2. Filtrado de señales .....	42
3.2.3. Concatenación de vectores .....	44
3.3. Organización de datos para creación de modelos.....	45
<b>CAPÍTULO IV .....</b>	<b>49</b>
<b>RESULTADOS.....</b>	<b>49</b>
4.1. Análisis de Componentes Principales (PCA).....	49

4.2. Modelo de Calibración .....	53
4.2.1. Determinación del número de componentes principales (PC) .....	53
4.2.2. Regresión de Componentes Principales (PCR) y Regresión por mínimos cuadrados parciales (PLS) para TNT .....	54
4.2.3. Límite de detección con TNT (LD).....	56
4.2.4. Regresión de Componentes Principales (PCR) y Regresión por mínimos cuadrados parciales (PLS) para pólvora.....	56
4.2.5. Límite de detección (LD) de pólvora .....	58
4.3. Análisis de Mezclas .....	59
4.3.1. Regresión de Componentes Principales (PCR) y Regresión por mínimos cuadrados parciales (PLS) de las dos sustancias explosivas .....	59
4.3.2. Límite de detección (LD) de las sustancias explosivas con mezclas .....	61
4.4. Discusión de resultados .....	62
4.5. Aplicabilidad de R en el proyecto realizado.....	63
<b>CAPÍTULO V .....</b>	<b>65</b>
<b>CONCLUSIONES Y RECOMENDACIONES .....</b>	<b>65</b>
5.1. Conclusiones .....	65
5.2. Recomendaciones .....	66
<b>BIBLIOGRAFÍA .....</b>	<b>68</b>

## ÍNDICE DE TABLAS

<b>Tabla 1</b>	<i>Resultados presentados por el Comando Conjunto de las Fuerzas Armadas</i>	6
<b>Tabla 2</b>	<i>Modelos Multivariantes más utilizados aplicados a datos</i>	14
<b>Tabla 3</b>	<i>Clasificación de los valores de <math>R^2</math></i>	31
<b>Tabla 4</b>	<i>Tabla de toma de datos por día de pólvora en estado puro y pólvora mezclada</i>	38
<b>Tabla 5</b>	<i>Tabla de toma de datos por día de TNT en estado puro y TNT mezclado</i>	39
<b>Tabla 6</b>	<i>Tabla de toma de datos de TNT utilizados en la validación cruzada</i>	46
<b>Tabla 7</b>	<i>Tabla de toma de datos de pólvora utilizados en la validación cruzada</i>	47
<b>Tabla 8</b>	<i>Tabla de toma de datos de TNT en estado puro y TNT mezclado</i>	47
<b>Tabla 9</b>	<i>Tabla de toma de datos de pólvora en estado puro y pólvora mezclado</i>	48
<b>Tabla 10</b>	<i>Número de componentes principales de cada sustancia con los 2 modelos</i>	53
<b>Tabla 11</b>	<i>Tabla de cifras de mérito para la sustancia TNT con los dos modelos</i>	55
<b>Tabla 12</b>	<i>Límite de detección para TNT utilizando PCR y PLS</i>	56
<b>Tabla 13</b>	<i>Cifras de mérito para la pólvora con los dos modelos</i>	57
<b>Tabla 14</b>	<i>Límite de detección para pólvora utilizando PCR y PLS</i>	58
<b>Tabla 15</b>	<i>Tabla de cifras de mérito para la sustancia TNT con los dos modelos</i>	61
<b>Tabla 16</b>	<i>Límite de detección de sustancias mezcladas con los modelos de regresión</i>	61
<b>Tabla 17</b>	<i>Resumen de cifras de mérito y límite de detección de las sustancias y mezclas</i>	62

## ÍNDICE DE FIGURAS

<b>Figura 1</b> Diagrama de bloques del prototipo Nariz Electrónica.....	4
<b>Figura 2</b> Prototipo Nariz Electrónica utilizado para el trabajo de investigación .....	10
<b>Figura 3</b> Esquema de las partes que tiene la nariz electrónica.....	12
<b>Figura 4</b> Scores de modelo PCA .....	17
<b>Figura 5</b> Regresión Lineal univariante .....	19
<b>Figura 6</b> Regresión total y de una sola muestra .....	20
<b>Figura 7</b> Hiperplano con dos variables explicativas y una dependiente .....	23
<b>Figura 8</b> Validación cruzada de 5 veces con 30 muestras.....	28
<b>Figura 9</b> Error Cuadrático Medio.....	30
<b>Figura 10</b> Distribución de valores alrededor de cero y nivel crítico .....	32
<b>Figura 11</b> Nivel crítico y límite de detección para una probabilidad $\alpha$ .....	33
<b>Figura 12</b> Nivel crítico y límite de detección para $\alpha=\beta$ .....	33
<b>Figura 13</b> Tiempo de medición. ....	36
<b>Figura 14</b> Señal original (Señal que sale de la nariz electrónica). ....	40
<b>Figura 15</b> Flujograma del algoritmo de alineación de líneas base .....	41
<b>Figura 16</b> Señales con líneas base alineadas .....	42
<b>Figura 17</b> Filtro de mediana con una ventana de $k=5$ .....	43
<b>Figura 18</b> Datos filtrados de un solo experimento en una sola matriz con 6 sensores.....	43
<b>Figura 19</b> Datos concatenados con 2100 características de 6 sensores de un experimento .....	44
<b>Figura 20</b> Datos concatenados de 149 experimentos con 6 sensores.....	45
<b>Figura 21</b> PCA de todos los datos sin suprimir el quinto sensor. ....	50
<b>Figura 22</b> Loading del segundo componente principal sin eliminar el quinto sensor.....	50
<b>Figura 23</b> PCA de todos los datos suprimiendo el quinto sensor.....	51
<b>Figura 24</b> Scores del primer componente principal de las muestras sin el quinto sensor.....	52
<b>Figura 25</b> Scores del primer componente principal de las muestras sin el quinto sensor.....	52
<b>Figura 26</b> Modelos multivariantes para TNT.....	54
<b>Figura 27</b> Mezclas proyectadas en el modelo PLS de TNT.....	55
<b>Figura 28</b> Modelos multivariantes para pólvora .....	57

<b>Figura 29</b> Predicción de mezclas de pólvora en el modelo PLS de pólvora pura.....	58
<b>Figura 30</b> Modelos multivariantes para TNT mezclado .....	59
<b>Figura 31</b> Modelos multivariantes para pólvora mezclada .....	60

## RESUMEN

En este proyecto de investigación se realizó un modelo de calibración multivariante para una nariz electrónica, que ayudó a determinar la mínima concentración de sustancia explosiva que el equipo puede analizar y predecir. Los experimentos se realizaron usando dos sustancias explosivas: TNT y pólvora en base doble, en su estado puro y mezclados con jabón y pasta dental. Las mezclas se realizaron con el propósito de ver el poder predictivo del modelo al tener las sustancias mezcladas con otras sustancias de uso común y que podrían ser usadas para el tráfico ilícito. Se realizaron 149 experimentos: 50 con TNT en estado puro, 49 con pólvora pura, 15 con TNT y Jabón, 15 con TNT y pasta dental, 10 con pólvora con jabón y 10 con pólvora y pasta dental, con diferentes concentraciones entre 0g a 3g. Se generaron modelos de calibración multivariante usando Regresión de Componentes Principales (PCR) y Regresión de Mínimos Cuadrados Parciales (PLS), para poder establecer el límite de detección del prototipo. Luego de realizar los modelos se determinaron los siguientes límites de detección del instrumento, utilizando el modelo PLS con TNT puro su límite fue de 1.42g, para pólvora en estado puro 0.99g, para TNT mezclado 1.02g y para pólvora mezclada 0.86g, determinando que con sustancias como pólvora se tiene un menor límite de detección que con sustancias como el TNT.

### **Palabras clave:**

- **NARIZ ELECTRÓNICA**
- **ALGORITMOS DE REGRESIÓN MULTIVARIANTES**
- **LÍMITE DE DETECCIÓN**
- **DETECCIÓN DE SUSTANCIAS EXPLOSIVAS**

## **ABSTRACT**

In this research project, it was made a multivariate calibration model for an electronic nose that helped to determine the minimum concentration of explosive substance that the device can consult and predict. The experiments were made using explosive substances such as TNT and gunpowder in double base, in its pure state and mixed with soap and toothpaste. The mixing was made with the purpose of analyzing the predictive power of the model when these substances are mixed with another common substances that could be used to for illicit traffic. There were accomplished 149 experiments: 50 with TNT in its pure state, 49 with pure gunpowder, 15 with TNT and soap, 15 with TNT and toothpaste, 10 with gunpowder and soap and 10 with gunpowder and toothpaste, with different concentrations between 0g and 3g. The multivariate calibration models were generated with Principal Component Regression (PCR) and Partial Least Squares Regression (PLS), in order to establish the detection limit of the prototype. After making the models, the following detection limits of the instrument were determined, using the PLS model with pure TNT its limit of 1.42g, for gunpowder in pure state 0.99g, for mixed TNT 1.02g and for mixed gunpowder 0.86g, determining that with gunpowder there is a lower limit of detection than with substances like the TNT.

### **Keywords:**

- **ELECTRONIC NOSE**
- **MULTIVARIATE REGRESSION ALGORITHMS**
- **DETECTION LIMIT**
- **DETECTION OF EXPLOSIVE SUBSTANCES**

# CAPÍTULO I

## INTRODUCCIÓN

### 1.1. Antecedentes

En la actualidad y gracias al avance de la tecnología, se han creado nuevos dispositivos bioinspirados, los cuales son capaces de imitar los sentidos humanos tales como visión, olfato, gusto y audición. Estos dispositivos están constituidos principalmente de un hardware, para adquirir información, y un software, para analizar las señales adquiridas. Los dispositivos en su gran mayoría pretenden resolver problemas biológicos o de cualquier índole, teniendo varias aplicaciones como: Visión artificial (Marvall, 1993), Narices electrónicas (Moreno, Caballero, Galán, Matía, & Jiménez, 2009), Lenguas electrónicas (Fernández de Labastida, 2011) y sistemas de audición (Valero & Francese, 2012).

Las narices electrónicas en un inicio se crearon con el fin de reemplazar a los expertos en la clasificación de aromas, ya que estos muchas veces eran costosos, difíciles de transportar, y el análisis es subjetivo a cada experto cuyo resultado se puede ver afectado por causas como el cansancio, estado de ánimo, y otros factores externos (Moreno, Caballero, Galán, Matía, & Jiménez, 2009). Algunas veces dichos expertos se han apoyado en cromatografías de gases y espectrómetros de masas, no sólo clasificando olores sino también cuantificándolos, todo esto con un elevado costo en tiempo y dinero (Moreno & Jiménez, 2009). Siendo la nariz un dispositivo sensorial con la capacidad de captar compuestos orgánicos e inorgánicos en tiempo real, pudiendo de esta forma reconocer o discriminar dentro de un conjunto de sustancias (Montenegro, Trejos, & Araúz, 2015). Hoy en día estas narices electrónicas no solo tienen aplicaciones en la clasificación de aromas, sino también en la detección y la cuantificación de olores, utilizándose en ramas como

la agroindustria (Zhang, Wang, Sheng, & Chang, 2008b), medio ambiente (Díaz, 2006), seguridad (Charumporn & Omatu, 2002), medicina (Persaud, Pisanelli, Evans, & Travers, 2006) e industria alimenticia (Saevels, y otros, 2003).

En temas de seguridad, las narices electrónicas son capaces de detectar, distinguir, clasificar y cuantificar muestras de sustancias peligrosas, gases tóxicos, explosivos entre otros (Menéndez, 2015). Las sustancias explosivas muchas veces están mezcladas con otros compuestos y son olfativamente indetectables (Lillenthal & T., 2006), esto se debe a que los explosivos al combinarse con otros componentes no afectan la estructura química de los mismos, por lo cual son camuflados para poder ser transportados y comercializados. En entrevistas realizadas a personal especializado del Comando Conjunto de las Fuerzas Armadas se investigó que las sustancias con las cuales se mezclan los explosivos para camuflarlos en su gran mayoría son alimentos, como latas de tomate, patatas, frutas, café, entre otros; otras sustancias utilizadas con este fin pueden ser la pasta de dientes o jabón, siendo de uso común.

Las narices electrónicas tienen una sensibilidad cruzada, es decir que no solo responden a un compuesto, sino a diferentes sustancias que se encuentran en el ambiente, para ello se realiza un preprocesado posterior de la señal captada por el dispositivo para la clasificación y cuantificación de las sustancias. Para poder determinar la mínima concentración de sustancia que un dispositivo, nariz electrónica, puede detectar se utilizan modelos de calibración multivariante, y ecuaciones para calcular el límite de detección, sobre todo cuando esta sustancia se encuentra mezclada con otros compuestos y es difícil determinar aproximadamente la concentración de sustancia en la mezcla.

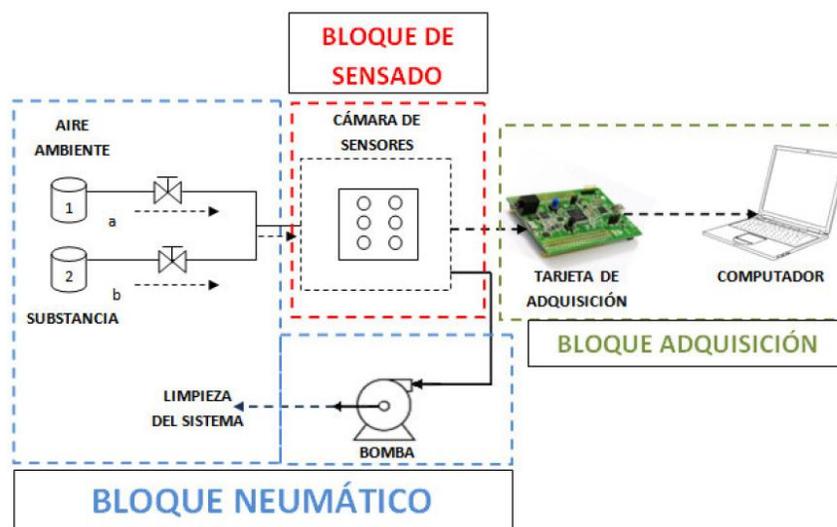
Para sustancias explosivas se ha investigado que los modelos más comunes son los modelos multivariantes como el análisis de componentes principales (PCA) (Lóp16), y para regresión se ha usado regresión de componentes principales (PCR), regresión por mínimos cuadrados parciales (PLS) y análisis mediante redes neuronales artificiales (ANN) (Olguín, 2017).

En la investigación realizada por Correa se utilizaron dos sustancias etanol y 1-propanol para predecir la concentración de cada una de las sustancias mediante la derivada de su recta de calibración respecto a la curva teórica, comparando la concentración calculada teóricamente que estaba medida en ppm. Donde se concluye que no es válido calibrar la sensibilidad de las narices electrónicas usando concentraciones crecientes sin introducir un factor de corrección (Correa, 2003).

Por otro lado en el laboratorio Nacional de Ciencia Forense de Suecia analizó la concentración de etanol en muestras de aliento utilizando una nariz electrónica, mediante sistemas neuroborrosos y redes neuronales artificiales (ANN), obteniendo un error estándar de 16,3 mol de ppm de Etanol (EtOH) en fase gaseosa (Santanu, Rajib, & Bhattacharyya, 2012).

En un estudio realizado por Paredes y otros (2006), se desarrolló una nariz electrónica que permite la detección de alcoholes, mediante 4 sensores de gas SnO<sub>2</sub>, un sistema neumático irregular, un hardware y software para la adquisición de datos. Se evaluó el comportamiento de la nariz y las condiciones de trabajo con muestras de vapor de alcoholes (metanol, etanol, n-butanol y 1-2 octanol) y se determinó que los alcoholes se pueden detectar con el arreglo de sensores preparado y pueden diferenciarse entre sí haciendo uso del análisis estadístico de componentes principales (PCA), además obtuvieron un límite de detección para cada una de las sustancias de prueba del experimento (Paredes-Doig, Sun-Kou, Picasso-Escobar, Doig-Camino, & Comina, 2006).

En un trabajo realizado por López (2016) se desarrolló una nariz electrónica para la detección de dinitrotolueno, vinagre, etanol, trinitrotolueno, pólvora en base doble y alcohol constituida por 3 bloques: (i) Bloque de sensado, (ii) Bloque neumático y (iii) Bloque de adquisición de datos, como se observa en la figura 1. Para la discriminación de las dos sustancias se utilizaron modelos multivariantes: uno de ellos el Análisis de Componentes Principales (PCA) con ello se logró tener un número menor de variables, también se utilizó un Análisis Discriminante Lineal (LDA) que garantizó la mayor separabilidad de las clases. Luego de ello utilizando el algoritmo del clasificador de los K vecinos más cercanos logró clasificar los datos que no pertenecían a ninguna de las 2 clases para obtener un error mínimo al momento de discriminar entre una u otra sustancia. Obteniéndose un 70% de eficiencia en la detección de sustancias explosivas como TNT y pólvora en base doble y un 86.66% de discriminación entre sustancias no explosivas (López, 2016).



**Figura 1** Diagrama de bloques del prototipo Nariz Electrónica

Fuente: (López, 2016)

Por otro lado, en el trabajo de titulación de Vallejo y Zurita (2017) realizado en la Universidad de las Fuerzas Armadas se creó un dispositivo con algunas mejoras del trabajo realizado por (López, 2016), (Vallejo & Zurita, 2017) donde se hicieron experimentos con mecánica de fluidos y transferencia convectiva de masa; así también desarrollaron un sistema de control de temperatura, con la finalidad de que el dispositivo pueda ser utilizado en cualquier tipo de ambiente otorgando portabilidad al mismo. Además, el prototipo lleva un método de dopaje automático que no altera la muestra a analizar. (Vallejo & Zurita, 2017) La nariz electrónica aumentó su portabilidad y mejoró considerablemente su funcionamiento en condiciones reales de análisis de sustancias.

Para el presente trabajo de investigación se desarrollaron algoritmos de preprocesamiento de las señales y se crearon modelos de calibración multivariante para discriminar y cuantificar dos sustancias explosivas: TNT y pólvora, también se realizó el cálculo del límite de detección para cada sustancia ya sea en estado puro como cuando se encuentran mezcladas con otros componentes como el jabón y la pasta dental, a diferentes concentraciones utilizando el prototipo realizado por Vallejo y Zurita (2017).

