MODELO DE INFERENCIA DE LA RESPUESTA DE UN SENSOR DE GAS DE ESTADO SÓLIDO PARA SISTEMAS DE OLFATO ELECTRÓNICO

JUAN CARLOS RODRÍGUEZ GAMBOA

Institución Universitaria
INSTITUTO TECNOLÓGICO METROPOLITANO -- ITM
MAESTRÍA EN AUTOMATIZACIÓN Y CONTROL INDUSTRIAL
FACULTAD DE INGENIERÍAS
MEDELLÍN, COLOMBIA
JUNIO, 2013

MODELO DE INFERENCIA DE LA RESPUESTA DE UN SENSOR DE GAS DE ESTADO SÓLIDO PARA SISTEMAS DE OLFATO ELECTRÓNICO

JUAN CARLOS RODRÍGUEZ GAMBOA

Trabajo de investigación para optar al título de
Magíster en Automatización y Control Industrial

Director:
M.Sc. Jorge Alberto Jaramillo Garzón
Co-Director:
D.Sc. Edilson Delgado Trejos

Asesor:
D.Sc. Cristhian Manuel Durán Acevedo

Línea de investigación:
Máquinas Inteligentes y Reconocimiento de Patrones

INSTITUTO TECNOLÓGICO METROPOLITANO -- ITM
MAESTRÍA EN AUTOMATIZACIÓN Y CONTROL INDUSTRIAL
FACULTAD DE INGENIERÍAS
MEDELLÍN, COLOMBIA
JUNIO, 2013

INFERENCE MODEL OF THE SOLID-STATE GAS SENSOR RESPONSE FOR ELECTRONIC NOSE SYSTEMS

JUAN CARLOS RODRÍGUEZ GAMBOA

A thesis submitted to the postgraduate program "Masters in Industrial Control and Automation" in partial fulfillment of the requirements for the Master's degree

Director:
M.Sc. Jorge Alberto Jaramillo Garzón
Co-Director:
D.Sc. Edilson Delgado Trejos

Advisor:
D.Sc. Cristhian Manuel Durán Acevedo

Research Line:
Machine Intelligence and Pattern Recognition

INSTITUTO TECNOLÓGICO METROPOLITANO
MASTERS IN INDUSTRIAL CONTROL AND AUTOMATION
FACULTY OF ENGINEERING
MEDELLÍN, COLOMBIA
JUNE, 2013

A mi amada esposa por acompañarme en todo momento y lugar, por su comprensión y apoyo incondicional.
A mi madre y a mi hermana, quienes siempre han sido ejemplo de superación y constante motivación para mi crecimiento personal y profesional

AGRADECIMIENTOS

Expreso mis agradecimientos a:

M.Sc. Jorge Alberto Jaramillo Garzón, docente investigador del Instituto Tecnológico Metropolitano - Colombia**,** por su valioso aporte académico y el acompañamiento en todo el proceso como director del trabajo de grado.

D.Sc. Edilson Delgado Trejos, decano facultad de ingenierías del Instituto Tecnológico Metropolitano - Colombia, quién ha estado presente en diferentes etapas de este proceso, inicialmente como director y finalmente como asesor, motivando siempre la finalización de este trabajo.

D.Sc. Cristhian Manuel Durán Acevedo, docente investigador de la Universidad de Pamplona - Colombia, gestor y orientador del camino tomado en esta línea de investigación de sistemas multisensoriales y reconocimiento de patrones.

El Instituto Tecnológico Metropolitano - Colombia (ITM) por ofrecer todas las garantías, espacios y recursos durante los estudios de maestría y para el trabajo de investigación materializado en el proyecto que relata este libro.

La Universidad de Pamplona - Colombia (UP) y muy especialmente al grupo de Investigación en Sistemas Multisensoriales y Reconocimiento de Patrones por permitir el acceso y trabajo con las bases de datos A-NOSE y B-NOSE, las cuales contienen mediciones realizadas con sistemas de olfato electrónico.

Mis compañeros de maestría en Automatización y Control del Instituto Tecnológico Metropolitano, a quienes les comparto este logro.

TABLA DE CONTENIDO

Pág.

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

GLOSARIO

- **Entrenamiento:** Procedimiento matemático relacionando con un algoritmo supervisado de reconocimiento de patrones, a través del cual se pretende encontrar un modelo o los parámetros más adecuados para un modelo a partir de un conjunto de datos.
- **Escalado: A**daptación de los datos a una escala o con respecto a un punto de referencia.
- **Espécimen:** En biología un espécimen es un individuo o parte de un individuo que se toma como muestra, en este trabajo un espécimen es una correspondiente medida de algún compuesto o elemento, tomada con algún sistema de olfato electrónico.
- **Estimador:** Es un modelo o función empleado para estimar un parámetro desconocido de una población, en este documento se relaciona con sensor virtual.
- **Filtro:** Dispositivo que elimina o selecciona ciertas frecuencias de un espectro.
- **Kernel:** Función matemática que define un producto punto es un espacio de dimensionalidad generalmente alta. Se usa en las máquinas de regresión por vectores de soporte para extender el modelo lineal a un modelo no lineal.
- **Matriz:** Este término se utiliza mucho en este documento cuando se hace referencia a un arreglo de sensores de gases, esto se debe a que los datos obtenidos de los mismos por lo general son almacenados en matrices.
- **Modelo:** En este documento hace referencia a los modelos matemáticos, los cuales buscan representar fenómenos o relaciones entre ellos a través de una formulación matemática.
- **Pre-procesamiento:** Es un proceso que consiste en la manipulación de los datos con el propósito de extraer información poco útil o impura que pueda interferir posteriormente; en algunos casos sirve para ajustar los datos a unos rangos o parámetros que faciliten su manipulación.
- **Regresión:** Es una técnica estadística utilizada para establecer la relación entre variables; en este trabajo se empleó para establecer el modelo que prediga el comportamiento de un determinado sensor.
- **Sensor:** Dispositivo capaz de detectar magnitudes físicas o químicas y transformarlas en magnitudes eléctricas, también puede ser visto como un dispositivo que convierte una forma de energía en otra.
- **Sensor Virtual:** Conocidos como *soft-sensor* o estimador, hace referencia a modelos o funciones matemáticas que permiten predecir o estimar el comportamiento de una variable física a partir de otros datos o mediciones.
- **Sintonización:** En este documento se utiliza el término para referirse al ajuste de parámetros de la máquina de vectores de soporte para regresión.
- **Sistema de Olfato Electrónico:** Un dispositivo que incluye el hardware y software necesario para emular el funcionamiento del sentido del olfato, capaz de detectar olores con mucha precisión y exactitud.

- **Validación:** Procedimiento estadístico que permite estimar la capacidad de generalización del modelo ante datos que no estén presentes en el conjunto de entrenamiento.
- **Volátil:** Sustancia que se transforma fácilmente en vapor o gas cuando está expuesta al aire.

ACRÓNIMOS

- **MOS:** Semiconductor de óxido metálico.
- **SOE:** Sistema de olfato electrónico también conocido como nariz electrónica.
- **SVM:** Máquina de vectores de soporte.
- **SVR:** Máquina de vectores de soporte para regresión.
- **VS:** Sensor virtual.

RESUMEN

Un Sistema de Olfato Electrónico (SOE) o nariz electrónica puede ser visto como un instrumento o equipo de medida de olfato artificial, el cual tiene una matriz de sensores de gases como parte fundamental, destinada a medir los volátiles presentes en los diferentes olores (Rodríguez, Durán, & Reyes, 2010; Tian et al., 2005)

Los SOE actuales son básicamente arreglos de sensores químicos conectados a sistemas de procesamiento o de cómputo, en los cuales se aplican técnicas avanzadas de procesamiento digital de señales y reconocimiento estadístico de patrones. Su objetivo fundamental es permitir la cualificación de olores a través de tareas de clasificación, discriminación, predicción e incluso cuantificación de productos, elementos o componentes de acuerdo a sus características organolépticas (Runu , Bipan , Laxmi , Arun , Nabarun , & Rajib , 2012; Rodríguez, Durán, & Reyes, 2010; Durán & Baldovino, 2009; Wilson & Baietto, 2009; Zhou et al., 2006).

Este tipo de sistemas usan complejos arreglos o matrices de sensores de gases para la detección de los diferentes olores. A mayor cantidad de sensores se puede obtener mejores representaciones de las muestras sometidas a medición, pero trae consigo aumento en el costo debido a los sensores, el sistema de adquisición de datos y otras partes del equipo; además, el tamaño del hardware y el consumo de potencia también se ven afectados (Ghasemi-Varnamkhasti, Mohtasebi, Razavi, Siadat, Ahmadi, & Dicko, 2012; Brattoli, M. & et. al. 2009; Durán Acevedo, 2005).