## **1.2. Justificación e importancia**

Desde el 2013, en el Ecuador, se creó un Plan Nacional de Seguridad Integral, cuyo principal objetivo es brindar seguridad y protección al ser humano (Ministerio de Coordinación de Seguridad, 2013). Enmarcados en este plan, las Fuerzas Armadas del Ecuador realizan constantes operaciones a nivel nacional para controlar la venta, porte de armas y sustancias ilícitas en el país. En el año 2015 se presentaron los resultados que se pueden observar en la Tabla 1, en donde el

número de explosivos incautados, decomisados o abandonados es de 8998 (Ministerio de Defensa Nacional del Ecuador, 2015) .

**Tabla 1**

*Resultados presentados por el Comando Conjunto de las Fuerzas Armadas*

<b>Resultados</b>	<b>Ene- Mar</b>	<b>Abr- Jun</b>	<b>Jul- Sep</b>	<b>Oct- Dic</b>	<b>Total 2015</b>
<b>Numero de armas decomisadas, incautadas, abandonadas</b>	1065	333	95	232	1725
<b>Numero de municiones decomisadas, incautadas, abandonadas</b>	67396	20076	5858	23825	117155
<b>Numero de explosivos decomisados, incautados, abandonados</b>	7402	851	303	442	8998

Fuente: (Ministerio de Defensa Nacional del Ecuador, 2015)

Para combatir estos índices presentados por el Comando Conjunto de las Fuerzas Armadas se pueden utilizar alternativas como el uso de espectrómetros por movilidad iónica para la detección de explosivos (Boggio, y otros, 2008) o el uso de narices electrónicas que sean capaces de detectar estas sustancias (López, 2016) y en general sustancias peligrosas (Menéndez, 2015). Las narices electrónicas se asemejan al sentido del olfato y probablemente son mejores detectando ciertos compuestos imperceptibles por los seres humanos (Lillenthal & T., 2006). Para los operativos que realiza periódicamente la Policía Nacional, se utilizan canes entrenados que detectan las sustancias explosivas e ilícitas por medio del olfato, pero la utilización de un dispositivo facilitaría este tipo de trabajos, haciéndolos menos peligrosos para estos animales y con un costo relativamente menor. El contar con un dispositivo que no solo detecte la sustancia,

sino que además pueda determinar el grado de concentración de la misma en la mezcla puede reemplazar a los procesos que antes se realizaban manualmente, así que es algo muy beneficioso para las personas que trabajan con estos explosivos, brindándoles mayor seguridad.

En este proyecto de investigación se pretende realizar modelos multivariantes para la calibración de una nariz electrónica, que ayudará a determinar la mínima concentración de sustancia explosiva que el equipo va a poder analizar y predecir. El presente trabajo está enmarcado en el proyecto de investigación 2016-pic-009, cuyo objetivo es discriminar dos sustancias explosivas como son TNT y pólvora en base doble para determinar la limitación que tiene el prototipo nariz electrónica, además obtener el límite de detección del dispositivo cuando estas sustancias se encuentran mezcladas con otros compuestos como jabón y pasta dental.

### **1.3. Objetivos**

#### **1.3.1. Objetivo general**

Generar un modelo de calibración multivariante para determinar el límite de detección del prototipo nariz electrónica para medición de sustancias explosivas.

#### **1.3.2. Objetivos específicos**

- Crear un modelo multivariante para la calibración de una nariz electrónica.
- Determinar la mínima concentración de sustancias explosivas que se encuentran mezcladas con otros componentes, que es capaz de diferenciar el prototipo.
- Realizar el análisis estadístico del modelo multivariante.
- Analizar los resultados obtenidos mediante el software R.

#### **1.4. Descripción del proyecto**

En este proyecto se desarrollan modelos de calibración multivariante para determinar la mínima concentración de sustancias explosivas TNT y pólvora que el prototipo nariz electrónica puede discriminar y detectar. El documento cuenta con 5 capítulos en los cuales se desarrollan cada una de las etapas que intervienen en el proyecto de investigación:

En el capítulo II se presenta el marco conceptual, describiendo algunas definiciones básicas como la definición de la nariz electrónica y el funcionamiento de la misma, los modelos de calibración multivariantes, validación cruzada, límite de detección y errores que se pueden obtener al utilizar estos algoritmos.

En el capítulo III se muestra la metodología experimental, es decir el procedimiento para la adquisición de los datos, el proceso para determinar las concentraciones de las sustancias de prueba, además del pre procesamiento de las señales para acondicionar la respuesta de la nariz electrónica.

Posteriormente en el capítulo IV se presentan los resultados obtenidos por los modelos de calibración multivariante y por los modelos estadísticos de límite de detección para determinar la mínima concentración que el dispositivo puede detectar, también presenta una discusión en donde se analizan los resultados obtenidos y se realiza un análisis comparativo de los modelos de calibración multivariante escogidos para estos propósitos, además se evalúa la viabilidad de utilizar el software R para el análisis de datos.

En el último capítulo se muestran las conclusiones de acuerdo a los objetivos planteados al inicio y las recomendaciones que pueden ser utilizadas para trabajos futuros relacionados con este proyecto.

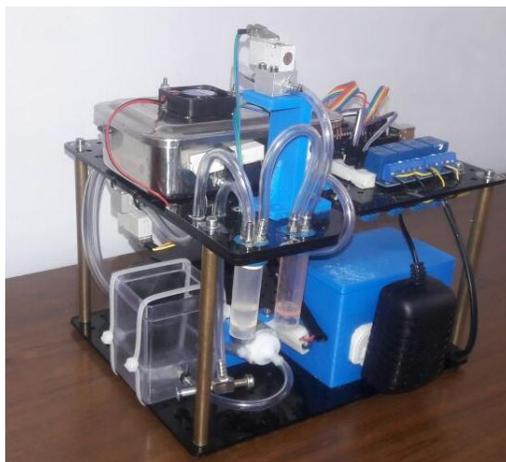
## **CAPÍTULO II**

### **MARCO TEÓRICO**

En este capítulo se presentan los conceptos básicos utilizados a lo largo de la investigación, donde se presenta una explicación rápida del funcionamiento del prototipo con el cual se realizó el análisis de las sustancias explosivas. Además, los diferentes modelos multivariantes que se usaron para analizar de forma cualitativa como el análisis de componentes principales (PCA) y cuantitativa para obtener una curva de calibración específicamente regresión de componentes principales (PCR) y regresión por mínimos cuadrados parciales (PLS) . Por otro lado se describe la metodología de validación cruzada usada para configurar los parámetros de los modelos y por último el concepto de límite de detección para determinar la mínima cantidad de sustancia a ser detectada.

#### **2.1. Prototipo Nariz Electrónica**

El prototipo que se utilizó en este proyecto fue realizado por Vallejo y Zurita (2017), el cual es un dispositivo compacto y portable que consta de una cámara de sensores compuesta por una caja metálica hermética para que los cambios externos influyan lo menos posible en la toma de datos. En la Figura 2 se observa un esquema del prototipo nariz electrónica el cual consta de las siguientes partes: Cámara de sensores, Bloque Neumático, Sistema de Dopaje y Bloque de Adquisición.



**Figura 2** Prototipo Nariz Electrónica utilizado para el trabajo de investigación  
Fuente: (Vallejo & Zurita, 2017)

- **Cámara de Sensores**

La cámara de sensores está hecha de una caja metálica de acero inoxidable herméticamente cerrada que tiene como objetivo eliminar el ruido y las interferencias que se pueda tener. Las dimensiones de la nariz electrónica es de 18.8cm de largo x 14.2cm de ancho, en la parte de arriba de la caja se tiene un ventilador que facilita la limpieza del sistema de dopaje, dentro de la caja metálica se encuentra una placa que contiene 6 sensores químicos de la familia T-GS (dos de ellos son del modelo T-GS2610 y los demás son de los modelos T-GS822,822,826,825), que se encargan de convertir los olores en señales análogas de voltaje en función del tiempo, Los sensores están distribuidos a manera de una matriz de 3x2, es decir dos columnas de tres sensores cada una como se observa en la figura 3.

También se tiene un sensor de temperatura DTH100 que es algo muy útil, ya que al momento en que el dispositivo empieza a limpiar hay que esperar a que la temperatura se estabilice para que esta influya lo menos posible en la toma de estos datos, además que si por alguna razón hay un

sobrecalentamiento de la cámara de sensores con este sensor se puede medir estas altas temperaturas y evitar posibles daños al dispositivo electrónico.

- **Sistema de dopaje automático**

El sistema de dopaje automático funciona como parte del bloque neumático, está constituido de un reservorio en el cual se almacena la sustancia dopante (alcohol), también tiene 2 contenedores más pequeños: uno para alcohol y otro para la sustancias de prueba como se muestra en la figura 3, este proceso funciona de la siguiente manera: se vacía el contenedor de alcohol, se enciende la bomba de alcohol por 2s y se inicia el proceso de respiración. Si el sistema ha cumplido con el ciclo de respiración el sistema automáticamente inyecta 2ml de alcohol al contenedor dopante.

- **Bloque Neumático**

El bloque neumático se encarga del proceso de respiración y limpieza del sistema, la respiración tiene un tiempo estimado de 60 segundos y la limpieza de la cámara de sensores que dura aproximadamente 7 minutos, que se activa de forma secuencial y automática por 4 ciclos. Además, se encarga del dopaje automático de alcohol, para todo este proceso utilizaron los siguientes elementos.

- 3 Electroválvulas: La primera permite el paso de aire del compresor a los contenedores, la segunda permite el paso de aire contaminado a la cámara de sensores y la tercera permite el paso de aire limpio a la cámara de sensores.

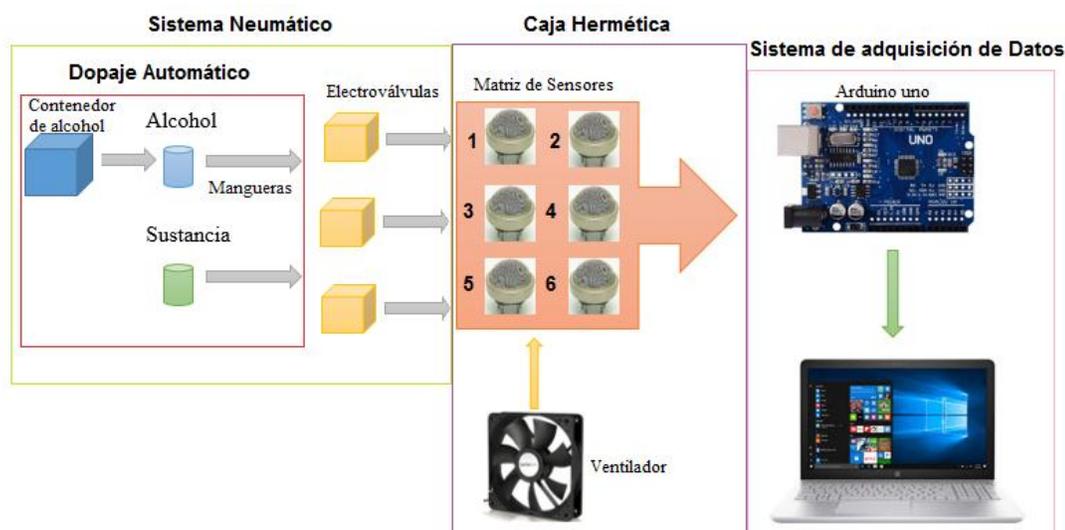
- 1 Compresor de aire: Entrega el aire al sistema de respiración.

- Mangueras de 4mm de diámetro: Por las cuales atraviesa el aire contaminado y el aire limpio.

- 2 contenedores: Donde se almacenan la muestra y el dopante.

- **Bloque de adquisición de datos**

El sistema de adquisición de datos se compone de un arduino UNO y una interfaz de adquisición de datos en el software NETBEANS que permite recolectar y exportar los datos obtenidos por los sensores químicos a un documento en Excel para posteriormente realizar el procesamiento de las señales.



**Figura 3** Esquema de las partes que tiene la nariz electrónica

### 2.1.1. Funcionamiento

Para poder tener un correcto funcionamiento del dispositivo se deben seguir algunos pasos los cuales se listan a continuación:

1. Primero se conecta la nariz electrónica al toma corriente con un cable de poder y se conecta el arduino uno a la computadora mediante un cable AB (cable de impresora).

2. Luego se verifica que la cantidad de alcohol en el primer recipiente sea de 2ml y que el reservorio de alcohol este aproximadamente en la mitad para cuando el alcohol del recipiente este menos de los 2ml este se dope automáticamente.
3. Se abre el programa que se tiene en la computadora de java en NETBEANS y se presiona el botón de inicio: se enciende el ventilador y empieza el proceso de limpieza de la nariz, esto se realiza por aproximadamente 7 minutos hasta que este proceso se detenga, luego se vuelve a presionar inicio en el programa y esperamos para que se establezca la temperatura en la nariz electrónica al valor de referencia de 29°C dependiendo de la temperatura ambiente, entonces la nariz electrónica empieza a limpiar mediante el ventilador por 7 minutos y luego por 1 minuto empieza a absorber los olores de la sustancia, se activan los relés, dando paso a la activación de las electroválvulas y por ende aquí es donde los sensores empiezan a captar los datos, otra vez limpia la nariz y vuelve a absorber, en un total de 16 minutos por muestra.
4. Se copian los datos obtenidos por la nariz en un documento de Excel y se guardan para ser procesados.

## **2.2. Modelos Multivariantes**

Los sensores químicos al igual que la nariz humana no son completamente específicos para un determinado olor, por lo que un sensor responde a una variedad de aromas ya que tiene una sensibilidad variable; para ello se requiere el uso de métodos de reconocimiento de patrones para procesar los datos y solucionar el problema de predicción de sustancia analizada (Moreno, Caballero, Galán, Matía, & Jiménez, 2009).

Los modelos multivariantes en la práctica se utilizan para cuando se tienen múltiples respuestas de sensores al mismo tiempo, en lugar de tener la respuesta de un sensor por cada experimento. Por lo tanto, el análisis de esta información se vuelve complejo, para lo cual se crean matrices de  $n$  experimentos  $\times$   $m$  respuestas de los sensores, obteniéndose una matriz  $X$  ( $n \times m$ ). Asociadas a esta matriz está la matriz de etiquetas  $Y$  de dimensiones ( $n \times 1$ ) y es usada en los modelos supervisados.

En la tabla 2 se pueden observar los modelos multivariantes más utilizados entre los cuales la regresión de componentes principales (PCR) y la regresión por mínimos cuadrados parciales (PLS) son modelos cuantitativos y el PCA es un modelo de análisis de patrones o cualitativo.

**Tabla 2**

*Modelos Multivariantes más utilizados aplicados a datos*

<b>Cuantitativo</b>		MLR (Regresión Lineal Múltiple)
		PLS (Mínimos Cuadrados Parciales )
		PCR(Regresión del Componente Principal)
<b>Análisis de Patrones</b>	No Supervisados	NFS (Sistema Neuroborrosos)
		Wavelets(Análisis de Ondas)
	Supervisados/ No Supervisados	DFA(Análisis de la Función Discriminante)
		LDA(Análisis Discriminante Lineal)
		PCA(Análisis de Componentes Principales)

Fuente: (Moreno, Caballero, Galán, Matía, & Jiménez, 2009)

### 2.2.1. Análisis de Componentes Principales (PCA)

El análisis de Componentes Principales permite evaluar rápidamente cualquier relación entre variables cuando se trata de conjuntos con muchas variables (Holland, 2016). La idea general es que se producen combinaciones lineales de las variables originales para generar nuevos ejes en una nueva dimensión también conocidos como componentes principales o PC. Estos nuevos ejes están asociados a la varianza que tenga los datos de tal forma que la varianza de algunos ejes puede ser grande mientras que otros pueden ser pequeños; de modo que los de menor varianza puedan estar relacionados con poca información lo que se conoce como la reducción de la dimensionalidad. Por ejemplo, se podría comenzar con treinta variables originales, pero terminar con solo dos o tres ejes significativas con la mayor varianza.

Dada una matriz de datos  $X$  ( $p \times n$ ) con  $p$  variables y  $n$  muestras, los datos se centran primero en la media de cada variable. Esto asegurará que la nube de datos se centre en el origen de los principales componentes, pero no afecta las relaciones espaciales de los datos ni las variaciones a lo largo de las variables. Los primeros componentes principales ( $Y_1$ ) vienen dados por la combinación lineal de las variables  $X_1, X_2, \dots, X_p$  tal como se ve en (3).

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad (3)$$

O, en notación matricial

$$Y_1 = a_1^T X \quad (4)$$

El primer componente principal se calcula de tal manera que representa la mayor cantidad posible de varianza en el conjunto de datos. Por supuesto, uno podría hacer que la varianza de  $Y_1$  sea lo más grande posible elegir valores grandes para los pesos  $a_{11}, a_{12}, \dots, a_{1p}$ . Para evitar esto, se calculan los pesos con la restricción de que su suma de cuadrados es 1.

$$a_{11}^2 + a_{12}^2 + \dots + a_{1p}^2 = 1 \quad (5)$$

El segundo componente principal se calcula de la misma manera, con la condición de que es no correlacionado con (es decir, perpendicular a) el primer componente principal y que tiene la siguiente mayor varianza.

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \quad (6)$$

Esto continúa hasta que se hayan calculado un total de  $p$  componentes principales, igual al número original de variables. En este punto, la suma de las varianzas de todos los componentes principales será igual a la suma de las varianzas de todas las variables, es decir, todas las originales donde la información ha sido explicada o contabilizada. Colectivamente todas estas transformaciones de las variables originales para los componentes principales son

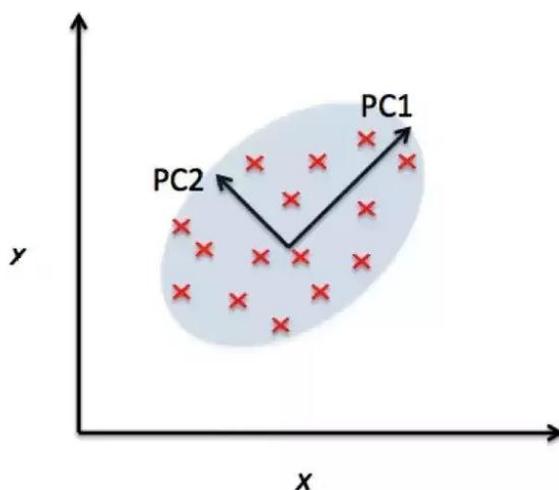
$$Y = XA \quad (7)$$

Se calcula la matriz de covarianza  $S_x$  de datos originales  $X$ , que viene dada por (8), también se calcula la matriz de covarianza  $S_y$  (9) de los componentes principales.

$$S_x = cov(X) \quad (8)$$

$$S_y = cov(CP) \quad (9)$$

Las filas de la matriz  $A$  se llaman vectores propios de la matriz  $S_x$ . Los elementos de un vector propio son los pesos  $a_{ij}$ , y son también conocidos como cargas. Los elementos en la diagonal de la matriz  $S_y$ , se conocen como los valores propios. Las posiciones de cada observación en este nuevo sistema de coordenadas de componentes principales son llamados scores y se calculan como combinaciones lineales de las variables originales (eje  $x$ ) y los pesos  $a_{ij}$  (eje  $y$ ) (véase figura 4). Los cuales están dados por la ecuación (10).