En este trabajo de grado de maestría se propone un modelo de inferencia de la respuesta de un sensor de gas de estado sólido tipo MOS (Semiconductor de óxido metálico) que permita obtener un sensor virtual (Zhang & Liu, 2013; Mielle, Marquis, & Latrasse, 2000; Chen, Song & Li, 2005; Liu & et. al, 2009; Ibarg & Reyes, 2006), con el fin de emplearse en sistemas de olfato electrónico que utilicen el modelo propuesto, principalmente para reducir costos y permitir la construcción de sistemas de olfato electrónico de tamaño más reducido.

El modelo propuesto está basado en la utilización de métodos de regresión (Chen, Sond & Li, 2005; Gu & Wang, 2008) con máquinas de vectores de soporte SVR, para inferir la respuesta de un sensor de gas de tipo MOS de una referencia (por ejemplo TGS822), a partir de la respuesta de otros sensores de gases MOS de otras referencias.

Con el modelo obtenido producto de esta investigación, se solucionan los problemas mencionados en cuanto a costos y tamaños de los sistemas de olfato electrónico, entre otros. Adicionalmente se espera que con el modelo propuesto se puedan obtener múltiples ventajas en la realización de pruebas bajo condiciones controladas y en tareas de calibración de sensores (Llobeta & et. al. 2001; Brattoli, M. & et. al. 2009; Knobloch & et. al., 2009; Peris & Escuder-Gilabert, 2003; Brattoli & et. al., 2011), posibilitando que se

utilicen con más frecuencia en el sector industrial y para múltiples aplicaciones (Rodríguez-Gamboa, Albarracín-Estrada & Delgado-Trejos, 2011).

Como resultado de esta investigación se logró una nueva alternativa para el desarrollo de sistemas de olfato electrónico, aportando una solución tecnológica que servirá para disminuir costos en los equipos desarrollados y en la operación de los mismos. Sin embargo, se debe tener en cuenta que el desarrollo de este tipo de sensores virtuales requiere un conocimiento previo del hardware o equipo de medida que se va a utilizar, adicionalmente se requieren datos históricos o bases de datos de medidas tomadas con el sistema de olfato electrónico y adicionalmente se encontró que debe existir un cierto grado de correlación entre el sensor que se desea modelar y los demás sensores a partir de los cuales se va inferir el modelo, en orden de facilitar la obtención del modelo de inferencia, esto último por lo general no supone un problema debido a que se ha encontrado que es común que haya una correlación fuerte entre sensores de gases tipo MOS de diferentes referencias.

ABSTRACT

Electronic nose system (SOE) or electronic nose can be seen as an instrument or measuring equipment of olfaction artificial, which has an array of gas sensors as fundamental part, designed to measure volatiles present in the different odors (Rodriguez , Durán, & Reyes, 2010, Tian et al., 2005).

The current SOE are basically arrays of chemical sensors, connected to processing systems or computer system, which apply advanced techniques of digital signal processing and statistical pattern recognition, the main objective is to allow the qualification of odors through classification task, discrimination, prediction and even quantification of products, components or components according to their organoleptic characteristics (Runu , Bipan , Laxmi , Arun , Nabarun , & Rajib , 2012; Rodriguez, Durán, & Reyes, 2010; Durán & Baldovino, 2009, Wilson & Baietto, 2009, Zhou et al. 2006).

These systems are using arrays complex or gas sensors matrix for detecting different odors. With a larger number of sensors is possible to obtain better representation of the samples being measured, but generate higher cost due to the sensors, the data acquisition system and other parts of the equipment, also the size of hardware and the power consumption are affected (Ghasemi-Varnamkhasti, y otros, 2012; Brattoli, M. y otros, 2009; Durán Acevedo, 2005).

This master thesis proposes an inference model from the response of a solid state gas sensor kind MOS (Metal Oxide Semiconductor) for obtain a virtual sensor (Mielle, Marquis, & Latrasse, 2000; Chen, Song & Li, 2005; Liu & et. al, 2009; Ibarg & Reyes, 2006), to use in electronic nose systems that employed the proposed model, mainly to reduce costs and allow the construction of electronic nose systems of smaller size.

The proposed model is based on the use of regression methods (Chen, Sond & Li, 2005; Gu & Wang, 2008) with support vector machines SVR to infer the response of a gas sensor of MOS type of a reference (i.e. TGS822) from the response of other MOS gas sensor other references.

With the model obtained as result of master's thesis, the problems referred in terms of costs and sizes of electronic nose systems among others are solved. Additionally, it is expected that the proposed model can obtain many advantages in testing under controlled conditions and sensor calibration tasks (Llobeta & et. all. 2001; Brattoli, M. & et. all. 2009; Knobloch & et. al., 2009; Peris & Escuder-Gilabert, 2003; Brattoli & et. al., 2011), allowing more frequently used in industry and for multiple applications (Rodriguez-Gamboa, Albarracin-Estrada & Delgado -Trejos, 2011).

As a result of this thesis was achieved a new alternative for the development of electronic nose systems, providing a technology solution that will help reduce costs in the developed

equipment and operation thereof. However, it should be noted that the development of this type of soft-sensors requires prior knowledge of the hardware or measurement equipment to be used, additionally it require historical data or databases measurements using the olfactory system electronic. Besides, it found that there must be some degree of correlation between the sensor to be modeled and from other sensors which will be used to infer the model, in order to provide the obtaining of inference model, this point usually not a problem because it has been found it is common a strong correlation between type MOS gas sensors of different references.

INTRODUCCIÓN

Un Sistema de Olfato Electrónico (SOE) es un instrumento que permite distinguir y reconocer olores o aromas. Este instrumento está conformado por una serie de módulos que trabajan de forma conjunta, los cuales permiten analizar muestras gaseosas, vapores y olores. Un instrumento o equipo de este tipo tiene al menos tres (3) partes o subsistemas, las cuales son: el sistema de medida o matriz de sensores de gases, el sistema de suministro de volátiles y el sistema de procesamiento (Rodríguez-Gamboa, Albarracín-Estrada & Delgado-Trejos, 2011; Durán & Baldovino, 2009; Tian et al., 2005).

Los sistemas de olfato electrónico se pueden llegar a utilizar ampliamente en nuestro país en múltiples industrias, tales como: alimentaria, farmacéutica, de productos químicos, militar, petrolera, de explotación de metales preciosos, entre otras; debido al gran potencial y riqueza que tiene Colombia con respecto a materias primas y recursos naturales. Con el fin de sacar el mayor provecho, a favor de la versatilidad y teniendo en cuenta la variedad de aplicaciones en las cuales se pueden utilizar, por ejemplo: para el control de calidad de los alimentos (Rodríguez-Gamboa, Albarracín-Estrada & Delgado-Trejos, 2011; El Barbri, Llobet, El Bari, Correig, & Bouchikhi, 2008; Durán, C. & Baldovino, D., 2009), detección de gases tóxicos, medición de nivel y factores de contaminación (Brattoli, M. & et. al. 2009; Zhou & et. al., 2005), detección de explosivos y narcóticos (Ross J. Harper, Jose R. Almirall, Kenneth G. Furton, 2005) o para el análisis y diagnóstico de enfermedades (Velásquez & et. al., 2009; Chen & et. al., 2005).

La falta de uso de esta tecnología en nuestro país se debe en gran medida al desconocimiento de la misma, a los costos que implica y a la incipiente investigación de esta línea en nuestro país. El desarrollo de este trabajo de grado estuvo encaminado en desarrollar una metodología que permitiera reducir los costos asociados al desarrollo de sistemas de olfato electrónico, adicionalmente buscando reducir el tamaño de estos equipos y con el propósito de fomentar aún más la investigación en estos temas, sin dejar de lado que favorece la aplicabilidad de los sistemas de olfato electrónico en el ámbito nacional.

En Colombia y en el mundo existe gran potencial de aplicabilidad de esta tecnología, por ejemplo, se encuentra que actualmente en Colombia los productos tradicionales ocupan el primer lugar en las exportaciones (ver **Figura 1**); en parte debido a los buenos precios internacionales de los bienes básicos (petróleo, carbón, café, ferroníquel), esta clasificación del DANE (Departamento Administrativo Nacional de Estadística) obedece a las exportaciones en Colombia desde el 2004 hasta el 2010. Los productos que exporta nuestro país sin mencionar el consumo interno, representan un gran potencial de industrias en las cuales podrían utilizarse sistemas de olfato electrónico para realizar tareas de clasificación, discriminación, cuantificación y predicción, posibilitando su implementación en procesos de control de calidad y en la producción en general.

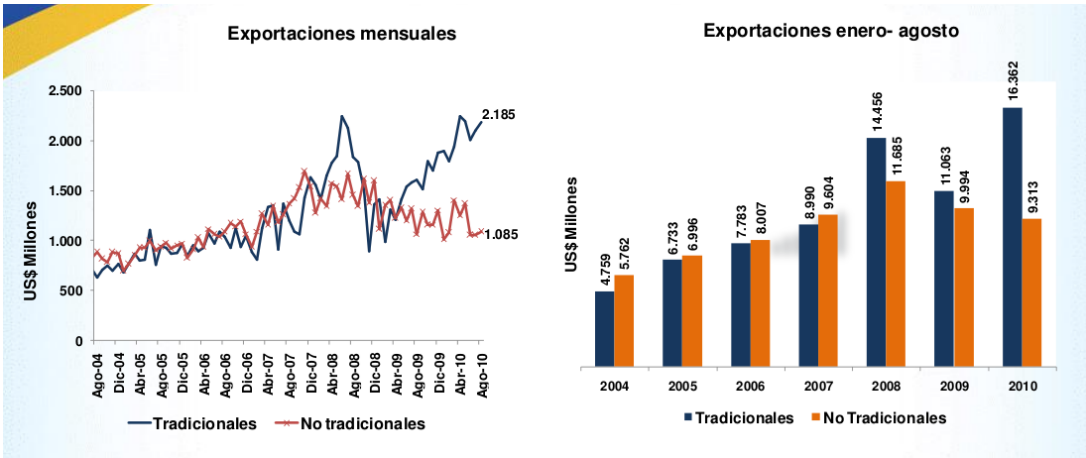


Figura 1. Exportaciones de Productos Tradicionales y no Tradicionales (Agosto 2004 - Agosto 2010). Imagen tomada del Informe de Exportaciones Colombianas, Proexport Colombia, Agosto de 2010

OBJETO DE ESTUDIO

Teniendo en cuenta la línea de investigación de SOE del cual hace parte este trabajo de grado, se realizó el mapa conceptual mostrado en la **Figura 2** donde se divide el problema de los sistemas de olfato electrónico en tres aspectos fundamentales, los cuales son: el Hardware de los sistemas de olfato electrónico, el Software de los sistemas de olfato electrónico y las aplicaciones realizadas con dichos sistemas. Cada una de estas se subdividió en otras más específicas. El cuadro resaltado en la **Figura 2**, indica en cuál temática se encuentra el objeto de estudio de este trabajo de grado y en la **Figura 3** se muestra de forma más detallada el marco dentro del cual se encuentra este trabajo.

En la **Figura 3** se encuentra resaltado el objeto del trabajo de grado de forma más específica. Este trabajo de grado de maestría se centró en la simulación (sensor virtual) de un sensor de gas de tipo MOS; para ello el estudio del estado del arte realizado se centró en las técnicas de modelado y caracterización de las respuestas obtenidas por los sensores de gases y las técnicas utilizadas para obtener sensores virtuales.

ACERCAMIENTO AL PROBLEMA

Esta tesis de maestría se centró en la matriz de sensores de gases o subsistema de medida, la cual es la encargada de realizar la detección de los compuestos volátiles en un Sistema de Olfato Electrónico. Este subsistema por lo general presenta derivas en los sensores, requiere calibración de los mismos y presenta problemas de repetitividad en las medidas

debido a la saturación de los sensores. Es importante mencionar que las matrices de sensores de gases usualmente utilizan sensores del mismo tipo pero de diferentes referencias (Ejemplo: TGS822, TGS821, TGS813, etc.), con el fin de obtener un mayor solapamiento entre las señales para facilitar las tareas de clasificación y detección de olores (Durán Acevedo, 2005).

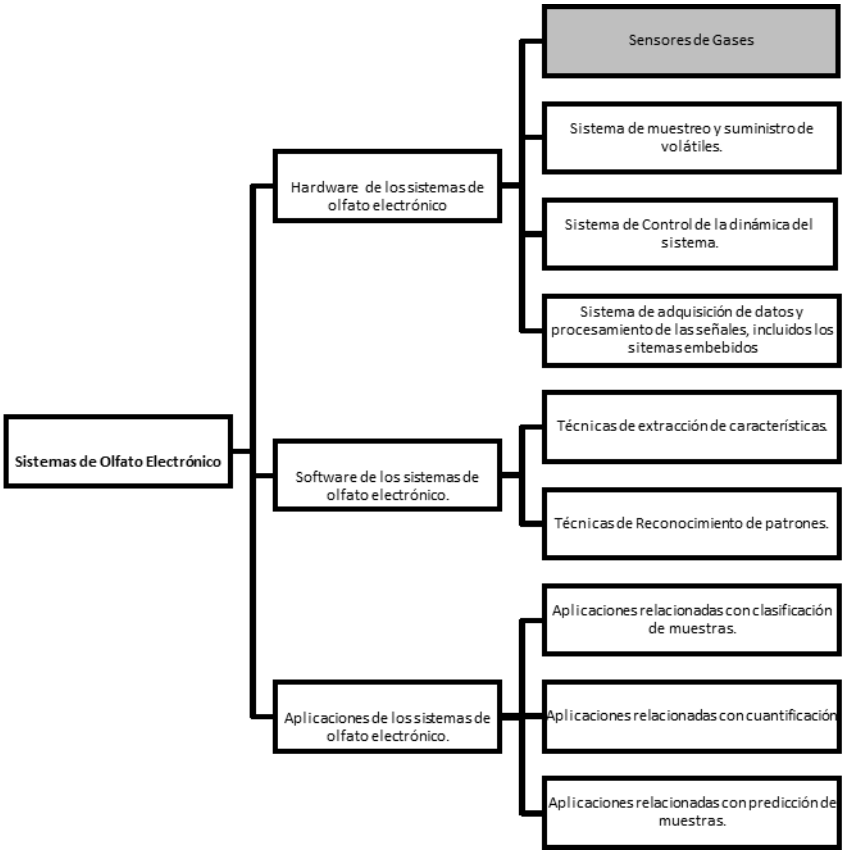


Figura 2. Mapa Conceptual General de los tópicos relacionados con los sistemas de olfato electrónico

Los problemas relacionados específicamente con los sensores de gases se evidencian aún más en matrices de sensores que utilizan una cantidad considerable de sensores, esto se debe a que, i.e., en un SOE con *N* sensores, el mínimo número de parámetros extraídos en cada medida será *N* (uno por sensor), aunque pueden ser muchos más cuando utilizamos información dinámica y por lo general se desea obtener un amplio número de variables descriptoras por experimento para realizar las tareas de clasificación, predicción, entre otras (Gualdron Guerrero, 2006). Razón por la cual se utilizan matrices de sensores de gases, adicionalmente porque un dispositivo sensor de este tipo, comúnmente no es selectivo a un solo compuesto y esto obliga a utilizar múltiples sensores con sensibilidades solapadas para poder solucionar los problemas de medidas en las cuales frecuentemente hay fuentes interferentes, es decir, otros analitos presentes (Durán Acevedo, 2005).

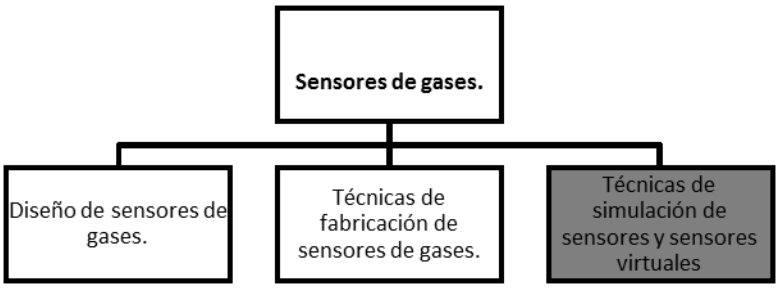


Figura 3. Aproximación al objeto del trabajo de grado

Otros problemas relacionados con los sensores de gases se presentan cuando los sensores tienen mucho tiempo de uso (envejecimiento), ocasionando por ejemplo la reorganización de la superficie del sensor durante largos periodos de tiempo; también cuando se utilizan con mucha frecuencia o por el contrario cuando se dejan de utilizar por largos periodos de tiempo (envenenamiento), i.e., cuando se da unión irreversible debido a la contaminación externa, estos problemas mencionados se conocen como la deriva real o de primer orden. Otras deriva se conocen como derivas de segundo orden, inherentes al sistema de medición, a las condiciones externas, incluyendo los cambios del ambientes (temperatura y humedad), el ruido de los sistemas y el proceso de acondicionamiento de la muestra y los efectos térmicos o el efecto memoria debido a mediciones anteriores. También se debe considerar que la utilización de un número considerable de sensores de gases en un SOE trae consigo algunos problemas, dentro de los cuales cabe mencionar el aumento en el costo del equipo (por los sensores, el sistema de adquisición de datos, entre otras partes), el tamaño del hardware se incrementa y el consumo de potencia también se ve afectado.