**Figura 4** Scores de modelo PCA

Fuente: (Raschka, 2015)

$$Y_{rk} = a_{1k}x_{r1} + a_{2k}x_{r2} + \dots + a_{pk}X_{rp} \quad (10)$$

Uno de los principales objetivos del análisis de componentes principales es la reducción de dimensionalidad, es decir, enfocarse en unos pocos componentes principales versus muchas variables.

En este trabajo de investigación se utilizó el análisis de componentes principales para ver los efectos experimentales que afectan a los modelos de calibración multivariantes y el poder de discriminación entre las dos sustancias explosivas con y sin mezclas.

### 2.2.2. Modelos de Calibración

Un modelo de calibración pretende realizar una cuantificación para poder predecir una cantidad determinada de sustancia, para este trabajo se utiliza TNT y pólvora, estos se los puede definir mediante cantidades, que a diferencia de los modelos cualitativos solo definen y clasifica por medio de cualidades. Los modelos de calibración pueden clasificarse en modelos univariante y multivariantes.

### 2.2.3. Modelos de regresión lineal univariante

Para análisis de regresión univariante, (Muños, 2015) se usan para explicar las variables:  $y$  en función de una variable  $x$ , en donde  $y$  es la variable dependiente y  $x$  la variable independiente y se representa (Morelisi, Modelo de regresion lineal simple, 2018) mediante (12)

$$y = f(x) \quad (12)$$

Si se considera que la función  $f$  es lineal, entonces se puede escribir mediante (13).

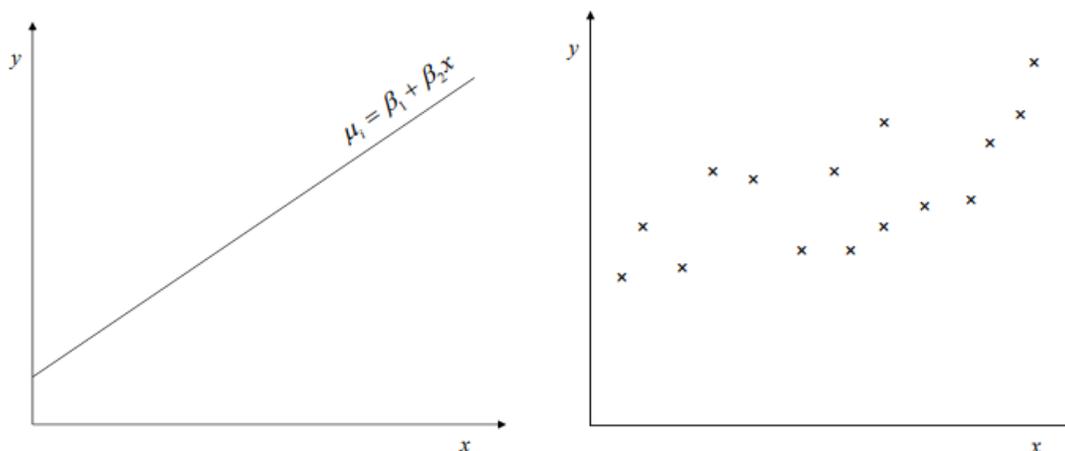
$$y = \beta_1 + \beta_2 x + u \quad (13)$$

Donde,  $\beta_1$  es el término independiente u ordenada al origen y  $\beta_2$  es la pendiente de la recta y  $u$  es denominado error o perturbación aleatoria (Uriel, 2013).

$$u_y = \beta_1 + \beta_2 x \quad (14)$$

La ecuación (14) es conocida como la función de regresión poblacional o recta poblacional. Por lo tanto, como puede verse en la figura 5a,  $u_y$  es el componente sistemático que es función lineal de  $x$ , los parámetros  $\beta_1$  y  $\beta_2$  son fijos y desconocidos.

La linealidad significa que al aumentar  $x$  el valor esperado de  $y - u_y$  varíe en  $\beta_1$  unidades. Si se dispone de una muestra aleatoria de tamaño  $n$   $(y_i, x_i): i = 1, \dots, n$  el diagrama de dispersión será algo como lo observado en la figura 5b, en donde se muestran datos hipotéticos de muestra.



**Figura 5** Regresión Lineal univariante

Fuente: (Uriel, 2013)

El modelo poblacional para cada observación de la muestra se puede expresar como (15)

$$y_i = \beta_1 + \beta_2 X_i + u_i \quad i = 1, 2, 3, \dots, n \quad (15)$$

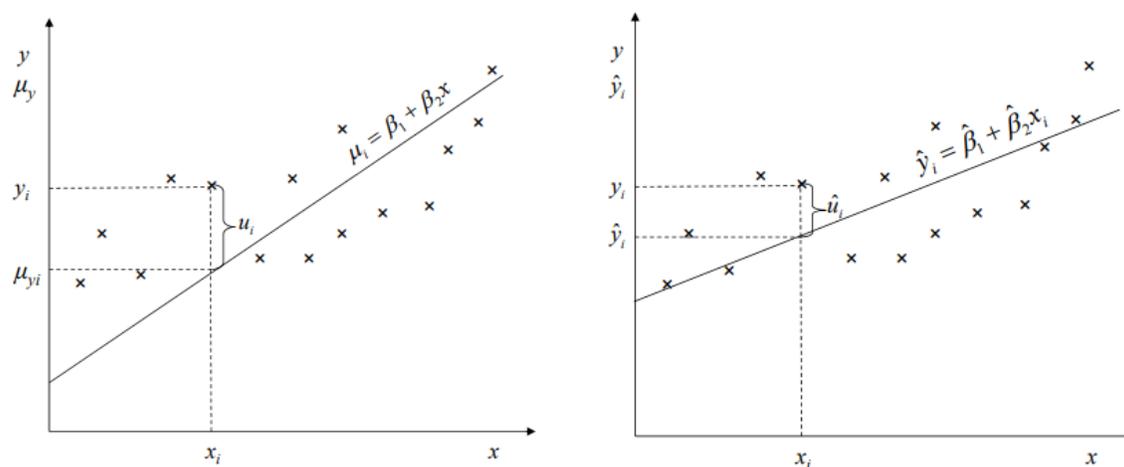
El objetivo principal (Morelisi, Modelo de regresion lineal simple, 2018) de la regresión es la determinación o estimación de  $\beta_1$  y  $\beta_2$  a partir de la información contenida en las observaciones ( $x$ ) de que se disponen. Esta estimación se puede llevar a cabo mediante mínimos cuadrados ordinarios que posteriormente se explica.

Si se designa (Morelisi, Modelo de regresion lineal simple, 2018) mediante  $\hat{\beta}_1$  y  $\hat{\beta}_2$  las estimaciones de  $\beta_1$  y  $\beta_2$ , respectivamente, la ordenada de la recta para el valor  $x_i$  vendrá dada por

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i \quad (16)$$

Según, (Morelisi, Modelo de regresion lineal simple, 2018), el problema que se tiene planteado es, hallar unos estimadores  $\hat{\beta}_1$  y  $\hat{\beta}_2$  tales que la recta que pasa por los puntos  $(\hat{x}_i, \hat{y}_i)$ , se ajuste lo mejor posible a los puntos  $(x_i, y_i)$ . Se denomina error o residuo a la diferencia entre el valor de muestra  $y_i$  y el valor ajustado  $\hat{y}_i$  como se observa en la figura 6, es decir,

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i \quad (17)$$



**Figura 6** Regresión total y de una sola muestra

Fuente: (Uriel, 2013)

- **Estimación por mínimos cuadrados**

Se utiliza este método con el fin de calcular  $\hat{\beta}_1$  y  $\hat{\beta}_2$  mediante las estimaciones de  $\beta_1$  y  $\beta_2$  que consiste en minimizar la suma de los cuadrados de los residuos, es decir,

$$\text{Min } S = \text{Min } \sum_{i=1}^n \widehat{u}_i^2 \quad (18)$$

Los estimadores (Morelisi, Modelo de regresion lineal simple, 2018) se denominan estimadores de mínimos cuadrados (MC), desde el momento en que se toman los cuadrados de los residuos, se está penalizando proporcionalmente a los residuos grandes frente a los pequeños. (Morelisi, Modelo de regresion lineal simple, 2018) siendo S una función de estimadores por lo tanto se tienen:

$$\text{Min } S = \text{Min } \sum_{i=1}^n \widehat{u}_i^2 = \text{Min } \sum_{i=1}^n (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 \quad (19)$$

Para minimizar S, se deriva parcialmente con respecto a  $\hat{\beta}_1$  y  $\hat{\beta}_2$ :

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)$$

$$\frac{\partial S}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) x_i$$

Los estimadores de MC se obtienen igualando las anteriores derivadas a cero:

$$\sum_{i=1}^n (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \quad (20)$$

$$\sum_{i=1}^n (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) x_i = 0 \quad (21)$$

Las ecuaciones (20) y (21) se denominan ecuaciones normales o condiciones de primer orden de MC. Operando las ecuaciones normales, se tiene que

$$\sum_{i=1}^n y_i = n \hat{\beta}_1 - \hat{\beta}_2 \sum_{i=1}^n x_i \quad (22)$$

$$\sum_{i=1}^n y_i x_i = \hat{\beta}_1 \sum_{i=1}^n x_i - \hat{\beta}_2 \sum_{i=1}^n x_i^2 \quad (23)$$

Dividiendo ambos miembros de (22) por n, se tiene

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{x} \quad (24)$$

Por tanto,

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} \quad (25)$$

Sustituyendo este valor de  $\hat{\beta}_1$  en la segunda ecuación normal (23), se obtiene que

$$\sum_{i=1}^n y_i x_i = (\bar{y} - \hat{\beta}_2 \bar{x}) \sum_{i=1}^n x_i + \hat{\beta}_2 \sum_{i=1}^n x_i^2$$

$$\sum_{i=1}^n y_i x_i = \bar{y} \sum_{i=1}^n x_i - \hat{\beta}_2 \bar{x} \sum_{i=1}^n x_i + \hat{\beta}_2 \sum_{i=1}^n x_i^2$$

Resolviendo para  $\hat{\beta}_2$  se tiene que :

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} \quad (26)$$

Escrito de otra forma:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (27)$$

Si se divide numerador y denominador de (27) por  $n$ , se puede ver que  $\hat{\beta}_2$  es el cociente entre la covarianza de las dos variables y la varianza de  $x$ . Por lo tanto, el signo de  $\hat{\beta}_2$  es el mismo que el signo de la covarianza. Una vez calculado  $\hat{\beta}_2$ , se puede obtener  $\hat{\beta}_1$  utilizando la ecuación (25), determinando los estimadores mediante mínimos cuadrados ordinarios (MCO) Uriel (2013).

#### 2.2.4. Modelos de regresión lineal multivariante

En la regresión lineal múltiple se utilizan más de una variable explicativa o independiente, esto ofrece la ventaja de utilizar más información en la construcción del modelo y, consecuentemente, realizar estimaciones más precisas Rojo (2007).

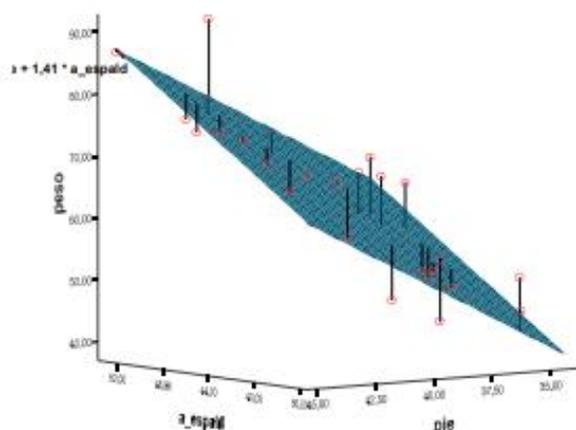
En el modelo de regresión lineal múltiple,  $y$  es la variable dependiente que es una función lineal de  $k$  regresores correspondientes a las variables independientes, si se designa  $x_2, x_3, \dots, x_k$  como variables independientes y  $u$  al error o perturbación aleatoria, el modelo de regresión lineal múltiple vendrá dado por la siguiente expresión:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u \quad (28)$$

Donde, los parámetros  $\beta_1, \beta_2, \beta_3, \dots, \beta_k$  son fijos y desconocidos. En el segundo miembro de (28) se pueden distinguir dos componentes: un componente sistemático  $\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k$  y la perturbación aleatoria  $u$ . Llamando  $u_y$  al componente sistemático, se puede escribir:

$$u_y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k \quad (29)$$

Esta ecuación es conocida como función de regresión o hiperplano poblacional. Cuando  $k=2$ , la función de regresión es específicamente una línea recta, cuando  $k=3$ , la función de regresión es específicamente un plano y, por último, cuando  $k>3$ , la función de regresión es denominada genéricamente hiperplano, que no es susceptible de ser representado físicamente (ver figura 7).



**Figura 7** Hiperplano con dos variables explicativas y una dependiente

Fuente: (Rojo, 2007)

La ecuación (29) se puede representar para una población  $n$ , donde cada una de las variables pueden quedar expresadas en forma matricial.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{21} & x_{31} & \dots & x_{k1} \\ 1 & x_{22} & x_{32} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{2n} & x_{3n} & \dots & x_{kn} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_k \end{bmatrix}$$

Donde,  $X$  es la matriz de variables independientes de dimensiones  $(k \times n)$ ,  $y$  representan un vector de variables dependientes de dimensiones  $(1 \times n)$ ,  $\beta$  representa un vector hiperplano que se

forma en función de  $X$  e  $Y$  de dimensiones  $(1 \times k)$  y  $u$  es el vector de error de  $(1 \times k)$ . Finalmente, la curva de regresión puede representarse de forma sintetizada en (30).

$$y = X\beta + u \quad (30)$$

La idea básica de la regresión consiste en estimar los parámetros  $\beta_1, \beta_2, \beta_3, \dots, \beta_k$ , a partir de una muestra dada Uriel (2013). En este trabajo la obtención de los parámetros  $\beta$  será realizado usando PCR y PLS.

### 2.2.5. Regresión Componentes Principales (PCR)

La Regresión Componentes Principales es una técnica de estimación para combatir la multicolinealidad lo que significa que una o más variables, son combinación lineal de otra Navarro (2009). La idea detrás de la regresión de componentes principales es realizar primero un análisis de componentes principales (PCA) que se especifica a continuación. En la ecuación (31) se presenta la formulación para una regresión usando PCR.

$$Y = X^* \beta^* + \varepsilon \quad (31)$$

Donde  $X$  representa la matriz de características de cada experimento de dimensiones  $(n \times k)$  (con variables independientes),  $X^*$  representa la matriz  $X$  centrada y escalada,  $Y$  de dimensiones  $(k \times 1)$  es un vector de variables dependientes estandarizadas,  $\beta^*$  es un vector de parámetros de orden  $(k \times 1)$  y  $\varepsilon$  un vector de errores aleatorios de orden  $(n \times 1)$ .

$$Z = X^*A \quad (32)$$

Siendo  $Z$  la matriz de componentes principales de  $X^*$  de dimensiones  $(n \times k)$  se usan solo los primeros  $k$  componentes principales para hacer la regresión, es decir se utilizan los componentes principales del PCA que son ortogonales dado que  $A$  es ortogonal,  $A$  es una matriz de dimensiones

( $k \times k$ ) donde la  $j$ -ésima columna es asociada al  $j$ -ésimo autovalor para lo cual se dice que  $AA^\perp = A^\perp A = 1$  se puede reescribir la ecuación (31) como:

$$Y = X^* AA^\perp \beta^* + \varepsilon \quad (33)$$

$$\gamma = A^\perp \beta^* \quad (34)$$

Si  $\gamma$  es un vector de coeficientes ( $k \times 1$ ), entonces reemplazando las ecuaciones (32) y (34) en la ecuación (31) queda expresada como:

$$Y = Z\gamma + \varepsilon \quad (35)$$

### 2.2.6. Regresión por Mínimos Cuadrados Parciales (PLS)

El modelo de regresión por mínimos cuadrados parciales es un modelo estadístico aplicable a conjuntos multivariados y consiste en obtener de forma iterativa direcciones ortogonales en los espacios de variables. Las variables independientes son estimadas como combinaciones lineales de las variables dependientes (Gaviria, 2016). La regresión por mínimos cuadrados parciales consiste en descomponer tanto la matriz de diseño  $X$  como la matriz de respuesta  $Y$  (Valdéz, 2010).

$$X = TP^T + E \quad (36)$$

Donde  $T$  de dimensiones ( $n \times a$ ) representan los “scores” (termino que puede ser traducido como “resultados”); y la matriz  $P$  de dimensiones ( $a \times n$ ) es denominada “loadings” (termino que puede ser traducidos como “cargas”). De esta manera la matriz  $X$  ( $n \times m$ ) queda descompuesta en un número de “variables latentes”, cada uno caracterizado por un vector  $t$  y un vector  $P^T$ ,  $E$  ( $n \times m$ ) es el error.

Es posible representar la matriz  $X$  por una matriz  $T$  con un número menor de columnas, esta descomposición se muestra en la ecuación (37).

$$X_{n \times m} = T_{n \times a} P_{a \times n}^T + E_{n \times m} = t(1)_{n \times 1} \cdot p^T(1)_{1 \times n} + t(2)_{n \times 1} \cdot p^T(2)_{1 \times n} + \dots + t(a)_{n \times 1} \cdot p^T(a)_{1 \times n} + E_{n \times m} \quad ; a < m \quad (37)$$

El modelo PLS se desarrolla de modo que las primeras variables latentes ( $t(1), t(2), \dots$ ) sean las más importantes para explicar el vector  $Y$  en la muestra. El número de variables latentes necesarias para explicar la matriz  $X$  es una medida de la complejidad del modelo. La matriz  $Y$  ( $n \times 1$ ) también se descompone en (38), en la cual se tiene el vector  $b$  de dimensiones ( $a \times 1$ ) (denominado “sensibilidades”) y la matriz  $T$  ( $n \times a$ ) que al igual que antes representa los “scores”.

$$Y = Tb + F \quad (38)$$

Donde  $b$  se calcula para minimizar los errores  $F$  de dimensiones ( $n \times m$ ). Los coeficientes del vector  $b$  son estimados mediante mínimos cuadrados, por lo cual el vector  $Y$  se obtiene mediante  $b$  y  $T$ .

$$\hat{Y}_{n \times 1} = T_{n \times a} b_{a \times 1} \quad (39)$$

Se toman en cuenta todas las variables latentes ( $a=m$ ), los coeficientes del vector  $b$  son idénticos a los coeficientes del modelo de regresión lineal múltiple:

$$b_{a \times 1} = \hat{\beta} = (X^T X)^{-1} X^T Y \quad (40)$$

### 2.2.7. Validación Cruzada

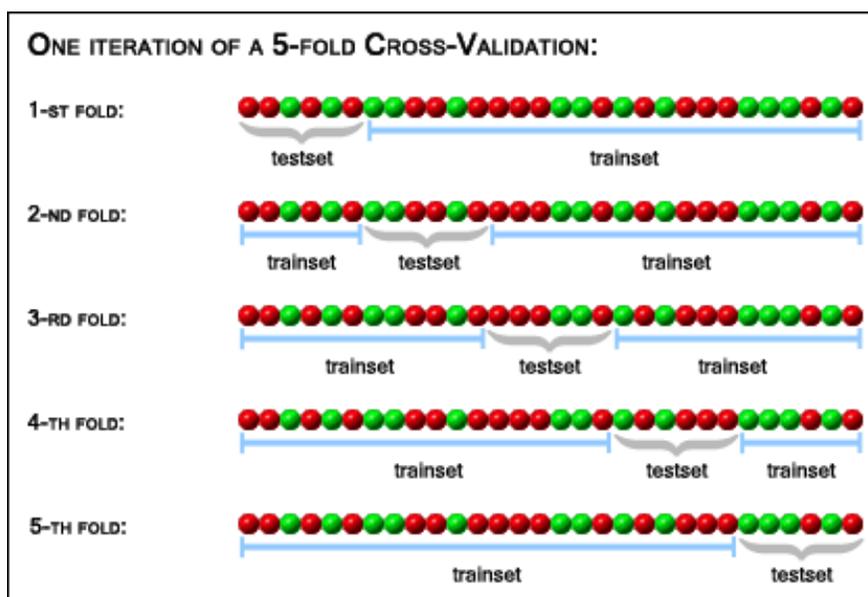
La validación cruzada es una técnica estadística de evaluación que compara algoritmos de aprendizaje para dividir los datos en dos segmentos: uno utilizado para aprender o entrenar un modelo y el otro usado para validar el modelo. Los conjuntos de entrenamiento y validación deben cruzar en rondas sucesivas de modo que cada punto tiene una posibilidad de ser validado en contra (Payam, 2008). Existen tres tipos de validación cruzada para la estimación de precisión de cuantificación:

- K-fold cross-validation
- Validación cruzada dejando uno fuera
- Validación cruzada aleatoria

Para este proyecto se utiliza el modelo de validación k-fold cross-validation, ya que se usa comúnmente en entornos donde se tiene que realizar una predicción en donde se pretende estimar la precisión con la que el modelo va a funcionar en la práctica.

### **2.2.8. K-fold Cross-Validation**

Es una técnica de validación de modelos de regresión para evaluar como los resultados de un análisis estadístico se generalizan a un conjunto de datos independientes (Joanneum, 2006). Una iteración de la validación cruzada K-fold se realiza de la siguiente manera: Primero, se genera una permutación aleatoria del conjunto de muestra y se divide en subconjuntos K de aproximadamente el mismo tamaño. De los subconjuntos K, se toma un grupo de datos para validación y el resto de ellos para entrenamiento. En otras palabras, como se observa en la figura 8, el entrenamiento y la evaluación del modelo se repiten  $k=5$  veces, para un  $k=1$  se toman los primeros 6 datos de validación y el resto son datos de entrenamiento, para un  $k=2$  se tiene los primeros datos de entrenamiento mientras que los 6 siguientes son de validación, y los datos siguientes a esos son también datos de entrenamiento, con cada uno de los subconjuntos K utilizados exactamente se realiza lo mismo hasta llegar a  $k=5$ .



**Figura 8** Validación cruzada de 5 veces con 30 muestras

Fuente: (Joanneum, 2006)

La estimación de precisión resultante depende de la permutación aleatoria que se generó al comienzo del proceso, ya que afecta la forma en que se divide el conjunto de muestra. Por lo tanto, para obtener una estimación más exacta de la precisión, tiene sentido repetir la validación cruzada varias veces y tomar el promedio de todas las estimaciones de precisión obtenidas después de cada iteración como la estimación de precisión resultante.

### 2.2.9. Figuras de mérito

Las cifras de mérito o en inglés *figures of merit* son valores que se utilizan regularmente con el propósito de calificar un determinado método y comparar sus propiedades analíticas con las provistas por otras técnicas (Olivieri, 2006), algunas de ellas son:

- Raíz del Error Cuadrático Medio (RMSECV) (usando datos de validación cruzada)
- Error Cuadrático Medio (MSECV) (Usando los datos de validación)
- Raíz del Error Cuadrático Medio (RMSEP) (usando datos de predicción)

- Error Cuadrático Medio (MSEP) (Usando los datos de predicción)
- Coeficiente de determinación  $R^2$

### **Error cuadrático medio (MSECV) y Raíz del Error Cuadrático Medio (RMSECV)**

#### **Usando los datos de validación**

En cada una de las K iteraciones del método de validación cruzada se realiza un cálculo de error, Error cuadrático medio (Mean Square Error, MSECV) y la Raíz del Error Cuadrático Medio (Root Mean Square Error, RMSECV). El resultado final se obtiene a partir de realizar la media aritmética de los K valores de errores obtenidos (43) y (44), partiendo de las fórmulas (41) y (42):

$$MSECV_i = (\hat{y}_i - y)^2 \quad (41) \quad RMSECV_i = \sqrt{(\hat{y}_i - y)^2} \quad (42)$$

Donde,  $\hat{y}_i$  es el valor predicho de la validación cruzada y  $y$  es el valor original, dando como resultado:

$$MSE = \frac{1}{K} \sum_{i=1}^K MSE_i \quad (43) \quad RMSE = \frac{1}{K} \sum_{i=1}^K RMSE_i \quad (44)$$

Es decir, se realiza la sumatoria de los K valores de error y se divide entre el valor de K.

#### **Error cuadrático medio (MSEP) y Raíz del Error Cuadrático Medio (RMSEP) Usando los datos de predicción**

El error cuadrático medio es la desviación estándar de los residuos (errores de predicción), es una medida de cuán lejos están los puntos de datos de la línea de regresión, se usa comúnmente para el análisis de regresión para verificar los resultados obtenidos. Los residuos son la diferencia entre los valores reales y los valores predichos (45). En donde  $y_i$  es el valor real y  $\hat{y}_i$  es el valor predicho (Barnston, 1992).

$$\hat{u}_i = \hat{y}_i - y_i \quad (45)$$

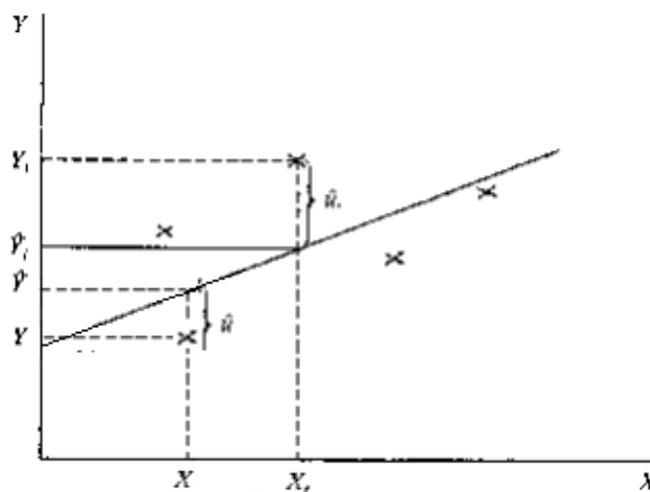
La ecuación del error cuadrático medio se puede obtener con la ecuación (46).

$$MSE = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n} \quad (46)$$

Pueden ser positivos o negativos ya que el valor predicho debajo o sobre estima el valor real. Cuadrar los residuos, promediando los cuadrados, y tomando la raíz cuadrada nos da el RMSE.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (47)$$

En la figura 9 se puede observar la resta entre los valores predicho y lo valores reales dan como resultado  $\hat{u}_i$  y  $\hat{u}$ , si se suman estos valores y se divide para los 2, entonces se tendrá como resultado el error cuadrático medio, si se saca la raíz se obtendrá en RMSE.



**Figura 9** Error Cuadrático Medio

Fuente: (Morelisi, Modelo de regresion lineal simple, 2018)

### **Coefficiente de determinación $R^2$**

Según, (Rojo, 2007) El coeficiente de determinación es la proporción de la varianza explicada por la regresión (VE) entre la varianza total (VT); de este modo, se define:

$$R^2 = \frac{VE}{VT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (48)$$

Por ser cociente de sumas de cuadrados, este coeficiente será siempre positivo. Si todos los puntos están sobre la recta de regresión, la varianza no explicada será 0, y por lo tanto:

$$R^2 = \frac{VE}{VT} = 1 = \frac{0}{VT} = 1$$

Este coeficiente es muy importante pues determina qué porcentaje de la varianza de la variable dependiente es explicado por el modelo de regresión. En general, se pueden clasificar los valores de  $R^2$  como se observa en la tabla 3. Mientras mayor sea el valor de  $R^2$  más cerca estarán los datos a la línea de regresión.

**Tabla 3**

*Clasificación de los valores de  $R^2$*

<b>Menor de 0.3</b>	<b>0.3 a 0.4</b>	<b>0.4 a 0.5</b>	<b>0.5 a 0.85</b>	<b>Mayor de 0.85</b>
Muy malo	Malo	Regular	Bueno	Sospechoso

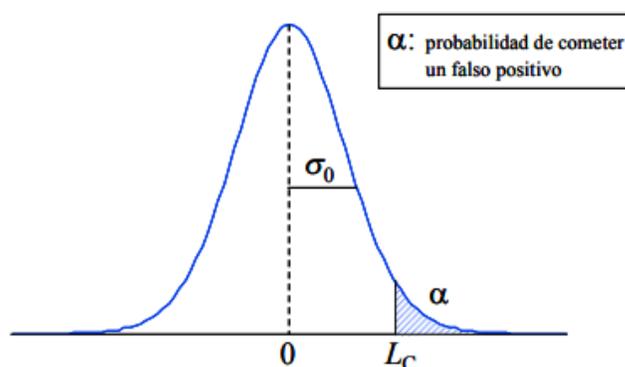
Fuente: (Rojo, 2007)

### 2.3. Límite de Detección

El límite de detección (LOD) es la capacidad de cuantificar un elemento, es un número expresado en unidades de concentración(o cantidad) que describe la concentración más baja que un analista puede determinar (American Chemical Society, 1993).

El LOD sería la concentración mínima obtenida a partir de la medida de una muestra que sería capaz de discriminar de la concentración obtenida a partir de la medida de un blanco, es decir de una muestra sin analito (blanco) presente.

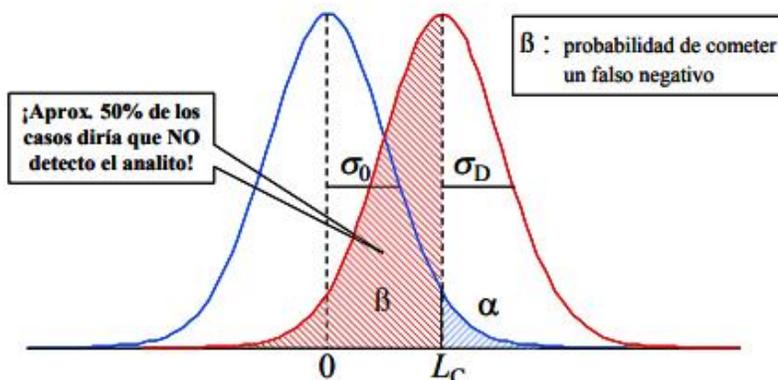
Si se analizan los blancos de una muestra, se obtendrían una distribución de valores como lo que se muestran en la figura 10 (Boqué, 2014). Los valores de concentración se distribuyen alrededor de cero con una desviación estándar  $\sigma_0$ . Quiere decir que como resultado de la medida de un blanco, y debido a los errores experimentales del método se puede obtener una concentración que no fuese cero.



**Figura 10** Distribución de valores alrededor de cero y nivel crítico  
Fuente: (Boqué, 2014)

El valor crítico,  $L_C$ , permitirá, una vez medida la muestra, tomar la decisión de si el analito se halla presente o no, si la concentración obtenida es superior a  $L_C$  entonces sin duda no corresponde a un blanco y se puede decir que el analito si está presente en la muestra.

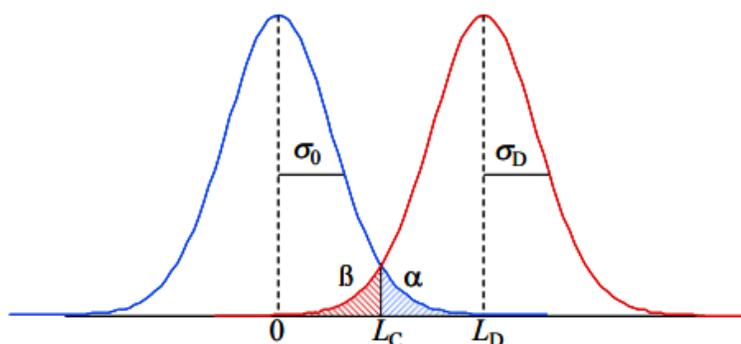
La probabilidad  $\alpha$  se denomina error de tipo I, error de primera especie, o error de falso positivo, que es el riesgo de concluir erróneamente que el analito está presente, cuando en realidad no lo está. El riesgo de concluir erróneamente que el analito no está presente cuando en realidad si lo está se expresa mediante la probabilidad  $\beta$  denominada error tipo II, error de segunda especie o, más comúnmente probabilidad de falso negativo como se observan en la figura 11.



**Figura 11** Nivel crítico y límite de detección para una probabilidad  $\alpha$

Fuente: (Boqué, 2014)

Es decisión del laboratorio la elección del valor de  $\beta$ . La única alternativa para reducir el riesgo de falsos negativos es aumentar el límite de detección, como en el ejemplo de la figura 12.



**Figura 12** Nivel crítico y límite de detección para  $\alpha=\beta$

Fuente: (Boqué, 2014)

Currie en 1968 (Currie, 1968) presentó un amplio estudio sobre hipótesis basado en la teoría del límite de detección, este estudio se basó en definir LOD en base a la hipótesis nula ( $H_0$ ), la cual prueba que el analito no está presente. Currie establece dos conceptos: un nivel crítico ( $L_C$ ) que es

el valor máximo aceptable para evitar detectar una sustancia cuando no lo es y un límite de detección ( $L_D$ ) que es la señal verdadera más pequeña que será detectada confiablemente contra una falsa conclusión de una observación en blanco.

La formulación matemática de estos conceptos están dados por (49) y (50).

$$L_C = k_\alpha \sigma_0 \quad (49)$$

$$L_D = L_C + k_\beta \sigma_D \quad (50)$$

Donde,  $k_\alpha$  y  $k_\beta$  corresponden a probabilidades de cometer un error de tipo 1 y de tipo 2 de acuerdo a la tabla de distribución t,  $\sigma_0$  es la desviación estándar de los blancos de la señal y  $\sigma_D$  es la desviación estándar la cual representa las mediciones que contiene el analito en el nivel del límite de detección.

In 1997 la Unión internacional de química pura y aplicada (IUPAC) publico una visión general para estandarizar de la determinación del límite de detección y cuantificación (Mocak, Bond, Mitchell, & Schollary, 1997). El LOD fue definido como límite de detección el cual debe reflejar el valor de la señal verdadera que es significativamente diferente de los valores de blancos de la señal. Cambiaron la formulación original (49) teniendo en cuenta un número finito de mediciones, por lo cual la nueva formulación se muestra en la ecuación (51).

$$L_D = \bar{y}_0 + t(v_0, \alpha)(1 + 1/n_0)^{1/2} s_0 \quad (51)$$

Donde,  $\bar{y}_0$  y  $s_0$  son la media y la desviación estándar de las muestras de blancos,  $t(v_0, \alpha)$  es el valor crítico de la distribución t con  $v_0$  grados de libertad los cuales son calculados como el número de blancos menos 1,  $(1 + 1/n_0)^{1/2}$  es una corrección de la incertidumbre de la determinación de  $\bar{y}_0$  y  $s_0$ .

## CAPÍTULO III

### METODOLOGÍA EXPERIMENTAL

En el presente capítulo se explicará el proceso realizado para la recolección de datos de cada uno de los experimentos, también los parámetros para determinar las concentraciones de cada sustancia y el acondicionamiento y pre procesamiento de las señales.

#### 3.1 Experimentos

En esta parte del proyecto se realizó una experimentación aleatoria para evitar que los datos generen un patrón o “efecto memoria” que pudiera influir en los resultados y generación de los modelos de calibración multivariante. En un inicio se tomaron datos de prueba de TNT y pólvora en estado puro con las cantidades de 0g, 0.1g, 0.2g, 0.3g, 0.4g, 0.5g, 0.6g, 0.7g, 0.8g, 0.9g, 1g, 2g, centrándose la toma de datos en concentraciones bajas, ya que el objetivo es determinar el límite de detección del dispositivo y se requiere tener suficientes mediciones con concentraciones pequeñas, posteriormente se añadieron 2 medidas más, teniendo un total de 12 concentraciones posibles (0g, 0.1g, 0.2g, 0.3g, 0.4g, 0.5g, 0.7g, 1g, 1.5g, 2g, 2.5g, 3g).

Para realizar la base de datos, se tomó en cuenta el tiempo mínimo de medición de cada experimento, como se observa en la figura 13 el tiempo en el cual la nariz respira y limpia ( $t_m$ ) se puede calcular mediante la fórmula (27) en donde,  $t_2$  es el tiempo final y  $t_1$  es el tiempo inicial.