Actualmente se han venido adelantando estudios para mejorar la respuesta, funcionamiento y operación de los sensores de gases utilizados en sistemas de olfato electrónico, donde se puede encontrar diseños novedosos como el mostrado en (Li, Dao, & Sugiyama, 2007), simples como en (Kato & Mukai, 2007), para aplicaciones específicas y generales como en (Peris & Escuder-Gilabert, 2009; Romain, Delva, & Nicolas, 2008), también existen desarrollos que buscan la estandarización como en (Ulivieri, Distante, Luca, Rocchi, & Siciliano, 2006), para mitigar o compensar el efecto de las derivas (Vergara, y otros, 2012) e incluso el concepto de sensores virtuales como en (Krutzler, y otros, 2012; Chen, y otros, 2005; Mielle , Marquis, & Latrasse, 2000). En la mayoría de estos casos se realizan experimentación y validación de los diseños; pero se encuentran falencias en los métodos de calibración de sensores y no se especifican procedimientos ni metodologías para lograr reproducibilidad en las medidas. La utilización de sensores virtuales se ha probado y ha mostrado ser viable pero no se ha definido un método para obtener o caracterizar los sensores para ser utilizados de forma virtual.

OBJETIVOS PROPUESTOS

OBJETIVO GENERAL

Proponer un modelo de inferencia de la respuesta de un sensor de gas de tipo MOS de una referencia específica, a partir de otros sensores de gases del mismo tipo de otras referencias, utilizando el método de regresión de soporte vectorial SVR para obtener un sensor virtual.

OBJETIVOS ESPECÍFICOS

- Caracterizar la respuesta de un sensor de gas de estado sólido usando técnicas de análisis estadístico multivariante, con el fin de proponer un esquema de representación de la información del sensor.

- Obtener el modelo de inferencia de la respuesta de un sensor de gas, utilizando el método de regresión por vectores de soporte SVR.

- Validar el modelo de inferencia propuesto frente a un marco experimental basado en sensores reales.

1. MARCO TEÓRICO Y ESTADO DEL ARTE

1.1 GENERALIDADES DE LOS SISTEMAS DE OLFATO ELECTRÓNICO

Los Sistema de Olfato Electrónico (SOE) actuales, también conocidos comúnmente como narices electrónicas, son básicamente arreglos de sensores químicos conectados a sistemas de procesamiento o de cómputo, en los cuales se aplican técnicas avanzadas de procesamiento digital de señales y reconocimiento estadístico de patrones. Su objetivo fundamental es permitir la cualificación de olores a través de tareas de clasificación, discriminación, predicción e incluso cuantificación de productos, elementos o componentes de acuerdo con sus características organolépticas (Durán & Baldovino, 2009; Wilson & Baietto, 2009; Zhou et. al, 2005).

1.1.1 Partes de un Sistema de Olfato Electrónico (SOE)

Un Sistema de Olfato Electrónico o nariz electrónica puede ser visto como un instrumento o equipo de medida de olfato artificial, conformado por una serie de módulos que trabajan de forma conjunta, los cuales permiten analizar muestras gaseosas, vapores y olores (Durán & Baldovino, 2009; Tian et. al, 2005). Un instrumento o equipo de este tipo tiene al menos 4 partes, en la **Figura 4** se presenta la representación de un sistema de olfato electrónico presentado en (Rodríguez-Gamboa, Albarracín-Estrada, & Delgado-Trejos, 2011). Note que en esta gráfica aparecen más de cuatro bloques o partes, lo cual se debe a que algunas de ellas pueden estar integradas en una sola o se pueden omitir de acuerdo al tipo de sistema de olfato que se requiera o a la aplicación para la cual haya sido desarrollado.

En los siguientes párrafos se detallan las funciones específicas de las 4 partes fundamentales:

Matriz de sensores de gases.

La gran mayoría de sistemas de olfato electrónico poseen como elemento principal una matriz de sensores de gases. Es conveniente que dicha matriz esté localizada en una cámara o compartimiento especial en el cual se garanticen unas condiciones adecuadas para su correcto funcionamiento. Principalmente se debe asegurar el adecuado aislamiento que impida que se introduzcan contaminantes y mantener la presión y temperatura adecuadas (estos parámetros son importantes o críticos en función del tipo de sensor utilizado) (Rodríguez-Gamboa, Albarracín-Estrada, & Delgado-Trejos, 2011).

Otra ventaja de utilizar una cámara de sensores es que facilita el proceso de medida debido a que los volátiles van a estar en mayor concentración y tienen más contacto con el elemento activo de los sensores, lo cual permite una mejor y más rápida respuesta de los

sensores[1]. También se ha encontrado experimentalmente que entre más hermética sea la cámara de sensores se aprovechan mejor las ventajas mencionadas. En la **Figura 5** se aprecia la fotografía de una cámara de sensores desarrollada en (Rodríguez, Durán, & Reyes, 2010).

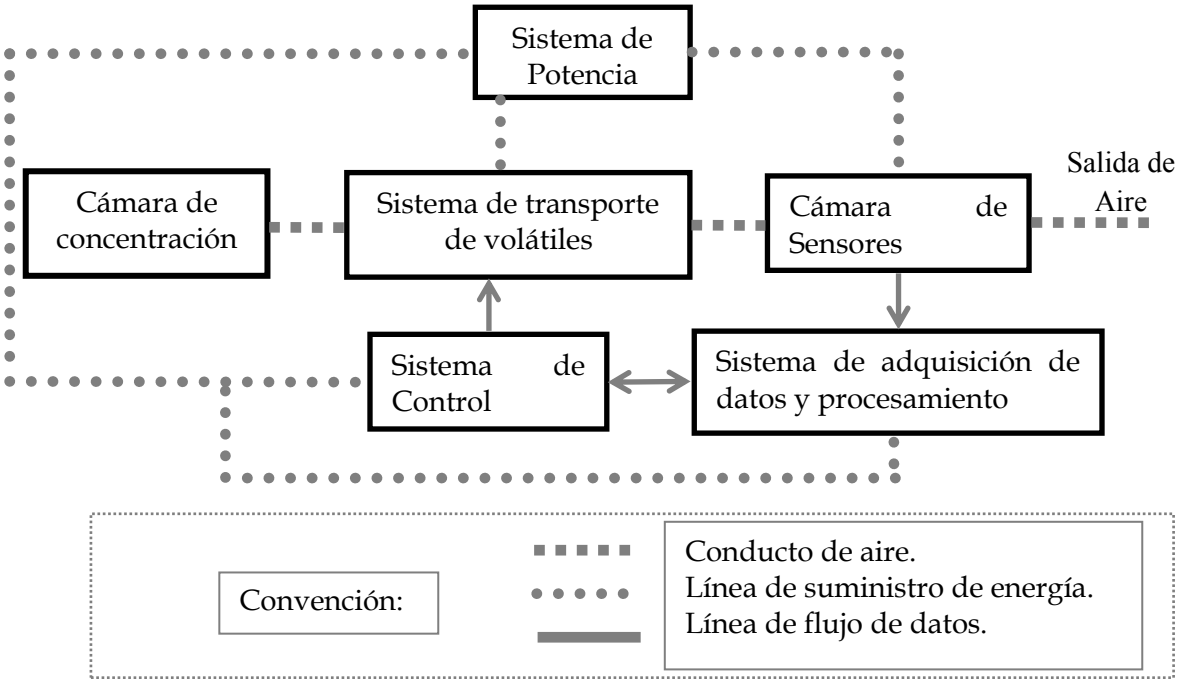


Figura 4. Diagrama de Bloques que representa un sistema de olfato electrónico

Este subsistema por lo general presenta derivas en los sensores, por lo que se requiere calibración de los mismos para evitar problemas de repetitividad en las medidas debido a la saturación que se puede presentar. Es importante mencionar que las matrices de sensores de gases usualmente utilizan sensores del mismo tipo pero de diferentes referencias (Ejemplo: TGS822, TGS821, TGS813, etc.), con el fin de obtener un mayor solapamiento entre las señales buscando facilitar las tareas de clasificación y detección de olores (Durán Acevedo, 2005).

Sistema de transporte de Volátiles

El sistema de transporte de volátiles es parte fundamental de los SOE, ya que condiciona el funcionamiento y permite que realicen los procesos de medición y purga de los sensores. Básicamente es un sistema que se encarga de transportar hacia la cámara de sensores los volátiles desprendidos por la muestra o elemento que se va a analizar. En algunas ocasiones se inyecta la muestra en la cámara de sensores de forma manual, con los consiguientes problemas de error y lentitud que ello implica. En otras ocasiones un sistema automático se encarga de transportar las moléculas olorosas o volátiles,

[1] *Puede encontrar más información de los sensores de gases en el ítem 1.2 de este documento.*

extrayéndolos de la zona en la que se encuentra la muestra a través de la inyección de algún tipo de gas o aire, hasta llevarlos a la cámara de sensores (Rodríguez, Durán, & Reyes, 2010).