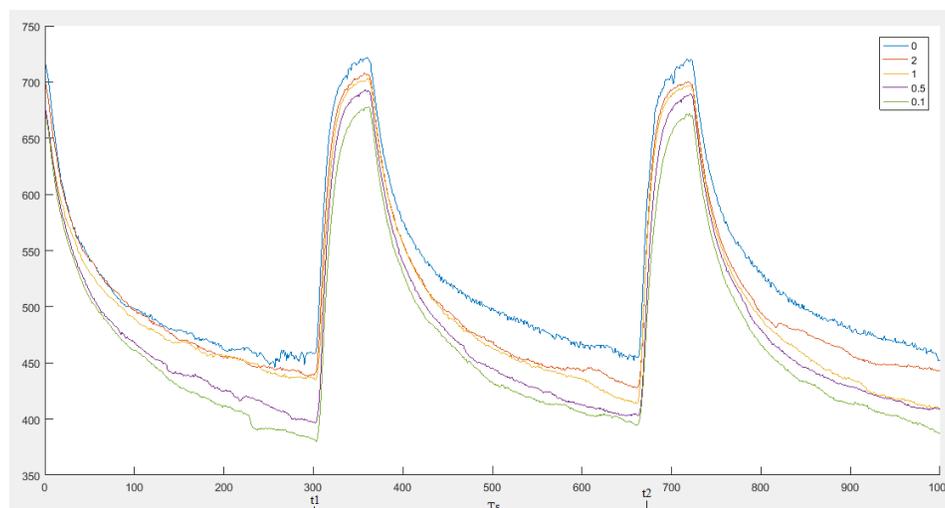
$$t_m = t_2 - t_1 \quad (52)$$

$$t_m = 650 - 300 = 350 \text{ s}$$

Se decidió realizar 3 réplicas, es decir tomar 3 veces el tiempo  $t_m$ , dando como resultado el tiempo total de medición de cada experimento (53).

$$t_T = 3t_m = 3 * 350 = 1050 \text{ s} \quad (53)$$

A  $t_T$  se sumaron 300s del inicio en los cuales la nariz se encontraba limpiando, por lo cual se realizaron medidas de 1350s por cada experimento, o en minutos da un tiempo aproximadamente de 22.5 minutos.



**Figura 13** Tiempo de medición.

Debido a que el jabón y la pasta dental son las sustancias comunes con las cuales los traficantes mezclan los explosivos se decidió utilizarlas para la realización de las mezclas haciendo mediciones con las sustancias explosivas en estados puros y combinados. La razón por la cual se utilizaron estas sustancias es porque las personas al viajar dentro y fuera del país llevan consigo útiles de aseo más que comida.

Se obtuvo un total de 149 experimentos: 50 con TNT en estado puro, 49 con pólvora pura, 15 con TNT y Jabón, 15 con TNT y pasta dental, 10 de pólvora con jabón y 10 con pólvora y pasta dental, todo esto en un periodo de 9 días , realizando 15 experimentos por día como en el siguiente apartado se explica.

### 3.1.1. Toma de datos

Al inicio de cada día antes de realizar la toma de datos se realiza un proceso de limpieza del instrumento, de aproximadamente 7 minutos, que tiene dos propósitos fundamentales: uno de ellos limpiar las mangueras de muestras residuales que pudieran haber quedado de procedimientos anteriores y el otro objetivo es conseguir que la temperatura de la cámara se estabilice a 29°C, para existan menos perturbaciones al momento de realizar las medidas.

Para cada experimento se introducen 2 sustancias en los contenedores del sistema de dopaje de la nariz electrónica, una de ellas es 2ml de alcohol que es la sustancias dopante, también se introduce la sustancia explosiva ya sea pura o mezclada de acuerdo a la concentración que se requiera medir. Además se tiene que tomar en cuenta que entre un experimento y otro se realiza otra limpieza del sistema de aproximadamente 7 minutos.

Para la toma de datos se planificó medir las diferentes concentraciones de manera aleatoria en el orden que se observa en las tablas 4 y 5, realizándose 15 mediciones por día con sustancias en estado puro en el siguiente orden POLVORA1, TNT1, POLVORA2, TNT2, POLVORA3 y TNT3, para concentraciones de 0g hasta 2g, aproximadamente desde las 8:30h hasta las 15:30h, de lunes a viernes, para medir las sustancias mezcladas POL Y 1g JABON, TNT Y 1g JABON, POL Y 1g PASTA, TNT Y 1g PASTA con concentraciones de 0g a 2g se utilizaron aproximadamente 2 días y medio, desde las 8:30h hasta las 17:30h, teniendo 20 mediciones por día aproximadamente, el ultimo día se realizó la toma de datos de los 10 últimos valores de TNT Y 1g PASTA y también se midieron 2.5g y 3g de sustancias solo puras, es decir de POLVORA1, TNT1, POLVORA2, TNT2, POLVORA3, TNT3.

Para este proceso de toma de datos se utilizó una balanza electrónica de marca CAMRY el grado de precisión del instrumento fue de 0.01g, por lo que se tuvieron medidas muy exactas al momento de realizar los experimentos.

**Tabla 4**

*Tabla de toma de datos por día de pólvora en estado puro y pólvora mezclada*

	<b>Primer día</b>	<b>Tercer día</b>	<b>Quinto día</b>	<b>Séptimo día</b>	<b>Octavo día</b>
	<b>Pólvora 1</b>	<b>Pólvora 2</b>	<b>Pólvora 3</b>	<b>Pol y 1g jabón</b>	<b>Pol y 1g pasta</b>
<b>Orden</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>
1°	0	1.5	1	1	Pasta
2°	0	0	0.1	Jabón	1
3°	2	0.5	0.2	0.5	0
4°	0.1	0.1	0	0	0.5
5°	0.3	0	0.5	0.1	0.1
6°	0	0.7	0	0.7	Pasta
7°	0	0.4	2	Jabón	Pasta
8°	0.7	0	0	Jabón	0.3
9°	1	0	0.3	0	0.7
10°	0.4	1	0	0.3	0
11°	1.5	0	1.5	0	0
12°	0	0.2	0	N/A	N/A
13°	0	0	0.4	N/A	N/A
14°	0.2	0.3	0.7	N/A	N/A
15°	0.5	2	0	N/A	N/A
16°	3	2.5	3	N/A	N/A
17°	2.5	3	2.5	N/A	N/A

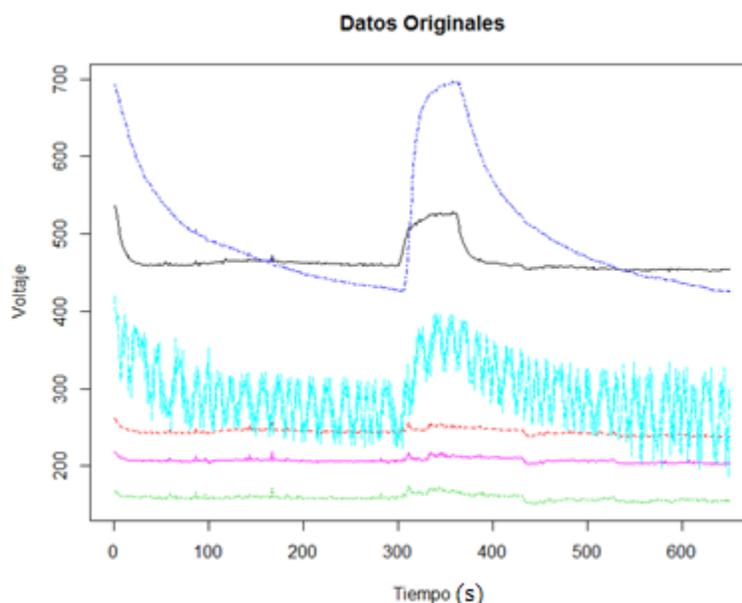
**Tabla 5**

*Tabla de toma de datos por día de TNT en estado puro y TNT mezclado*

	Segundo día	Cuarto día	Sexto día	Séptimo día	Octavo día
	TNT 1	TNT 2	TNT 3	TNT y 1g	TNT y 1g
Orden	Cant. (g)	Cant. (g)	Cant. (g)	jabón Cant. (g)	de pasta Cant. (g)
1°	0.7	0.7	0.5	0	2
2°	0	1	1	Jabón	Pasta
3°	0.3	1.5	0.7	0.5	0
4°	1	0	2	1	0.7
5°	0	0	0.3	0	1
6°	0.1	0	0	0.7	0.2
7°	2	2	0.2	1.5	0.5
8°	0	0.4	0	0.2	Pasta
9°	0	0.1	0	Jabón	0.1
10°	1.5	0	0.1	0.1	1.5
11°	0.4	0.2	0	2	Pasta
12°	0	0	0	0.4	0.4
13°	0.2	0.5	1.5	Javon	0
14°	0.5	0	0	0.3	0.3
15°	0	0.3	0.4	0	0
16°	3	3	2.5	N/A	N/A
17°	2.5	2.5	3	N/A	N/A

### 3.2. Preprocesamiento de la señal

Luego de la experimentación, se realiza un preprocesamiento de las señales, debido a que estas no tienen la misma línea base, es decir que ninguna de ellas empieza en cero en el eje vertical tal como se observa en la figura 14. Se realiza este proceso con el objetivo de comparar cada uno de los picos de la señal a diferentes concentraciones, además es necesario realizar un proceso de filtrado para suavizar la señal y eliminar el ruido que estas puedan tener, al final es necesario la concatenación de los datos para ser utilizados en los modelos de calibración multivariante. En la figura 14 se observa la señal obtenida de cada uno de los seis sensores de un experimento.



**Figura 14** Señal original.

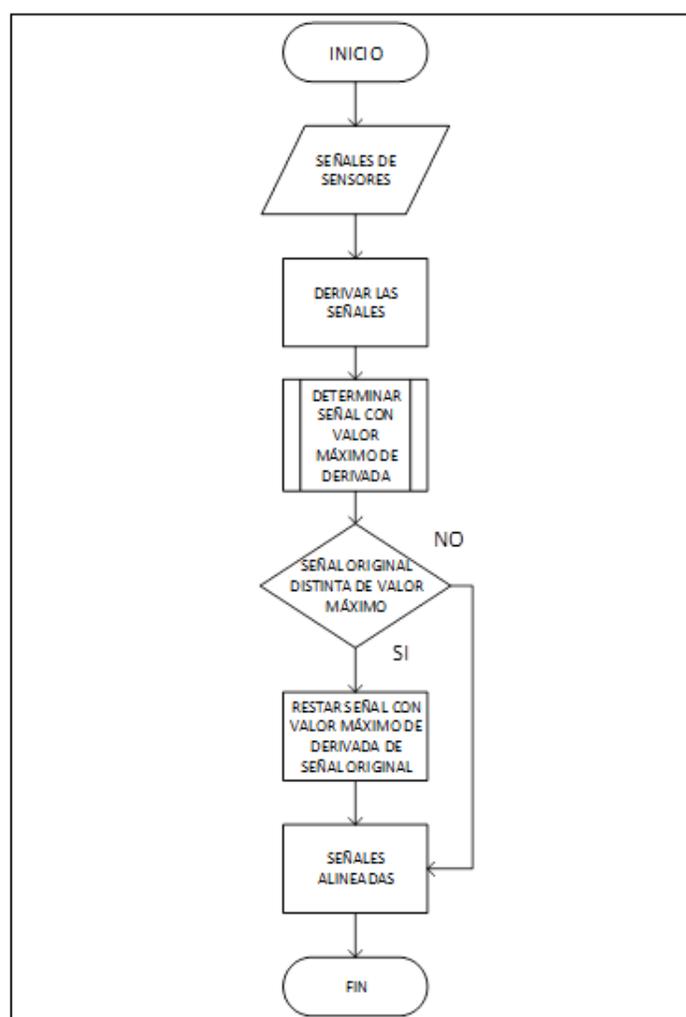
Para realizar este algoritmo se trabajó con listas de matrices, haciendo más sencillo el uso de estas, se utilizó el software R studio ya que es un software libre, que tiene varias funciones útiles para el preprocesamiento de las señales.

### **3.2.1. Corrección de líneas base de cada uno de los experimentos**

En la siguiente fase del proyecto se estableció una misma línea base para todos los experimentos, ya que al realizar los modelos de calibración multivariantes se comparan cada uno de los picos de la señal a diferentes concentraciones, y estas señales deben ser comparables entre ellas de lo contrario existirán errores en la fase de calibración, que afectará en la estimación del límite de detección que el dispositivo pueda tener.

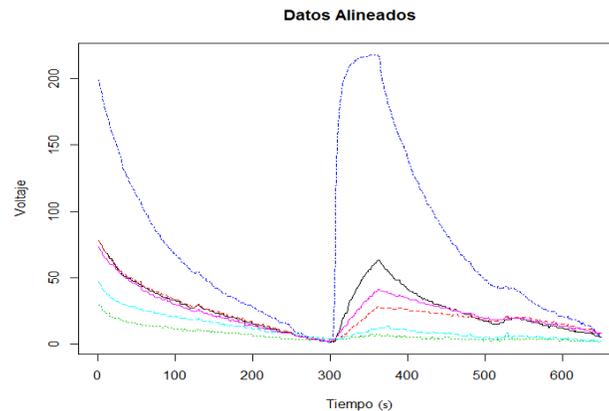
Para esto se realizó un algoritmo en el cual se utiliza la derivada de la señal, como sabemos la derivada de una señal puede representar el punto en el cual la señal crece para un tiempo anterior y decrece en un tiempo posterior o viceversa, obteniéndose puntos máximos y mínimos de una

señal o función, es este caso se utilizó para determinar el punto máximo de las señales y posteriormente determinar la señal que tiene el mayor valor de punto máximo, el proceso se muestra en la figura 15.



**Figura 15** Flujograma del algoritmo de alineación de líneas base

En la figura 16 se pueden observar el resultado de realizar la corrección de las líneas base, las señales de los diferentes sensores todas a un mismo nivel. El eje y está en función del voltaje (V) y el eje x en función del tiempo (s), pero como se observa existe distorsión en algunas señales.



**Figura 16** Señales con líneas base alineadas

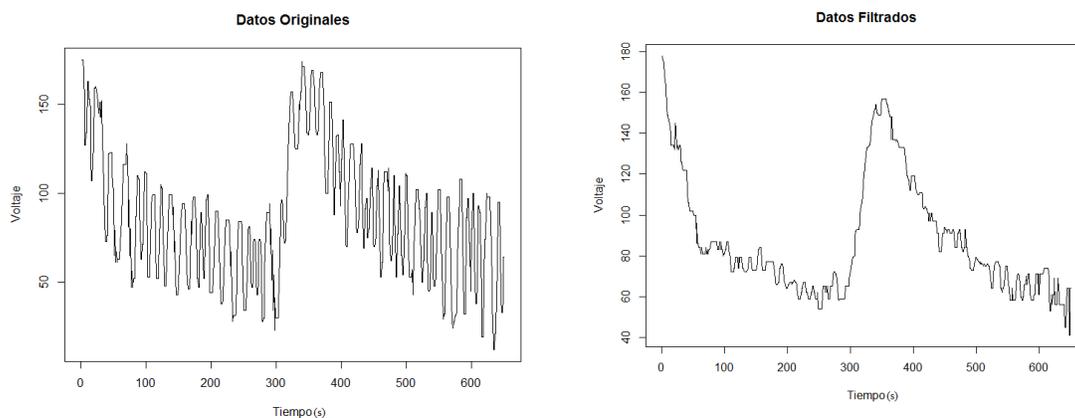
### 3.2.2. Filtrado de señales

Se filtraron los datos para obtener señales más puras, eliminando el ruido de ellas, ya que como se observa en la figura 16 existen perturbaciones en el sistema, esto puede darse debido a señales externas, por más que la matriz de sensores se encuentre aislada con una caja metálica herméticamente cerrada existe ruido en el ambiente que puede influir al momento de realizar las medidas. Otra de las razones puede ser por el ruido térmico que los componentes como resistores o transistores de la placa PCB generan y una tercera razón puede ser por las vibraciones que produce el dispositivo cuando se realiza el proceso de limpieza y cuando respira la nariz electrónica.

Lo que se quiere es eliminar el ruido de alta frecuencia, se utilizó un filtro de mediana siendo un método fácil de implementar y calcular.

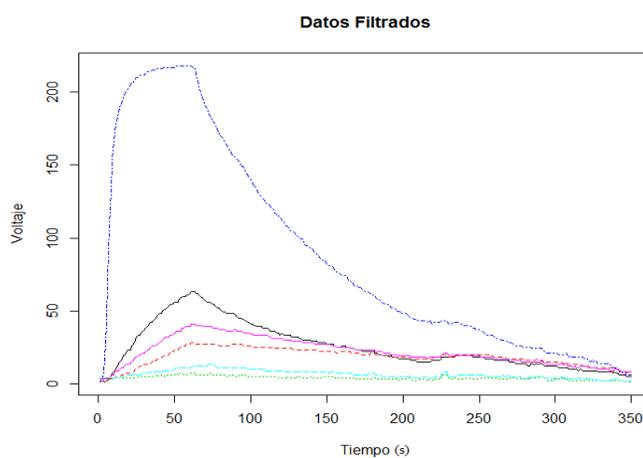
El filtro de mediana es un método que consiste en tomar los últimos  $k$  valores recibidos y se calcula la media de ellos, el resultado es la señal suavizada, únicamente se requiere sumar los  $k$  elementos de la ventana y dividir para  $k$ , el ancho de la ventana debe ser impar, para un alisamiento robusto “mínimo” que elimine a los valores atípicos (Llamas, 2017). Una de las ventajas de este

tipo de filtros es que no solo elimina el ruido sino también variaciones de la señal. En la figura 17a se observa la señal sin ser filtrada, mientras que en la figura 17b se presentan los datos ya filtrados, para ello se utilizó una ventana de  $k=5$ .



**Figura 17** Filtro de mediana con una ventana de  $k=5$

A demás se eliminaron los valores en los cuales la nariz realiza el proceso de limpieza, es decir todos los valores anteriores a 300s, para lo cual las señales filtradas de cada uno de los sensores se muestran en la figura 18.

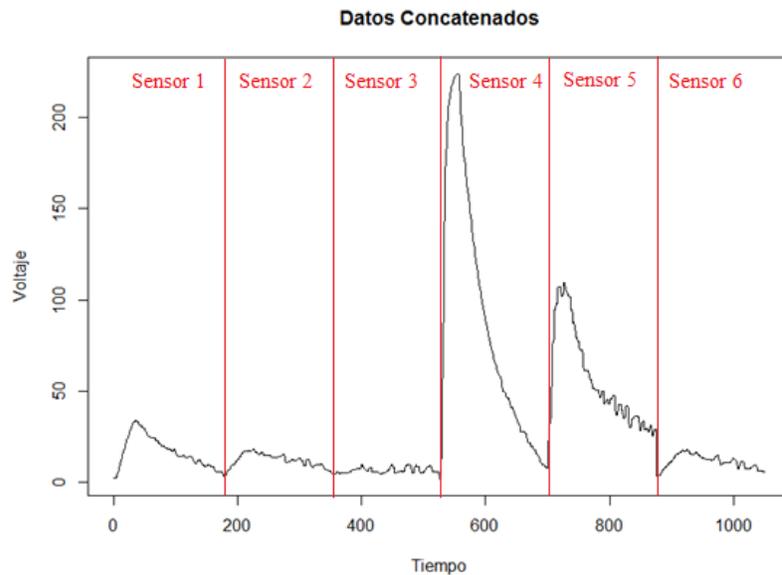


**Figura 18** Datos filtrados de un solo experimento en una sola matriz con 6 sensores

### 3.2.3. Concatenación de vectores

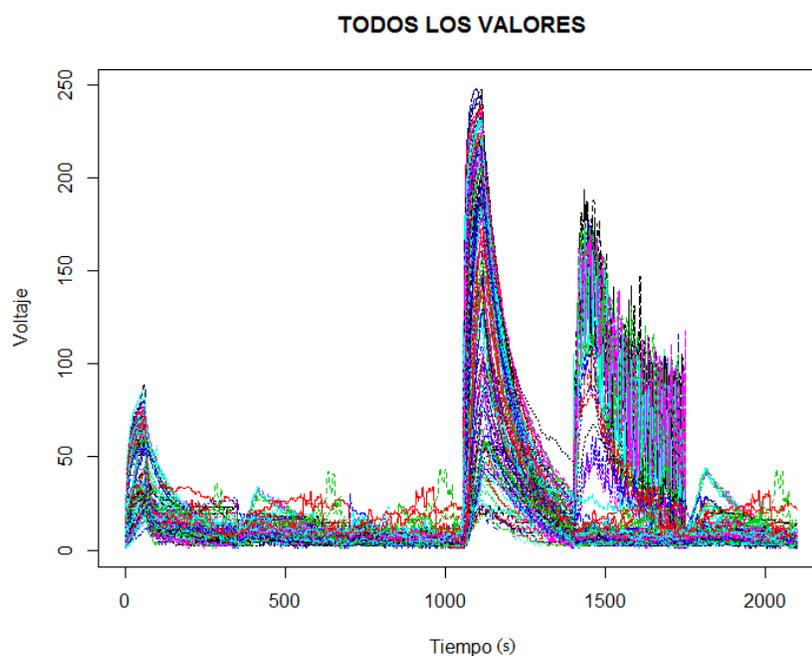
Para realizar los modelos de calibración multivariante es necesario tener los datos ordenados, debido a que se tiene la respuesta de 6 sensores al mismo tiempo, estos datos deberían estar ordenados por sensores y características, para ello se realiza la concatenación de los datos.

Teniendo originalmente 149 experimentos (matrices de  $N$  sensores con  $P$  muestras), es decir se tenían matrices de  $P \times N$ , al concatenar se crearon vectores de cada una de las matrices se obtuvieron vectores de  $1 \times (N \times P)$ , estos vectores se volvieron a guardar en una nueva matriz  $C$  de dimensiones  $149 \times M$  siendo  $M = N \times P$ , donde  $M$  son las características de los 149 experimentos. En la figura 19 se puede observar cómo queda un experimento concatenado donde existen 6 sensores.



**Figura 19** Datos concatenados con 2100 características de 6 sensores de un experimento

Como se tienen 149 experimentos al crear la matriz total de datos se obtienen las señales que se observan en la Figura 20.



**Figura 20** Datos concatenados de 149 experimentos con 6 sensores.

### 3.3. Organización de datos para creación de modelos

En este trabajo se abordó el problema de calibración creando cuatro modelos, una de TNT en estado puro, la cual tenía una dimensión de  $50 \times 1750$  siendo 50 el número de experimentos y 1750 las características de los 5 sensores, para pólvora pura se creó una matriz de  $49 \times 1750$ , para TNT mezclado una matriz de  $30 \times 1750$  y por ultimo para pólvora mezclada una matriz de dimensiones  $20 \times 1750$ .

Para crear y validar el modelo se dividieron los datos de cada matriz mencionada anteriormente: para TNT se tienen 50 experimentos de los cuales aleatoriamente se tomaron 35 datos para entrenamiento, mediante la técnica de validación cruzada para determinar el número de componentes principales, con estos mismos datos se entrenaron los modelos de regresión lineal como se observa en la tabla 6 que son los recuadros de color celeste y para validación de los

modelos se utilizaron los 15 restantes que son los recuadros que no se encuentran pintados incluidos los valores de ceros con asteriscos. Además para la realización del límite de detección se utilizaron 10 muestras de 0g de los datos de validación que en la tabla se encuentran marcadas con un asterisco, cabe mencionar que del segundo experimento se eliminó un dato 0.4g, ya que la curva de este era distinta a la de los otros debido a un error experimental que en el capítulo IV se profundiza.

**Tabla 6**

*Tabla de toma de datos de TNT utilizados en la VC*

	TNT 1	TNT 2	TNT 3
Orden	Cant. (g)	Cant. (g)	Cant. (g)
1°	0	0	0*
2°	0	0	0*
3°	0	0	0*
4°	0	0	0*
5°	0*	0*	0*
6°	0*	0*	0*
7°	3	3	3
8°	2.5	2.5	2.5
9°	2	2	2
10°	1.5	1.5	1.5
11°	1	1	1
12°	0.7	0.7	0.7
13°	0.5	0.5	0.5
14°	0.4	0.4	0.4
15°	0.3	0.3	0.3
16°	0.2	-	0.2
17°	0.1	0.1	0.1

Del mismo modo que para TNT se realizó para pólvora, ya que el número de experimentos de pólvora fue de 49 debido ya que del segundo experimento se eliminaron los datos de 0.4g y 0.2g porque existió un error en la experimentación, como se observa la tabla 7 los datos de color celeste son utilizados para entrenamiento, los de color blanco son utilizados para validación de los modelos y los que tienen un asterisco utilizados para los cálculos del límite de detección.

**Tabla 7**

*Tabla de toma de datos de pólvora utilizados en la VC*

	<b>Pólvora 1</b>	<b>Pólvora 2</b>	<b>Pólvora 3</b>
<b>Orden</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>
1°	0	0	0*
2°	0	0	0*
3°	0	0	0*
4°	0	0	0*
5°	0*	0*	0*
6°	0*	0*	0*
7°	3	3	3
8°	2.5	2.5	2.5
9°	2	2	2
10°	1.5	1.5	1.5
11°	1	1	1
12°	0.7	0.7	0.7
13°	0.5	0.5	0.5
14°	0.4	-	0.4
15°	0.3	0.3	0.3
16°	0.2	-	0.2
17°	0.1	0.1	0.1

Para TNT y TNT mezclado se determinaron los datos que son utilizados en cada uno de los procesos. Como se observa en la tabla 8 los datos de color celeste se utilizaron para entrenamiento de los modelos de regresión (43 experimentos), los de color blanco para validación (37 experimentos) y los ceros que tienen un asterisco para el cálculo del límite de detección. Siendo un total de 80 experimentos.

**Tabla 8**

*Tabla de toma de datos de TNT en estado puro y TNT mezclado*

	<b>TNT 1</b>	<b>TNT 2</b>	<b>TNT 3</b>	<b>TNT Y 1g</b>	<b>TNT Y 1g</b>
<b>Orden</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>	<b>Jabón Cant. (g)</b>	<b>Pasta Cant. (g)</b>
1°	0	0	0	0*	0*
2°	0	0	0	0*	0*
3°	0	0	0	0*	0*
4°	0	0	0	Jabón	Pasta
5°	0	0	0	Jabón	Pasta



6°	0	0	0	Jabón	Pasta
7°	3	3	3	2	2
8°	2.5	2.5	2.5	1.5	1.5
9°	2	2	2	1	1
10°	1.5	1.5	1.5	0.7	0.7
11°	1	1	1	0.5	0.5
12°	0.7	0.7	0.7	0.4	0.4
13°	0.5	0.5	0.5	0.3	0.3
14°	0.4	-	0.4	0.2	0.2
15°	0.3	0.3	0.3	0.1	0.1
16°	0.2	0.2	0.2	N/A	N/A
17°	0.1	0.1	0.1	N/A	N/A

En la tabla 9 se observa la distribución de los datos para la validación cruzada y entrenamiento de los modelos utilizándose 37 experimentos que se encuentran pintados de color celeste, para la validación con los modelos multivariantes se tomaron los datos que no se encuentran pintados siendo 32 experimentos y para el límite de detección se utilizaron los ceros que tienen asteriscos, teniendo un total de 69 datos.

**Tabla 9**

*Tabla de toma de datos de pólvora en estado puro y pólvora mezclada*

	<b>Pólvora 1</b>	<b>Pólvora 2</b>	<b>Pólvora 3</b>	<b>Pólvora 1g jabón</b>	<b>Pólvora 1g pasta</b>
<b>Ord.</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>	<b>Cant. (g)</b>
1°	0	0	0	0*	0*
2°	0	0	0	0*	0*
3°	0	0	0	0*	0*
4°	0	0	0	Jabón	Pasta
5°	0	0	0	Jabón	Pasta
6°	0	0	0	Jabón	Pasta
7°	3	3	3	0.7	0.7
8°	2.5	2.5	2.5	0.5	0.5
9°	2	2	2	0.3	0.3
10°	1.5	1.5	1.5	0.1	0.1
11°	1	1	1	N/A	N/A
12°	0.7	0.7	0.7	N/A	N/A
13°	0.5	0.5	0.5	N/A	N/A
14°	0.4	-	0.4	N/A	N/A
15°	0.3	0.3	0.3	N/A	N/A
16°	0.2	-	0.2	N/A	N/A
17°	0.1	0.1	0.1	N/A	N/A

## CAPÍTULO IV

### RESULTADOS

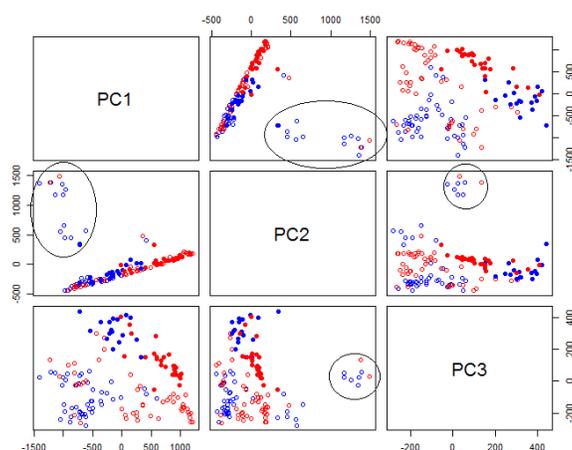
En este capítulo se presentan los resultados obtenidos por los modelos de calibración multivariante y el límite de detección para cada sustancia. También se muestra una discusión en donde se analizan los resultados y se realiza un análisis comparativo de los modelos de calibración multivariante escogidos para determinar el límite de detección y la aplicabilidad que tiene el software R para estos propósitos.

#### 4.1. Análisis de Componentes Principales (PCA)

Se realizó el análisis de componentes principales para observar la separación de las sustancias, además de corregir errores que podrían afectar al momento de desarrollar los modelos multivariantes de regresión de componentes principales y de regresión de mínimos cuadrados parciales, y que por lo tanto podrían influir en los cálculos del límite de detección de cada sustancia.

Para realizar el análisis de componentes principales, se utilizó la matriz generada de datos concatenados que se determinó en el capítulo anterior la cual contiene los 149 experimentos y las características de los 6 sensores (ver figura 20).

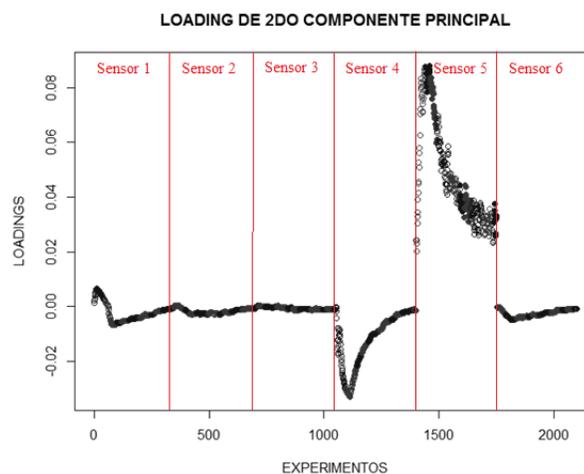
En la figura 17 se observa el PC1 vs PC2 donde se aprecia que existen datos separados del conjunto total los mismos que se encuentran encerrados en una circunferencia. Además, se observa que no hay una separación de las clases en función de las sustancias explosivas analizadas; del mismo modo ocurre cuando se comparan el PC2 vs PC3. Por lo cual se deben analizar los Loading del PC2.



**Figura 21** PCA de todos los datos sin suprimir el quinto sensor.

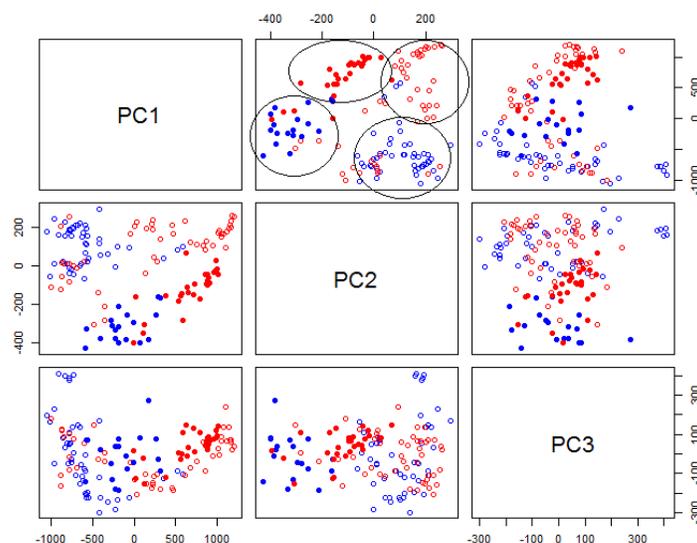
Siendo los círculos de color rojo muestras de TNT, los círculos de color azul muestras de pólvora, y los círculos pintados las mezclas con cada uno de los elementos explosivos.

Al analizar los Loading del PC2 se observa que la respuesta del quinto sensor es distinta a la de los demás, donde cabe suponer que hubo un error en este sensor al momento de realizar las mediciones.



**Figura 22** Loading del segundo componente principal sin eliminar el quinto sensor

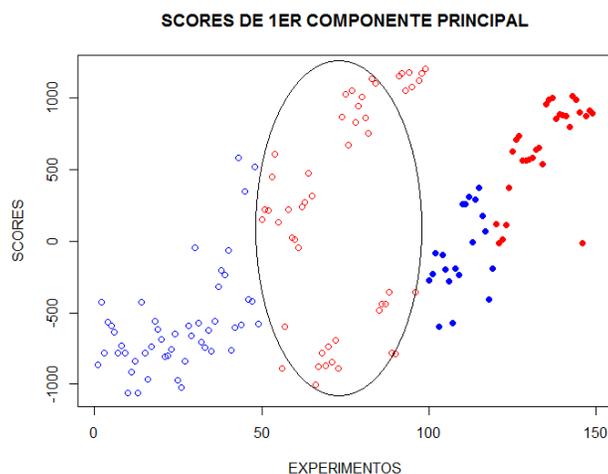
Luego de suprimir las características asociadas al sensor cinco de toda la base de datos, se realiza un nuevo modelo PCA dando como resultado una mejor separación de las clases, tal como se observa en la figura 23. En el PC2 se realiza esta separación sustancias y existe información discriminante entre ellos.



**Figura 23** PCA de todos los datos suprimiendo el quinto sensor.

Siendo los círculos de color rojo muestras de TNT, los círculos de color azul muestras de pólvora, y los círculos pintados las mezclas con cada uno de los elementos explosivos.

Analizando, los scores del primer componente principal se observa una separación de clases (ver figura 24). Existen muestras de la misma clase TNT, pólvora, las mezclas de TNT y pólvora que presentan variabilidad diferente, lo cual refleja una deriva instrumental debido a un factor día, pueden darse estos problemas por diferentes razones como los cambios de temperatura ambiente, sobrecalentamiento de elementos electrónicos, una incorrecta limpieza del dispositivo, o un problema en el sistema de dopaje automático.

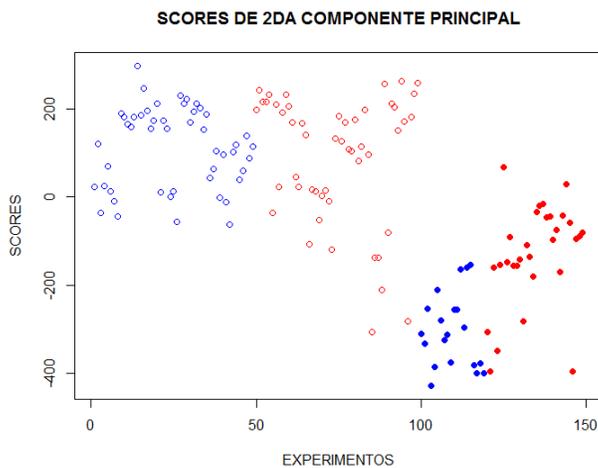


**Figura 24** Scores del primer componente principal

de las muestras sin el quinto sensor.

Siendo los círculos de color rojo muestras de TNT, los círculos de color azul muestras de pólvora, y los círculos pintados las mezclas con cada uno de los elementos explosivos.

Al observar la segunda componente principal, la información con sustancias puras, en este caso existe variabilidad y dispersión de clases.



**Figura 25** Scores del primer componente principal

de las muestras sin el quinto sensor.

Siendo los círculos de color rojo muestras de TNT, los círculos de color azul muestras de pólvora, y los círculos pintados las mezclas con cada uno de los elementos explosivos.

## 4.2. Modelo de Calibración

### 4.2.1. Determinación del número de componentes principales (PC)

Mediante la técnica de validación cruzada se determinó el número de componentes principales  $n$  y variables latentes cuyo resultado se muestra en la tabla 10. Para lo cual se estableció un número de iteraciones de  $k=10$ , variando el número de componentes  $n$  de 1 a 5, obteniéndose el menor valor de RMSECV, para determinar el número de componentes principales de cada modelo.