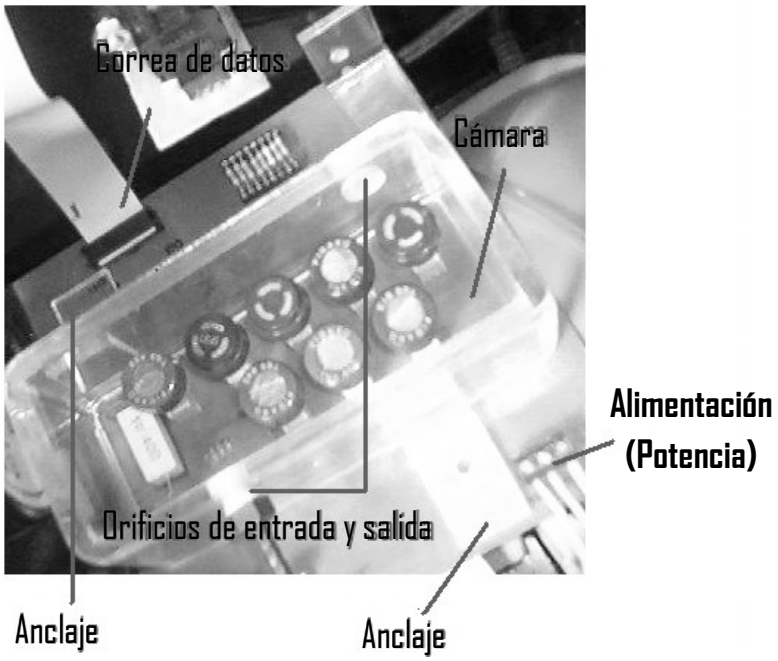


Figura 5. Cámara de sensores de gases, ofrece hermeticidad y garantiza la realización de medidas confiables

Además, los sistemas de olfato electrónico en su mayoría cuentan con algún tipo de mecanismo de limpieza de la cámara de sensores de forma que las medidas sucesivas se hagan partiendo de las mismas condiciones iníciales y se garantice la repetibilidad de los resultados. Se recomienda utilizar otra cámara o compartimento hermético para alojar la muestra que se desea analizar, siempre que las condiciones del medio y del experimento lo permitan.

Sistema de Control y de Adquisición de Datos

El sistema de control se encarga de manejar adecuadamente el sistema de suministro de volátiles, por ejemplo, las electroválvulas, la bomba de aire y demás dispositivos que hagan parte de este sistema. También está encargado de controlar los demás sistemas adicionales que pueda tener el Sistema de Olfato Electrónico, como es el sistema de control de temperatura y humedad, entre otros (Durán Acevedo, 2005).

El sistema de adquisición de datos se encarga de capturar las señales proporcionadas por los sensores de gases y las entrega al sistema de cómputo, el cual con algún software especializado procesa dicha información.

Los sistemas de control y de adquisición de datos pueden estar integrados por algunos o varios de los siguientes dispositivos: una tarjeta de adquisición de datos, un micro controlador, un DSP (Procesador Digital de Señales) o un computador con su adecuada etapa de potencia para manejar los elementos que demanden más corriente, así como la configuración adecuada de memoria para lograr almacenar la enorme cantidad de datos que se obtienen de los sensores (Rodríguez, Durán, & Reyes, 2010).

Se recomienda trabajar con una tarjeta de adquisición de datos y un computador para lograr buena capacidad de almacenamiento de información, procesamiento adecuado de la información y representación gráfica. Aunque en algunas ocasiones por necesidad de portabilidad se opte por trabajar con un DSP o un micro-controlador.

Parte importante del sistema de control es la fuente de potencia, la cual debe ser de unos cuantos amperios, dependiendo de la cantidad de sensores de gases y de los elementos adicionales que se utilicen. Una fuente de 3 Amperios puede ser suficiente si se está trabajando con un SOE que contiene una matriz de 8 sensores de gases.

Sistema de Procesamiento

El sistema de procesamiento en la mayoría de los casos está compuesto por un computador con el software adecuado para manipular los datos obtenidos por los sensores. A los datos obtenidos se les aplican técnicas de pre-procesamiento para extraer los parámetros estáticos de las medidas y reducir la cantidad de información a analizar. Después, se aplican técnicas de análisis multivariado como Análisis de Componentes Principales (PCA) y de reconocimiento de patrones como Redes Neuronales Artificiales (RNA), Máquinas de Vectores de Soporte (SVM), entre otras, para realizar tareas tales como: clasificación, discriminación, predicción, cuantificación de muestras de acuerdo a sus características organolépticas (Wilson & Baietto, 2009; Berna, 2010).

1.1.2 Funcionamiento de los Sistemas de Olfato Electrónico

El funcionamiento de los Sistemas de Olfato Electrónico depende de las partes que lo componen, así como de las funcionalidades que tenga dicho equipo. Para realizar medidas con un SOE básicamente se comienza realizando un proceso de adecuación de la muestra que se va a examinar, esto depende del tipo de elemento que se quiere analizar, el cual en algunas ocasiones debe calentarse, realizarle cortes, mezclarlo con otros elementos o simplemente basta con colocarlo cerca de la matriz de sensores o en la cámara de concentración dispuesta para este propósito (Durán & Baldovino, 2009).

Después de esto es conveniente esperar unos cuantos minutos para que la muestra expida suficientes volátiles y posteriormente se inicia el proceso de medida, para lo cual se deben depositar o transportar los volátiles expedidos por la muestra hacia la cámara de sensores. Durante la realización del proceso de medida, el sistema de adquisición de datos registra paso a paso los cambios de la señal de salida de cada uno de los sensores de gases. Una vez terminado el proceso de medida, se inicia la limpieza de la cámara de sensores e inmediatamente se pueden procesar y analizar los datos almacenados (procesamiento *off line*), utilizando el software especializado para el pre-procesamiento y procesamiento de

las señales, con el propósito de obtener la huella olfativa que representa dicha muestra y así poder realizar las respectivas tareas de clasificación, discriminación, entre otras (Berna, 2010; Wilson & Baietto, 2009).

1.2 SENSORES DE GASES

En general, los sensores de gases son dispositivos que constan de dos partes principales, la primera es un elemento activo cuyas propiedades físicas o químicas cambian en presencia del analito que se desea detectar y la segunda parte es un elemento transductor que convierte los cambios de las propiedades del elemento activo en una señal eléctrica. Estos sensores generalmente tienen una membrana selectiva que impide el paso de partículas o material indeseable, actuando como un primer filtro de ruido. En la **Figura 6** se puede observar un esquema simplificado de un dispositivo de este tipo, en el cual se pueden apreciar las principales partes de un sensor de gas y la naturaleza de las entradas y salidas (Grupo E-Nose, 2011, Tian et al., 2005).

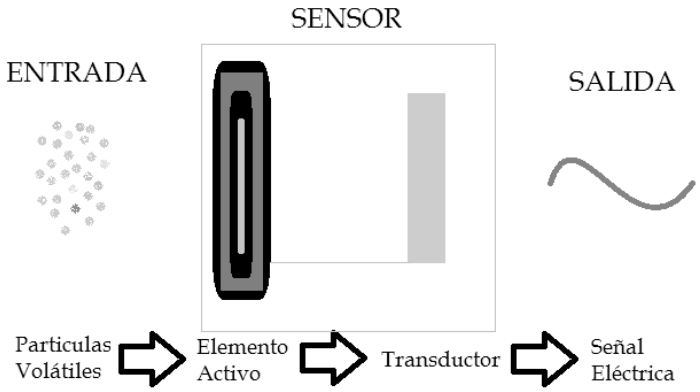


Figura 6. Esquema simplificado de un sensor de gas

Existen diferentes tipos de sensores de gases para emplear en SOE, los más utilizados son: MOX (*Metal Oxide Semiconductor*), QCM (*Quartz Crystal Microbalance*), SAW (*Surface Acoustic Waves*), MOSFET (*Metal Oxide Semiconductor Field Effect Transistor*), CP (*Conducting Polymers*), FO (*Fiber Optics*). En esta tesis de maestría se trabaja específicamente los sensores MOX, construidos con materiales semiconductores como el óxido de estaño (SnO_2), Óxido de Zinc (ZnO), Oxido de Titanio (TiO_2), entre otros. Su principio de funcionamiento se basa en el cambio de la conductividad de un material sensible cuando éste reacciona con los gases presentes en su entorno. En la **Figura 7** se aprecian varios sensores comerciales de este tipo (Berna, 2010).

Los sensores de gases son en su mayoría para propósito general y suelen poseer alta sensibilidad, llegando a detectar niveles muy bajos de concentración de volátiles, pero presentan desventajas cuando se desean determinar concentraciones de un solo

componente, debido a que la señal de salida no puede asignarse unívocamente a dicho componente por su misma generalidad (Durán Acevedo, 2005). Aunque en pruebas realizadas con el sistema de olfato electrónico desarrollado en (Rodríguez-Gamboa, Albarracín-Estrada, & Delgado-Trejos, 2011) se encontró que con la utilización de métodos de reconocimiento de patrones para entrenamiento supervisado y un adecuado entrenamiento, se pueden obtener muy buenos resultados para problemas de cuantificación de concentraciones.

Figura 7. Sensores de gases comerciales fabricados por Fígaro y FIS. (Imágenes tomadas de las páginas web de los fabricantes)

El tipo de señal obtenido a partir de los sensores de gases es mostrado en la **Figura 8**, donde se puede apreciar una medida de acetona (Base de Datos B-NOSE) obtenida con un sensor TGS826. Como se mencionó anteriormente, en los sistemas de olfato electrónico se utilizan matrices o arreglos de sensores de gases con las cuales se obtienen señales como las mostradas en la **Figura 9**.

1.3 PRE-PROCESAMIENTO DE LAS SEÑALES DE LOS SENSORES DE GASES

El pre-procesamiento de los datos se emplea comúnmente para realizar las siguientes tareas: descartar datos anómalos (*outliers*), normalizar la información que se va a procesar, realizar la compensación para mitigar las derivas y/o extraer los parámetros descriptivos de la respuesta del arreglo de sensores, a fin de preparar el vector de características para el futuro procesamiento de la información (Gutierrez-Osuna, 2002).

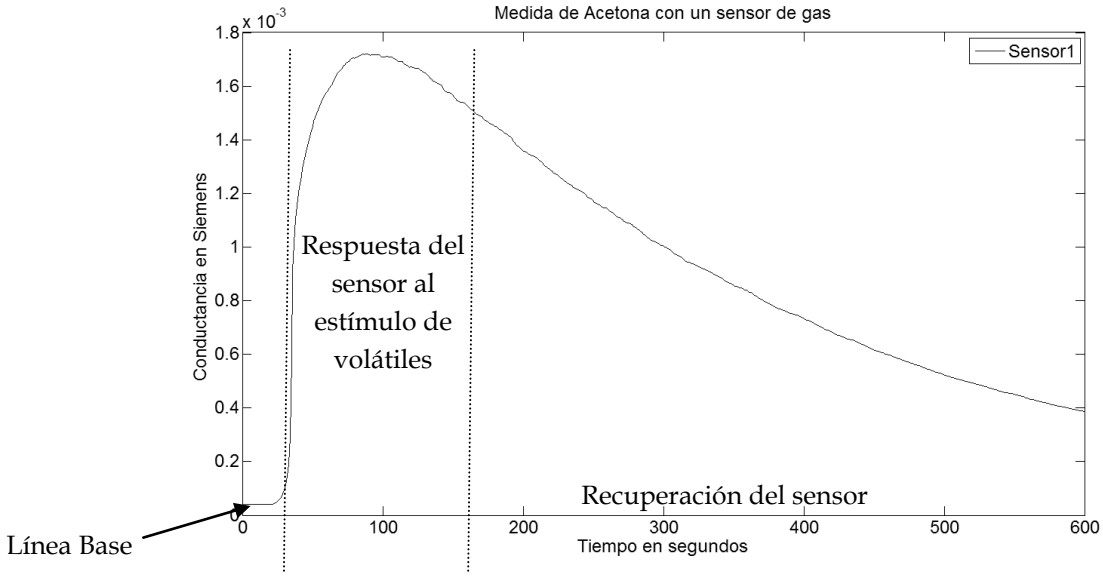


Figura 8. Señal de un sensor de gas TGS826 para una medida de acetona

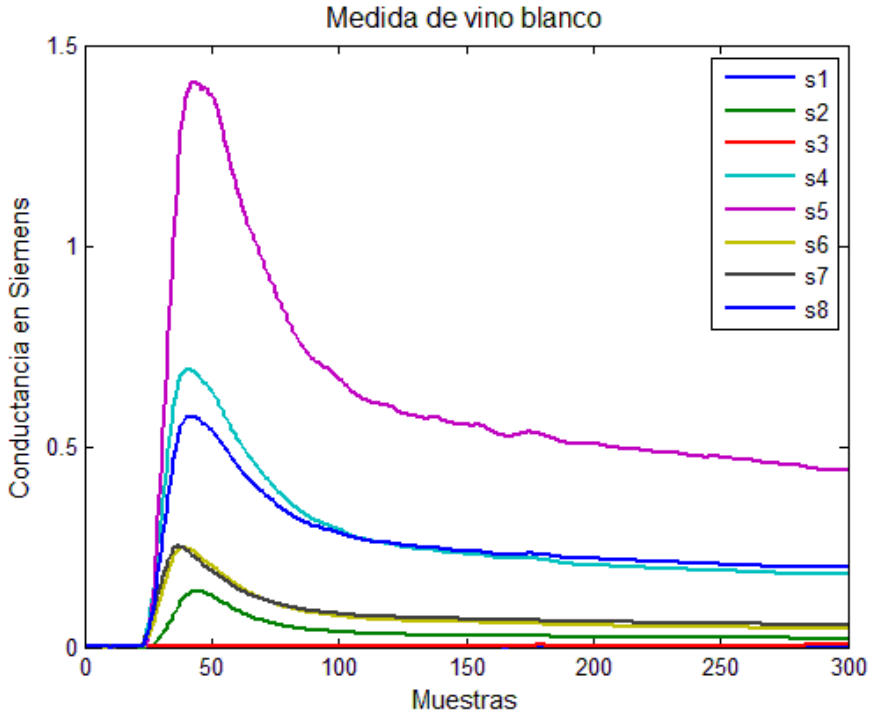


Figura 9. Señal de una matriz de 8 sensores de gases en una medida de vino blanco (Base de datos A-NOSE)

1.3.1 Remoción de datos anómalos

El tipo de señales que manejan los sensores de gases permite la pronta detección de datos anómalos realizando una inspección visual de los datos tomados. Sin embargo, como no es frecuente encontrar datos anómalos en este tipo de señales, entonces no se aplica ningún método para detectarlos y removerlos. En caso de que exista algún sensor defectuoso, o que este entregando una respuesta inadecuada, es mejor omitir el efecto de ese sensor, lo que en la práctica sería eliminar la columna correspondiente en la matriz de datos.

1.3.2 Escalado y normalización

Los métodos de escalado y normalización más comunes son los siguientes: (I) Media centrada: la media se resta de cada variable, (II) Auto-escalado: cada variable primero se centra y luego se divide por su desviación estándar, (III) Normalización: las variables se dividen por la raíz cuadrada de la suma de los cuadrados de las variables; (IV) Suma de fila constante: cada variable se divide por la suma de todas las variables en cada muestra; (V) Variable de normalización: las variables se normalizan con respecto a una sola variable, (VI) Transformación de rango: el valor mínimo de una variable se establece en 0, el valor máximo a 1, y todos los valores intermedios se encuentran a lo largo de un rango lineal entre 0 y 1 (Berrueta, Alonso-Salces, & Héberger , 2007).

1.3.3 Manipulación de la Línea Base

En los sensores de gases se pueden aplicar técnicas de pre-procesamiento para manipular la línea base, produciendo una transformación de la respuesta del sensor relativa a esta línea, con el objeto de mitigar o compensar los efectos de las derivas, como se presenta en el trabajo de (Gonzalez-Jimenez, Monroy, & Blanco, 2011), donde los autores reportan que al inicio de cada experimento las diferencias en la línea de base de los sensores fueron corregidas mediante la adición de una compensación a cada sensor. Para ello un factor de multiplicación fue calculado para cada sensor, con el fin de asegurar la obtención de valores de línea base idénticos. Debido al comportamiento no lineal de los sensores se seleccionó una ganancia media calculada a partir de tres diferentes concentraciones. Cabe destacar que bajo el supuesto de los experimentos de corto plazo, la deriva de referencia por humedad, temperatura o incluso intoxicación, es insignificante, por lo cual no se tuvieron en cuenta para este estudio. Los métodos comúnmente empleados para este propósito son: diferencial, relativo y fraccional.

El método diferencial consiste en restar la línea base al conjunto de datos y puede ser utilizado para eliminar la deriva aditiva de la respuesta del sensor (Bahraminejad, Basri, Isa, & Hambli, 2012).

$$\Delta Gs(k) = Gs(k) - Gs(0) \qquad \textbf{(1)}$$

Dónde:
Gs(k), es la salida original del sensor.
Gs(0), es el valor inicial de la línea de base.
ΔGs(k), es el valor de la salida del sensor ajustado.

El método relativo divide por la línea base el conjunto de datos, removiendo la deriva multiplicativa y generando una respuesta adimensional.

El método fraccional resta y divide por la línea base el conjunto de datos, generando una respuesta adimensional y normalizada.