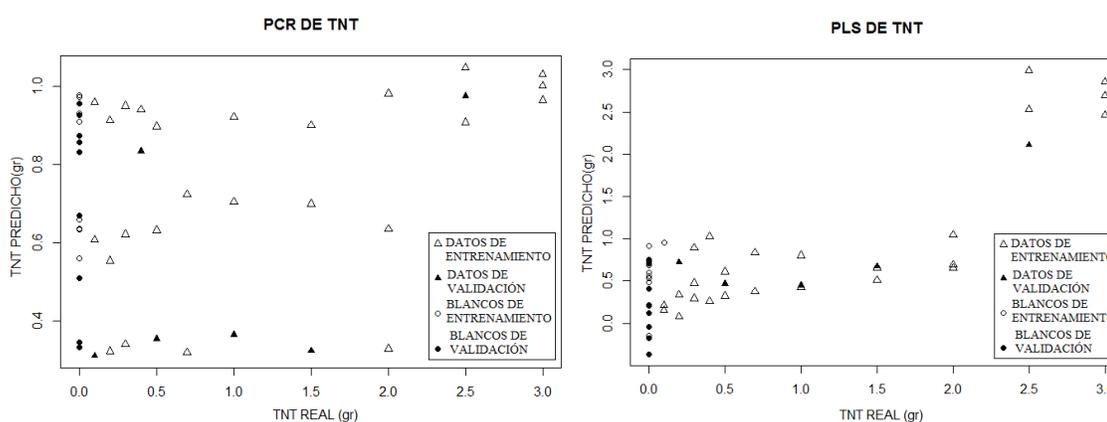
**Tabla 10**

*Número de componentes principales de cada sustancia con los 2 modelos*

Modelos /Sustancias	n=1	n=2	n=3	n=4	n=5
	RMSECV	RMSECV	RMSECV	RMSECV	RMSECV
PCR TNT	<b>1,04</b>	1,08	1,08	1,33	1,16
PLS TNT	1,46	1,21	1,39	1,35	<b>0,90</b>
PCR PÓLVORA	1,12	<b>0,87</b>	1,09	1,14	0,89
PLS PÓLVORA	1,20	<b>0,78</b>	1,09	1,16	1,01
PCR TNT MEZCLADO	1,26	1,28	1,31	<b>1,25</b>	1,27
PLS TNT MEZCLADO	0,81	0,98	1,52	1,19	<b>0,62</b>
PCR PÓLVORA MEZCLADA	<b>0,80</b>	0,90	0,96	0,91	0,86
PLS PÓLVORA MEZCLADA	1,02	<b>0,70</b>	1,01	1,02	0,83

#### 4.2.2. Regresión de Componentes Principales (PCR) y Regresión por mínimos cuadrados parciales (PLS) para TNT

Se utilizó la regresión de componentes principales con un  $n=1$  y la regresión por mínimos cuadrados parciales con un  $n=5$ , cuya respuesta se observa en la figura 26. En donde los triángulos blancos representan los valores utilizados para entrenamiento, los triángulos negros son los datos de validación, los círculos blancos representan los ceros de entrenamiento y los círculos negros representan los ceros de validación.



**Figura 26** Modelos multivariantes para TNT.

Al realizar los modelos de calibración se obtuvieron las cifras de mérito que se observan en la tabla 11, en donde se puede notar que para esta sustancia el modelo que mejor funciona es el modelo de regresión por mínimos cuadrados parciales, obteniéndose un valor mayor de error cuadrático medio y menor de coeficiente de determinación .

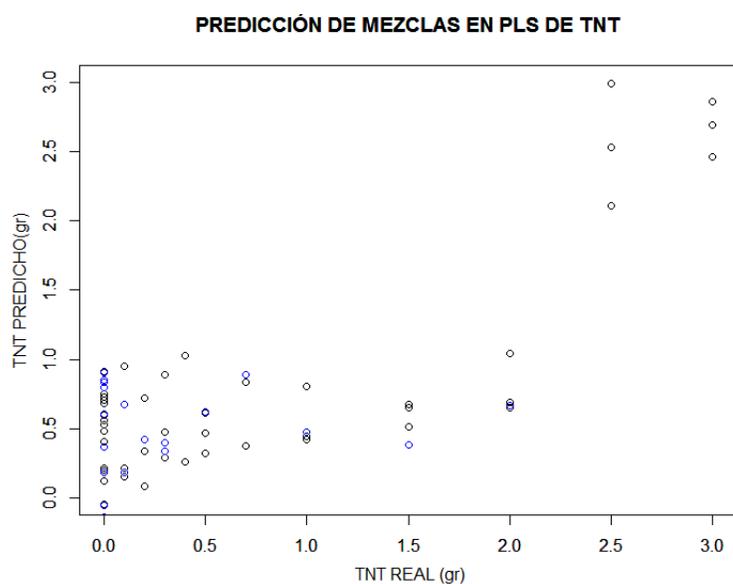
**Tabla 11**

*Tabla de cifras de mérito para la sustancia TNT con los dos modelos*

MODELO	n	MSEP	RMSEP	R <sup>2</sup>	Calificación
PCR	1	0.88	0.94	0.10	Muy malo
PLS	5	0.26	0.51	0.73	Bueno

De acuerdo a los datos presentados en la tabla 11 y en la figura 26 el método que mejor predicción realiza frente a esta sustancia es el modelo PLS.

Además, se realizó la predicción de las mezclas de jabón y pasta dental con TNT en el modelo que mejor comportamiento presentó, es decir el modelo PLS, en la figura 27 se observan los datos utilizados para la validación que son los círculos negros, y los datos de predichos de las mezclas que están representados por los círculos azules, el modelo tuvo un error cuadrático medio de 0.49 para estas sustancias mezcladas.



**Figura 27** Mezclas proyectadas en el modelo PLS de TNT

### 4.2.3. Límite de detección con TNT (LD)

El cálculo del límite de detección que tuvo el prototipo nariz electrónica con la sustancia explosiva TNT se lo realizó utilizando 10 muestras de 0gr. de TNT y se utilizó el valor de 2.262 de la tabla de distribución T, obteniéndose los valores presentados en la tabla 12.

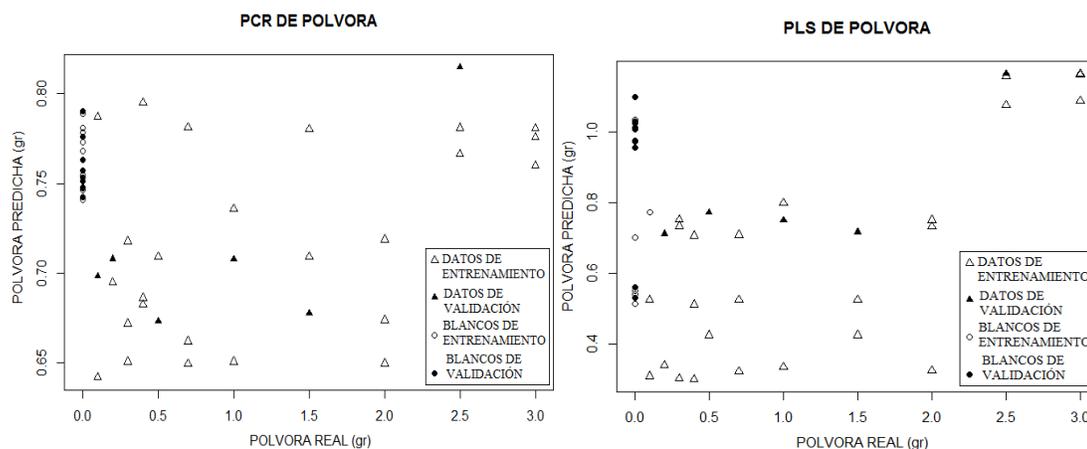
**Tabla 12**

*Límite de detección para TNT utilizando PCR y PLS*

<b>SUSTANCIAS</b>	<b>T distribución</b>	<b>Numero de ceros</b>	<b>LD PARA PLS (g)</b>
<b>TNT</b>	2.262	10	1.42

### 4.2.4. Regresión de Componentes Principales (PCR) y Regresión por mínimos cuadrados parciales (PLS) para pólvora

Para la regresión de componentes principales y la regresión por mínimos cuadrados parciales se utilizaron n=2 componentes principales, obteniéndose la figura 28, en donde los círculos blancos representan los ceros de entrenamiento, los círculos negros representan los ceros de validación, los triángulos blancos los valores de entrenamiento a diferentes concentraciones diferentes de cero y los triángulos negros los valores de validación diferentes de cero.



**Figura 28** Modelos multivariantes para pólvora

Se realizó el cálculo de las cifras de mérito en donde se obtuvieron los valores presentados en la tabla 13, para esta sustancia el modelo de regresión por mínimos cuadrados parciales es el que mejor predice al tener un error cuadrático medio de 0.39 en comparación con el otro modelo, ambos modelos son poco confiables al medir esta sustancia.

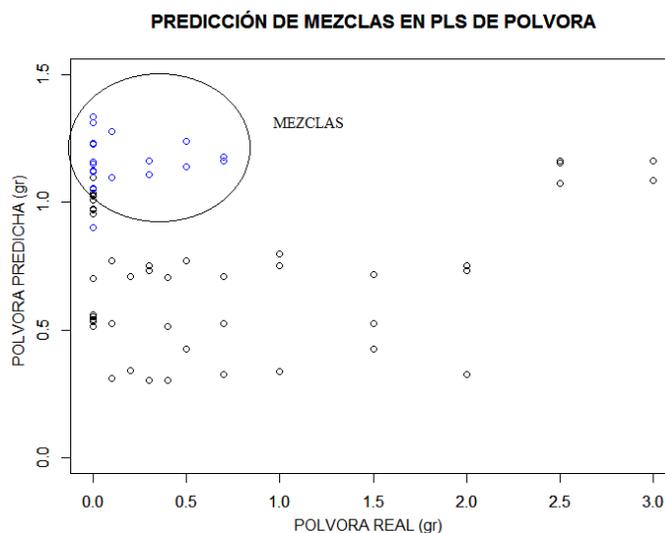
**Tabla 13**

*Cifras de mérito para la pólvora con los dos modelos*

MODELO	n	MSEP	RMSE	R <sup>2</sup>	Calificación
PCR	2	0.87	0.93	0.11	Muy malo
PLS	2	0.59	0.79	0.39	Malo

Las mezclas de pólvora con jabón y pasta dental fueron proyectadas en el modelo de regresión por mínimos cuadrados parciales (PLS), obteniéndose la figura 29, los círculos en negro son las muestras de validación, mientras que las muestras en azul son las mezclas, se puede observar que las sustancias mezcladas no son predichas del mismo modo que las sustancias puras de pólvora, obteniéndose un error cuadrático medio de predicción de 0.87. Puede deberse

principalmente a la composición química de esta sustancia, ya que al mezclarla con otras sustancias la nariz percibe de manera diferente que a la sustancia original.



**Figura 29** Predicción de mezclas de pólvora en el modelo PLS de pólvora pura

#### 4.2.5. Límite de detección (LD) de pólvora

Para el cálculo del límite de detección de la pólvora se utilizaron los datos de T-distribución y de número de ceros que se muestran en la tabla 14. Obteniéndose casi 1g que el dispositivo puede detectar y discriminar esta sustancia.

**Tabla 14**

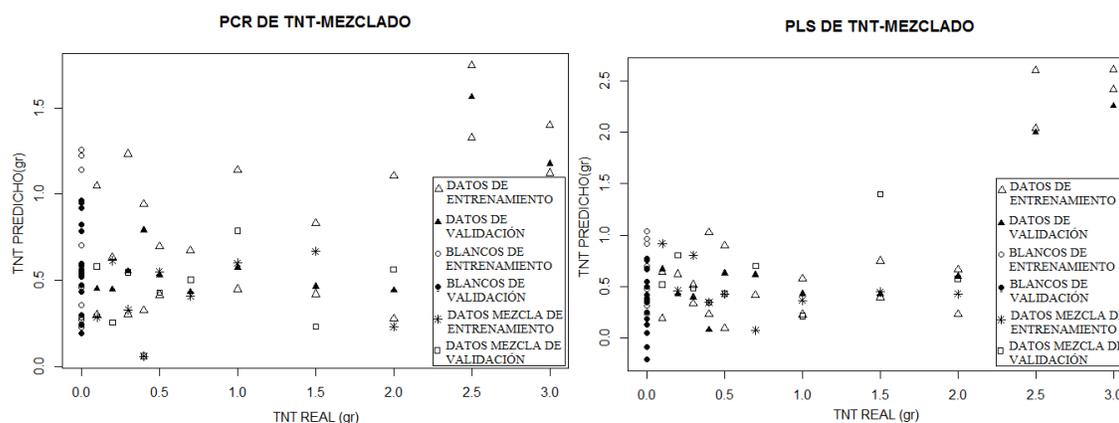
*Límite de detección para pólvora utilizando PCR y PLS*

SUSTANCIAS	T distribución	Numero de ceros	LD PARA PLS (g)
POLVORA	2.262	10	0.99

### 4.3. Análisis de Mezclas

#### 4.3.1. Regresión de Componentes Principales (PCR) y Regresión por mínimos cuadrados parciales (PLS) de las dos sustancias explosivas

Para la regresión de componentes principales y la regresión por mínimos cuadrados parciales con mezclas de TNT se utilizaron  $n=4$  componentes principales y  $n=5$  componentes principales respectivamente, obteniéndose la figura 30, en donde los círculos blancos representan los ceros de entrenamiento, los círculos negros representan los ceros de validación, los triángulos blandos los valores de entrenamiento a diferentes concentraciones distintos de cero, los triángulos negros los valores de validación, los asteriscos son las mezclas de las sustancias con TNT que se utilizaron de entrenamiento y los rectángulos son los datos de validación de las mezclas de TNT.

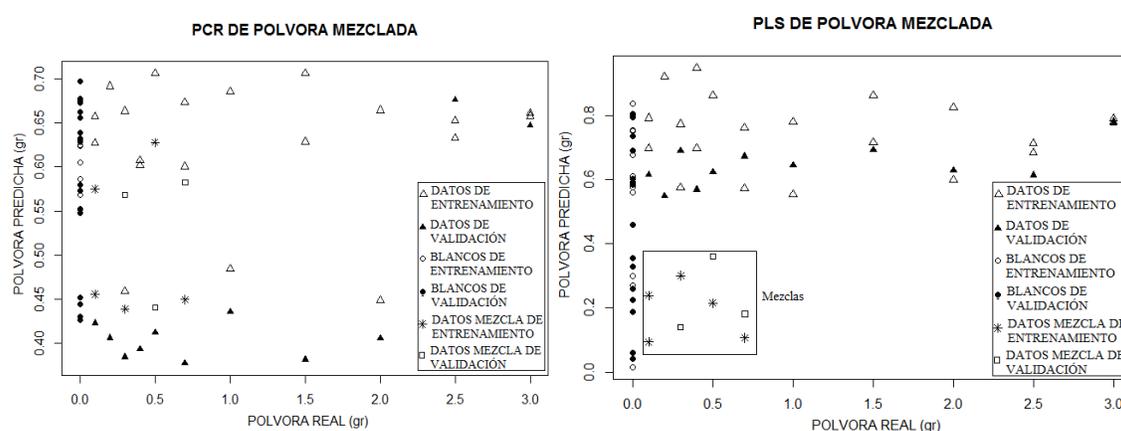


**Figura 30** Modelos multivariantes para TNT mezclado

En la figura 30 se observa que en concentraciones menores el modelo PLS funciona de mejor manera prediciendo que el modelo PCR para TNT mezclado.

Con pólvora mezclada se realizaron los modelos de calibración multivariantes, para el modelo de regresión de componentes principales se utilizaron  $n=1$  componentes principales y para el modelo de regresión por mínimos cuadrados parciales se utilizó un  $n=2$  componentes

principales, por lo que se obtuvieron las gráficas que muestran en la figura 31, en donde los círculos blancos representan los ceros de entrenamiento, los círculos negros representan los ceros de validación, los triángulos blandos los valores de entrenamiento a diferentes concentraciones distintos de cero, los triángulos negros los valores de validación, los asteriscos son las mezclas de las sustancias con pólvora que se utilizaron de entrenamiento y los rectángulos son los datos de validación de las mezclas de pólvora.



**Figura 31** Modelos multivariantes para pólvora mezclada

En la figura 31 se observa que ambos modelos reaccionan de manera similar al predecir los datos, con la única diferencia que en modelo PLS separa de manera más notoria los datos de sustancia pura de los datos de mezcla.

Se determinaron las cifras de mérito de los modelos de ambas sustancias mezcladas, las cuales se muestran en la tabla 15. El modelo PLS para ambas sustancias fue mejor prediciendo las concentraciones que el modelo PCR, ya que los valores se acercaron más a la curva de calibración y por ende se tuvo un menor error cuadrático medio.

**Tabla 15**

*Tabla de cifras de mérito para la sustancia TNT con los dos modelos*

<b>Mezcla de TNT</b>					
<b>MODELO</b>	<b>n</b>	<b>MSEP</b>	<b>RMSEP</b>	<b>R<sup>2</sup></b>	<b>Calificación</b>
<b>PCR</b>	4	0.53	0.73	0.31	Malo
<b>PLS</b>	5	0.34	0.59	0.55	Bueno
<b>Mezcla de pólvora</b>					
<b>MODELO</b>	<b>n</b>	<b>MSEP</b>	<b>RMSEP</b>	<b>R<sup>2</sup></b>	<b>Calificación</b>
<b>PCR</b>	1	0.55	0.74	0.28	Muy malo
<b>PLS</b>	2	0.23	0.48	0.70	Bueno

#### **4.3.2. Límite de detección (LD) de las sustancias explosivas con mezclas**

Para los 2 modelos de calibración se calculó el límite de detección que podía detectar y predecir la nariz electrónica al encontrarse las sustancias explosivas mezcladas con otros compuestos, obteniéndose los valores que se observan en la tabla 16, tanto para sustancias como TNT y pólvora mezcladas el modelo PCR tiene un límite de detección menor que el del modelo PLS.