1.3.4 Filtrado de la Señal

Debido a que las señales presentan ruido inherente a la adquisición y el proceso de digitalización, se hace necesario emplear algún tipo de filtro que mejore la componente de ruido aleatorio de las señales. En este caso se analizan los siguientes filtros: el de media móvil en el tiempo y el filtro *Butterworth*. En este caso en particular es necesario aplicar un filtro para atenuar el ruido aleatorio de alta frecuencia, para lo cual se implementa un filtro pasa bajas.

Filtro de media móvil en el tiempo

Es un filtro tipo ventana que se desplaza a través de los datos. La expresión matemática de un filtro de media móvil viene dada por:

$$y(i) = \frac{1}{M}\sum_{j=0}^{M-1} s[i+j] \quad (2)$$

Donde $s[i]$ es la señal de entrada, $y[i]$ es la correspondiente señal de salida y M es el número de puntos escogidos para promediar (tamaño de la ventana) denominado factor de anchura del filtro. El procedimiento corresponde a calcular los promedios de los M puntos de los datos de entrada e ir desplazándose por los demás datos de entrada repitiendo el procedimiento (Guiñón, Ortega, García-Antón, & Pérez-Herranz, 2007).

Filtro Butterworth.

Los filtros *Butterworth* provienen como muchos otros elementos del análisis de series temporales cuando existe la necesidad de procesar señales. Los filtros *Butterworth* permiten tanto la estimación de tendencias a largo plazo como la extracción directa de una señal cíclica (Bógalo & Quilis, 2003).

Los filtros *Butterworth* de paso bajo son operadores ARMA cuya función de ganancia obedece a la siguiente expresión:

$$|G(\omega)|^2 = \frac{1}{1+\left[\frac{\tan(\omega/2)}{\tan(\omega_c/2)}\right]^{2d}} \quad (3)$$

Donde w es la frecuencia expresada en radianes y está entre $0 \leq \omega \leq \pi$. Esta función está controlada por dos parámetros: la frecuencia de corte (w_c) y el grado del filtro d (Bógalo & Quilis, 2003).

1.4 SENSORES VIRTUALES

En la industria y en general en plantas o procesos existe la necesidad de instrumentos que contengan sensores, el propósito de estos últimos es obtener datos acerca de variables físicas o de los cambios en el entorno. Teniendo en cuenta la importancia de los sensores en el monitoreo y control de diversos procesos, en ocasiones se hace necesario obtener más datos o mejorar las formas de sensado. Es por ello, que desde hace dos décadas investigadores de todo el mundo trabajando con bases de datos de diferentes procesos han intentado construir modelos predictores basando en dichos datos, dichos modelos predictores son llamados *soft-sensors*. Desde ese entonces, los *soft-sensors* se establecieron como una valiosa alternativa a los medios tradicionales para la adquisición de variables en procesos críticos, monitoreo de procesos y otras tareas en las cuales se deben medir variables físicas o controlar procesos (Kadleca , Gabrys , & Strandtb, 2009).

Los *soft-sensors* también conocidos como *smart sensors,* estimadores o sensores virtuales (VS), son sensores desarrollados mediante software especializado, basándose en el modelamiento obtenido del comportamiento de sensores reales (ver **Figura 10**). Estos sensores toman lecturas de sensores físicos y calculan la salida o salidas usando los modelos del proceso (Liu , Kuo, & Zhou, 2009). En la **Figura 10** las entradas corresponden a las señales obtenidas con los sensores físicos y la salida es la predicción del sensor simulado o virtual.

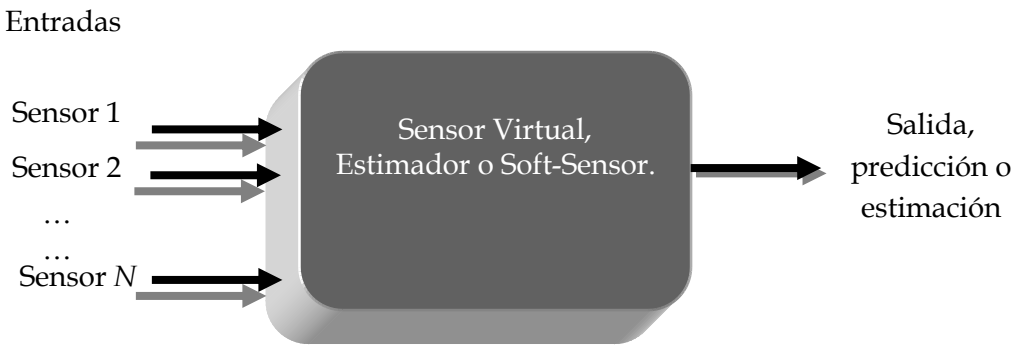


Figura 100. Esquema General de un Sensor Virtual

Obtener un sensor de gas virtual supone algunas ventajas y desventajas en cuanto a su implementación en sistemas multisensoriales de olfato electrónico como se muestra en la **Tabla 1.**

Variable a analizar	Sensor Real	Sensor Virtual
Costos	20 dólares aproximadamente por sensor	Una vez desarrollada la metodología es relativamente económico desarrollar tantos sensores virtuales como se quiera
Tamaño de los sistemas de olfato electrónico	Depende de la cantidad de sensores y del hardware asociado	Podría llegar a ser menor debido a que requiere menor sensores y por lo tanto menos hardware
Portabilidad	Depende del hardware, por lo general los sensores requieren fuentes de voltaje de unos cuantos amperios	Se pueden llegar a construir sistemas más portables y de menor consumo
Consumo de energía	Mayor	Nulo en cuanto al hardware.
Instalación	Si se tiene el hardware solo se debe conectar en el socket correspondiente	La implantación del sensor virtual se hace a nivel de software, por lo tanto el modelo del sensor virtual se debe incorporar sobre el software que se utilice para el procesamiento
Deterioro del sensor y vida útil.	Este tipo de sensores se deterioran con el tiempo y a medida que se les da uso o se dejan de utilizar. La vida útil puede ser de un par de años.	Debido a la dependencia del sensor con respecto a los sensores virtuales su respuesta se vería afectada, pero, esto se podría resolver realizando una nueva sintonización del sensor y esto alagaría su tiempo de vida.
Dependencia con respecto a los demás sensores	Ninguna	Completamente
Referencias comerciales	Variadas y para múltiples aplicaciones	Se podrían llegar a obtener sensores virtuales de variadas referencias y para múltiples aplicaciones

Tabla 1. Comparación entre los sensores de gases reales y los sensores de gases virtuales

Los VS son usados comúnmente para estimar variables en procesos en los cuales se hace difícil la medición directa u *online* o cuando los sensores son escasos o costosos. Una de las dificultades cruciales de los *soft-sensors* es la predicción precisa en ambientes o plantas cambiantes. Para hacer frente a este problema se puede optar por un modelo de regresión el cual se pueda actualizar. Sin embargo, si el modelo se actualiza con una muestra anormal la capacidad de predicción se puede deteriorar (Kaneko, Arakawa, & Funatsu, 2009).

En la **Figura 11** se muestra una metodología base que contiene los pasos más comúnmente empleados en la práctica para el desarrollo de soft-sensors (Kadleca , Gabrys , & Strandtb, 2009). La metodología presentada en esta figura es bastante general y por lo tanto se puede aplicar tanto para procesos continuos y por lotes, así como para múltiples aplicaciones o áreas.

Las tendencias en el desarrollo de soft sensors en el 2009 se muestran en la **Figura 12**. De acuerdo a lo mostrado en esta figura las metodologías más populares para la construcción o desarrollo de soft sensors son las técnicas de estadística multivariante, i.e. PCA y PLS, con una cobertura del 38% para ambas técnicas en las aplicaciones presentadas en (Kadleca , Gabrys , & Strandtb, 2009). Otras técnicas comúnmente aplicadas para la obtención de sensors virtuales son las redes neuronales basadas en métodos como MLP, RNN, entre otros. Entre los métodos empleados más recientemente para el desarrollo de sensores virtuales se encuentran los neuro-fuzzy, los cuales tienen la ventaja de proveer un mecanismo intrínseco de adaptación/evolución, así mismo los métodos basados SVM proveen muy buena generalización y han probado funcionar en diferentes áreas.

Evidentemente del 2009 hasta la fecha este panorama ha cambiado, sobre todo en cuanto a las dos metodologías mencionadas anteriormente (neuro-fuzzy y SVM) en las cuales actualmente es común encontrar cada vez más trabajos de sensores virtuales, también se mantienen las técnicas de estadística multivariante basadas en PLS y PCA, mientras que las metodologías basadas en redes neuronales han bajado su popularidad en relación con este tipo de aplicaciones y esto se debe en parte a la falta de generalización de las redes neuronales.