**Tabla 16**

*Límite de detección de sustancias mezcladas con los modelos de regresión*

<b>SUSTANCIAS</b>	<b>T distribución</b>	<b>Numero de ceros</b>	<b>LD PARA PLS (g)</b>
<b>TNT-MEZCLA</b>	2.571	6	1.02
<b>POLVORA-MEZCLA</b>	2.571	6	0.86

#### 4.4. Discusión de resultados

Los datos predichos por el modelo de regresión de componentes principales (PCR) presentaron mayor error cuadrático medio por ende un coeficiente de determinación  $R^2$  menor en todos los casos, obteniendo un desempeño “malo” y “muy malo” como se observa en la tabla 17, lo cual quiere decir que los datos se encontraban más alejados de la recta de calibración. Esto puede darse porque este modelo está basado en el análisis de componentes principales (PCA) en el cual se observó que los datos de una misma clase se encontraban divididos. Esto pudo afectar a la predicción de datos puesto que existió una deriva instrumental, pese a haber seguido el mismo procedimiento todos los días se obtuvo este comportamiento que pudo afectar en la estimación del modelo PCR. En la tabla 17 se observa un resumen de los resultados obtenidos en cada uno de los casos.

**Tabla 17**

*Resumen de cifras de mérito y límite de detección de las sustancias y mezclas*

<b>Modelos / Sustancias</b>	<b>n</b>	<b>RMSECV</b>	<b>MSEP</b>	<b>RMSEP</b>	<b>R<sup>2</sup></b>	<b>LD (g)</b>	<b>CALIF.</b>
<b>PCR TNT</b>	1	1.04	0.88	0.94	0.10	N/A	Muy malo
<b>PLS TNT</b>	5	0.90	0.26	0.51	0.73	1.42	Bueno
<b>PCR PÓLVORA</b>	2	0.87	0.87	0.93	0.11	N/A	Muy malo
<b>PLS PÓLVORA</b>	2	0.78	0.59	0.79	0.39	0.99	Malo
<b>PCR TNT</b>	4	1.25	0.53	0.73	0.31	N/A	Malo
<b>MEZCLADO</b>							
<b>PLS TNT</b>	5	0.62	0.34	0.59	0.55	1.02	Bueno
<b>MEZCLADO</b>							

CONTINÚA



<b>PCR PÓLVORA</b>	1	0.80	0.55	0.74	0.28	N/A	Muy malo
<b>MEZCLADA</b>							
<b>PLS PÓLVORA</b>	2	0.70	0.23	0.48	0.70	0.86	Bueno
<b>MEZCLADA</b>							

Para corregir los errores experimentales obtenidos para la creación de la base de datos y disminuir el error cuadrático medio de predicción de los modelos de calibración, se podrían utilizar técnicas de rectificación en el software. Por otra parte se podría verificar el hardware del prototipo para mejorar los resultados obtenidos.

En la recolección de los datos se obtuvieron valores que presentaban mucha distorsión, por lo cual se determinó que el quinto sensor tenía problemas, no se verificó con exactitud cuál es el inconveniente con este sensor, debido a que no se encontraba en el alcance del trabajo de investigación.

Los límites de detección obtenidos para el instrumento con cada sustancia fueron: Para TNT en estado puro 1.42g, para TNT mezclado 1.02g, para pólvora pura 0.99g y para pólvora mezclada 0.86g. Según estos resultados la sustancia que tiene un mejor límite de detección fue la pólvora, pese a tener un mayor error cuadrático medio al realizar predicciones.

#### **4.5. Aplicabilidad de R en el proyecto realizado**

El software R-studio fue muy útil para la realización de este proyecto ya que cuenta con muchas funciones para la creación de algoritmos de preprocesamiento de señales y de modelos de calibración multivariante, al ser un software libre, la descarga de paquetes fue muy eficiente y no se necesitaron permisos adicionales o licencias.

En un inicio es complicada la adaptación a un nuevo lenguaje de programación y a una nueva interfaz de desarrollo, pero con el paso del tiempo se va teniendo más experiencia en la realización de algoritmos, también es muy importante recalcar que el help o ayuda de este programa tiene respuestas a muchas interrogantes y ejemplos que pueden servir al momento de utilizar una función. Los datos son cargados de manera fácil solo utilizando una función, también fue posible realizar gráficos claros que podían ser exportados como archivos de imagen lo cual fue útil.

## CAPÍTULO V

### CONCLUSIONES Y RECOMENDACIONES

#### 5.1. Conclusiones

Se generó el modelo de calibración multivariante de regresión por mínimos cuadrados parciales para determinar el límite de detección del prototipo nariz electrónica para medición de las sustancias explosivas para TNT de 1.42gr y pólvora de 0.99gr.

Se determinó la mínima concentración de sustancia explosiva que el dispositivo nariz electrónica puede predecir al encontrarse mezclada con sustancias de uso común como jabón y pasta dental, para pólvora fue de 0.86g mediante el modelo de regresión por mínimos cuadrados parciales y para la sustancia explosiva TNT la mínima cantidad en gramos que el dispositivo pudo predecir es de 1.02g.

Se realizó un análisis estadístico de los modelos utilizados para la detección de sustancias explosivas en el cual se determinó que el modelo de regresión lineal multivariante de componentes principales (PCR) fue menos eficiente en comparación al modelo de regresión por mínimos cuadrados parciales (PLS), ya que en todos los experimentos realizados se obtuvo un coeficiente de determinación en un rango desde 0.10 a 0.31 que es calificado como “muy malo”. Comparando con los coeficientes de determinación del modelo PLS que oscilan en un rango de 0.39 a 0.73 que es calificado entre “regular” y “bueno”, por lo tanto el modelo que mejor estimaciones realiza para sustancias explosivas es el modelo PLS.

Se concluyó que el modelo de análisis de componentes principales ayudó a determinar las causas por las cuales existían errores en la predicción de los datos al utilizarse los modelos de

regresión multivariantes. También se pudo corregir el error de dispersión y variabilidad que se tenían en los experimentos eliminando los valores del quinto sensor en cada muestra.

Se concluyó que el software utilizado en este proyecto fue muy beneficioso al ser una herramienta estadística para la realización del acondicionamiento de las señales, algoritmos de regresión multivariante y del cálculo de los límites de detección, ya que este cuenta con varias funciones que sirven para estos fines, además de ser un programa con una interfaz amigable para el usuario.

## **5.2. Recomendaciones**

Se recomienda que para la creación de los modelos de regresión multivariante se realicen algoritmos de preprocesamiento de señales, que permitan acondicionar y ordenar los datos experimentales de forma que estos puedan ser utilizados de manera rápida en los modelos multivariantes.

Se recomienda utilizar como técnica estadística la validación cruzada, ya que se usa comúnmente en entornos donde se requiere estimar la precisión con la que el modelo va a funcionar en la práctica siempre y cuando no se tenga una cantidad de datos demasiado elevada, ya que el costo computacional de este método puede ser también elevado.

Se recomienda que para futuros trabajos de investigación con esta nariz electrónica se utilicen valores de concentraciones superiores a 1g, ya que como se observó en los resultados el límite de detección mínimo que se obtuvo en los experimentos fue de 0.7g esto garantizaría una mejor discriminación de las sustancias que se analicen. Además sería interesante observar si los modelos de regresión multivariante predicen de manera más exacta sustancias mayores a 1g de concentración.

Se recomienda que para trabajos futuros se realicen experimentos con los cuales se permita determinar la cantidad óptima de alcohol que el sistema de dopaje debe tener para que no existan errores experimentales.

Se recomienda que al utilizar una función del software R-studio primero se investiguen los parámetros de entrada y salida de dicha función, mediante la ayuda (help) que presenta el mismo programa.

Se recomienda realizar un análisis correctivo en la parte física del prototipo nariz electrónica, para determinar errores que afecten la realización de trabajos futuros, sobre todo verificar el funcionamiento del quinto sensor y si es necesario reemplazarlo, ya que en este trabajo de investigación eso se encontraba fuera del alcance del mismo.

## BIBLIOGRAFÍA

(n.d.).

American Chemical Society. (1993). Limit of Detection. *American Chemical Society*.

Andrew, W. (n.d.). Cross-Validation for detecting and preventing overfitting. *Carneigie Mellow University*.

Baldwin, E. B., & Dea, S. (2011). Electronic Noses and Tongues: Applications for the Food and Pharmaceutical Industries. *Sensors*.

Barnston, A. (1992). Correspondence among the Correlation, RMSE, and Heidke Forecast Verification Measures; Reinement of the Heidke Score. *Climate Analysis Center, Washington, D.C.*

Barrera, O., & Barat, J. (2016). Estudio de la capacidad de la nariz y lengua electrónica para la caracterización y clasificación de cafés. *Universitat Politècnica de València*.

Boggio, N., Alonso, P., Ortiz, J., Lamagna, A., Boselli, A., & Rinaldi, C. (2008). Desarrollo de un espectrómetro por movilidad iónica para la detección de compuestos orgánicos (Explosivos, drogas y contaminantes). *ANALES AFA*.

Boqué, R. (2014). El Límite de Detección de un Método Analítico. *Universitat Rovira i Virgili*.

Brezmes, J., & Correig, X. (2001). Diseño de una nariz electrónica para la determinación no destructiva del grado de maduración de la fruta. *Universitat Politècnica de Catalunya*.

Brezmes, J., Cabré, P., Rojo, S., Llobet, E., Vilanova, X., & Correig, X. (2005). Discrimination Between Different Samples of Olive Oil Using Variable Selection Techniques and Modified Fuzzy Artmap Neural Networks. *IEEE SENSORS JOURNAL*.

- Caygill, J., Davis, S., & Higson, J. (2011). Current trends in explosive detection techniques. *Elsevier*.
- Charumporn, B., & Omatu, S. (2002). Classifying Smokes Using an Electronic Nose and Neural Networks . *IEEE Proceedings of the 41st SICE Annual Conference* . , 2661 – 2665.
- Correa, E. (2003). *Desarrollo de una aplicación a la detección de aromas para una nariz electrónica del tipo Microbalanzas de Cristal de Cuarzo (QCM) para su adaptación a la determinación de calidad en frutas*. Madrid.
- Currie, L. (1968). *Limits for Qualitative Detection and Qualitative Determination: Application to Radiochemistry,Anal.*
- Díaz, J. (2006). Vulcanología: invento de una nariz electrónica capaz de oler gases. *Centro de Alta Tecnología de San José, Costa Rica. Turcon*.
- Faraway, J. (2002). Practical Regression and Anova using R.
- Fernández de Labastida, M. (2011). Aplicación de lenguas electrónicas en la monitorización de procesos de biosorción de metales. *Universitat Politècnica de Catalunya*.
- Fonollosa, J. V., & Huerta, R. (2012). Sensor Failure Mitigation Based on Multiple Kernels. *IEEE*.
- Gallardo, A. (2006). Geoestadística. *Asociación Española De Ecología Terrestre*.
- García, A. (2008). Análisis de regresión: Casos univariante y multivariante.
- García, P. C., & P., T. (n.d.). Determinación de la calidad interna de mango mediante espectroscopía visible y análisis multivariante. *Universidad Politecnica de Valencia*.
- Gaviria, C. (2016). Regresión por Mínimos Cuadrados Parciales PLS Aplicada a Datos Variedad Valuados. *Universidad Nacional de Colombia* .
- Gonzalez, J. M., & Blanco, J. (2011). The Multi-Chamber Electronic Nose—An Improved Olfaction Sensor for Mobile Robotics. *Sensors*.

- Gutierrez, R. P., & Sun, P. (2001). Chemosensory Adaptation In an Electronic Nose. *IEEEXPLORE*.
- Gutierrez, R., & Gutierrez, A. (2003). Habituation in the KIII Olfactory Model With Chemical Sensor Arrays. *IEEE*.
- Gutierrez, R., & Troy, H. (1999). A Method for Evaluating Data-Preprocessing Techniques for Odor Classification with an Array of Gas Sensors. *IEEE*.
- Gutierrez-Osuna, R. (n.d.). Leave-one-out Cross Validation. *Wright State University*.
- Holland, S. (2016). Principal Components Analysis (PCA). *Department of Geology, University of Georgia, Athens, GA 30602-2501*.
- Joanneum, F. (2006). Cross-Validation Explained. *Instituto de Genómica y Bioinformática*.
- Lillenthal, J. L., & T., D. (2006). Airbone Chemical Sensing with Mobile Robots. *Sensors*.
- Llamas, L. (2017, marzo 15). *Tutoriales Arduino Avanzados*. Retrieved from <https://www.luisllamas.es/arduino-filtro-media-movil/>
- López, I. (2016). *Desarrollo de un prototipo Electrónico de sensado químico, para la detección de trinitrotolueno (TNT) y pólvora base doble en un ambiente controlado*. Quito.
- Martens, H., & Geladi, P. (2004). Multivariate Calibration. *Encyclopedia of Statistical Sciences*.
- Marvall, A. (1993). Reconocimiento de Formas y Visión Artificial. *Universidad Autonoma de Nayarit (UAN), Mexico*.
- Menéndez, A. (2015). Desarrollo de un sistema de sensores para la deteccion de sustancias peligrosas. *Universidad Autónoma de Madrid*.
- Ministerio de Coordinación de Seguridad. (2013). Plan Nacional de Seguridad Integral. *Quito, Pichincha, Ecuador*.

- Ministerio de Defensa Nacional del Ecuador. (2015). *Operaciones militares para control de armas y sustancias ilícitas*. Retrieved febrero 22, 2018, from <http://www.defensa.gob.ec/operaciones-militares-para-el-control-de-armas/>
- Ministerio de Defensa Nacional del Ecuador. (2017). *Operaciones militares para control de armas y sustancias ilícitas*. <http://www.defensa.gob.ec/operaciones-militares-para-el-control-de-armas/>.
- Mocak, J., Bond, A., Mitchell, S., & Schollary, G. (1997). *A Statistical Overview of Standard (IUPAC and ACS) and New Procedures for Determining the Limits of Detection and Quantification: Application to Voltammetric and Stripping Techniques*. *Pure Appl. Chem.*
- Montenegro, A., Trejos, M., & Araúz, C. (2015). Nariz electrónica inalámbrica móvil con monitoreo en tiempo real. *Revistas Académicas UTP, Universidad Tecnológica de Panamá*.
- Montes de Oca, R. G., & Hernández, A. (2009). Uso de técnicas de análisis multivariable aplicadas en la obtención de modelos de predicción de propiedades relacionadas con los sistemas agrícolas. *Revista Ciencias Técnicas Agropecuarias*.
- Morelisi. (2018). Modelo de regresion lineal simple. *Universitat de València*. Retrieved from <https://www.uv.es/uriel/material/Morelisi.pdf>
- Morelisi. (2018). Modelo de regresion lineal simple. *Universitat de València*. Retrieved from <https://www.uv.es/uriel/material/Morelisi.pdf>
- Moreno, I. C., & Jiménez, A. (2009). La Nariz Electrónica: Estado del Arte. *Revista Iberoamericana de Automática e Informática Industrial*, 2-15.
- Moreno, I., Caballero, R., Galán, R., Matía, F., & Jiménez, A. (2009). La Nariz Electrónica: Estado del Arte. *Revista Iberoamericana de Automática e Informática Industrial*, 2-15.

- Muños, A. (2015). *Análisis de regresión: Casos univariante y multivariante*. Retrieved febrero 16, 2018, from <http://ocw.uc3m.es/estadistica/aprendizaje-del-software-estadistico-r-un-entorno-para-simulacion-y-computacion-estadistica/analisis-de-regresion-casos-univariante-y-multivariante>
- Nahvi, A., & Hollerbach, J. M. (1996). The Noise Amplification Index for Optimal Pose Selection in Robot. *IEEE*.
- Navarro, O. (2009). Selección de Variables de Regresión Componentes Principales. *LACCEI*.
- Olguín, C. (2017). Aproximación al biomimetismo mediante lenguas y narices electrónicas en medios complejos: detección de explosivos, agentes nerviosos y control de alimentos. *Universitat Politècnica de València*.
- Olivieri, A. (2006). Regresión Lineal en Química analítica. *Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario*.
- Paredes-Doig, A., Sun-Kou, M. R., Picasso-Escobar, G., Doig-Camino, E., & Comina, G. (2006). Implementación y evaluación de una nariz electrónica para la detección de alcoholes lineales. *Rev. Colomb. Quim.*  
doi:<http://dx.doi.org/10.15446/rev.colomb.quim.v45n2.60393>.
- Payam, R. (2008). k-fold Cross-Validation. *Arizona State University*.
- Pearce, T., Schiffman, S., Nagle, H., & Gardner, J. (2003). Handbook of Machine Olfaction. *Electronic Nose Technology*.
- Persaud, K., Pisanelli, A., Evans, P., & Travers, P. (2006). Monitoring urinary tract infections and bacterial vaginosis. *Sensors and Actuators B: Chemical*.
- Raschka, S. (2015, 07 25). *Quora*. Retrieved 02 22, 2018, from <https://www.quora.com/What-is-the-difference-between-LDA-and-PCA-for-dimension-reduction>

- Rodríguez, E. (2015). Modelo de Regresión Lineal: Mínimos Cuadrados Parciales(PLS) Como Metodo Cuasi-Newton. *Revecitec Urbe*.
- Rodríguez, M. (2001). Análisis sensorial de alimentos Nariz electrónica. *Dialnet*.
- Rojo, J. (2007). Regresión Lineal Múltiple. *Instituto de Economía y Geografía*.
- Saevels, S., Berna, Z., Lammertyn, J., Di Natale, C., & B., N. (2004). Characterisation of QMB sensors by means of the BET adsorption isotherm. *Sensors and Actuators B: Chemical*.
- Saevels, S., Lammertyn, J., Berna, A., E., V., Di Natale, C., & Nicolai, B. (2003). Electronic nose as a non-destructive tool to evaluate the optimal harvest date of apples. *Elsevier*.
- Sameer, M. (2003). Explosives detection systems (EDS) for aviation security. *Elsevier*.
- Santanu, P., Rajib, B., & Bhattacharyya, N. (2012). Multi-class Support Vector Machine for Quality Estimation of Black Tea Using Electronic Nose. *Sixth International Conference on Sensing Technology*.
- Uriel, E. (2013). Introducción a la Econometría. *Universidad de València*.
- Valdéz, D. (2010). Regresión por Mínimos Cuadrados Parciales. *Varianza*.
- Valero, X., & Francese, A. (2012). Análisis de la señal acústica mediante coeficientes cepstrales bio-inspirados y su aplicación al reconocimiento de paisajes sonoros. *Grup de Recerca en Tecnologies Mèdia, La Salle-Universitat Ramon Llull*.
- Vallejo, Z., & Zurita, D. (2017). *Diseño e implementación de un sistema mecatrónico portable con dopaje automático para detección de muestras explosivas*. Quito.
- Zhang, H., Wang, J., Sheng, Y., & Chang, M. (2008b). Evaluation of peach quality indices using an electronic nose by MLR, QPST and BP network. *Sensors and Actuators B: Chemical*.