Cuando se trabaja con sensores virtuales es importante tener en cuenta la dependencia de los VS con respecto a los sensores físicos, lo cual conlleva a que estos primeros también degraden su respuesta, añadiendo a esto la dificultad de identificar situaciones anormales. En ocasiones se puede optar por establecer umbrales de permisibilidad para la predicción de la variable de salida, a fin de detectar errores o alguna situación anormal (Kaneko, Arakawa, & Funatsu, 2009).

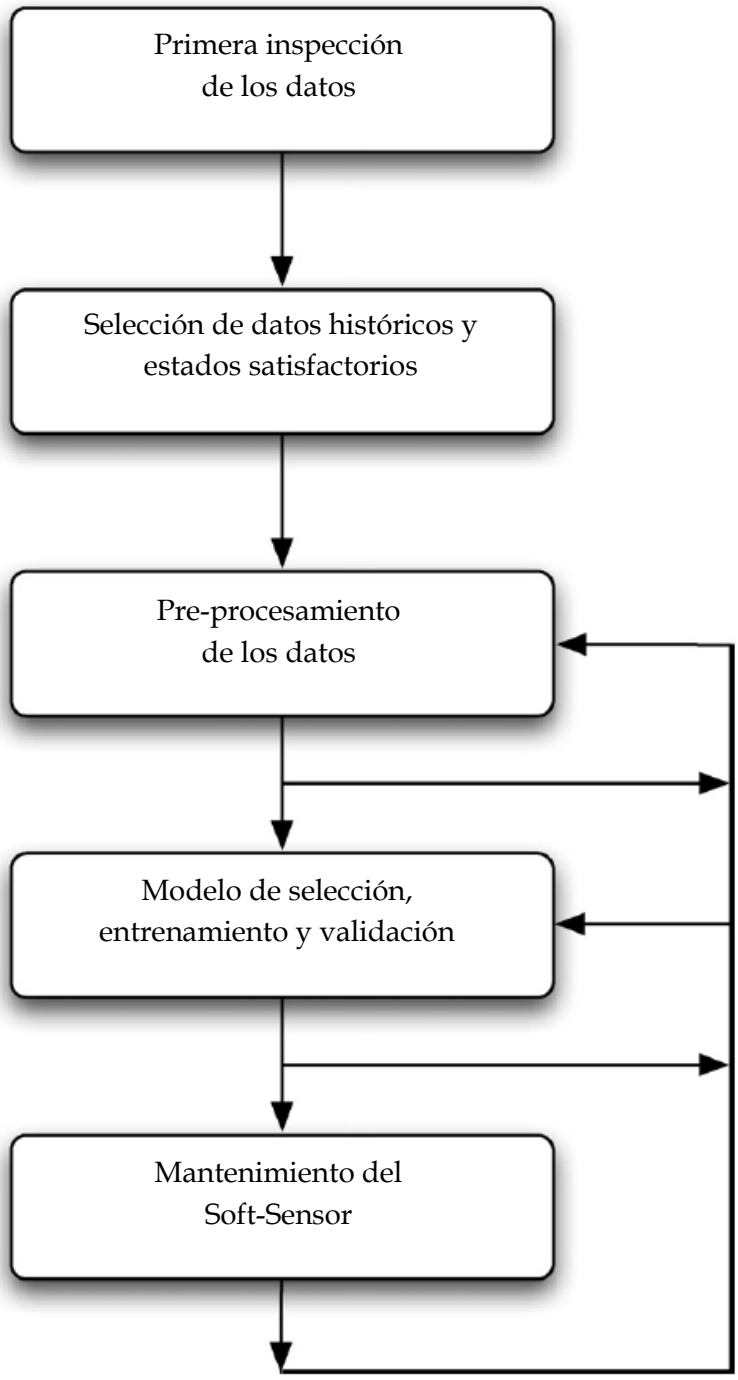


Figura 11. Metodología para el desarrollo de Soft-Sensor

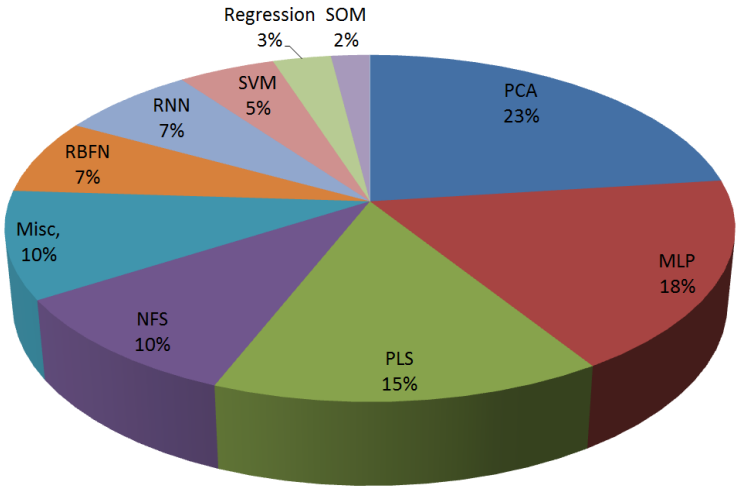


Figura 12. Distribución de los métodos de aprendizaje computacional en soft sensors.

En trabajos como (Chen & et. al., 2005; Mielle, Marquis, & Latrasse, 2000) se estudia de manera detallada el concepto de sensores virtuales. Se encuentra que en la mayoría de los casos se realizan experimentación y validación de los modelos propuestos; pero se encuentran falencias en los métodos de calibración de sensores, puesto que no se especifican procedimientos ni metodologías para lograr que las mediciones sean reproducibles.

Existen otros trabajos donde se analiza el desempeño de sensores virtuales, como en (Schütze, Gramm, & Rühl, 2004) donde se presenta un sistema multisensorial de tipo virtual, en el cual aplican una modulación en temperatura para optimizar la respuesta de los sensores, el cual también permite realizar tareas de calibración de la respuesta de los sensores. Por otro lado, en (Mielle, Marquis, & Latrasse, Electronic noses: specify or disappear , 2000) se discute sobre las limitaciones para el estudio del modelado de un único sensor que hace parte de una matriz de sensores, pero los autores muestran que en la mayoría de los casos, una matriz virtual (un único sensor modulado en temperatura) puede remplazar una matriz física de forma satisfactoria. Sin embargo, para poder hacer aplicable esto, se requiere calcular una función de transferencia para cada químico que se desee analizar y para cada temperatura, debido a que en este caso los cambios de temperatura afectarían drásticamente la respuesta del sensor (Mielle & Marquis, 2001).

En relación con la creación de sensores virtuales, en (Ibargüengoytia & Reyes, 2006) se muestra la construcción de sensores de temperatura virtuales utilizando razonamiento probabilístico, basándose en las lecturas de varios sensores reales, para posteriormente obtener un modelo que se entrena de forma automática con un algoritmo basado en redes bayesianas. Después el VS se ejecuta en línea para la estimación del correspondiente valor. En este trabajo se usaron datos reales de los sensores de temperatura de una turbina de gas en una central eléctrica. Con todas las lecturas, una variable que se supone ausente es estimada, de modo que la evaluación y la estimación se pueden realizar. Los autores

reportan que el prototipo construido fue probado con resultados prometedores, pero los resultados mostrados fueron muy pocos (solamente 10), en donde siete de los 10 casos, tuvieron una precisión superior al 88%, y en los tres casos restantes, la posterior distribución de probabilidad fue amplia, con precisiones entre el 37% y el 67%, que según los autores se debe a la perdida de información en el proceso de discretización de los datos.

Un tipo de sensor que ha sido modelado utilizando regresiones basadas en vectores de soporte se estudia en (Gu & Wang, 2008), donde se propone un algoritmo distribuido SVR para redes de sensores inalámbricos. La computación distribuida del algoritmo propuesto consiste en tres componentes principales, a saber, el uso de las funciones del kernel con soporte finito, el enfoque de gradiente descendiente para el aprendizaje SVR y el uso de un algoritmo especializado para evaluar o validar la función de regresión aprendida. Para este algoritmo distribuido SVR sólo es necesario el intercambio de información entre los nodos de sensor vecino y la información global puede ser difundida en toda la red a través de los intercambios de información local. Además, es escalable a medida que la adición de más nodos no afecta el rendimiento de los algoritmos; es robusto, ya que puede producir los resultados deseados, incluso si ocurren fallas en algunos de los nodos. Un campo que queda abierto es la intención de aplicar este algoritmo a una red de sensores prácticos para investigar su funcionamiento.

En esta misma vía y como referente para trabajos posteriores se destaca el trabajo presentado en (Chen , Song , & Li, 2005) donde se propone el análisis de componentes independientes dinámicas (DICA). Este método es capaz de extraer las características dinámicas importantes del proceso, para encontrar componentes estadísticamente independientes de las entradas de auto-correlación y correlación cruzada. Para hacer frente a la estimación por regresión, se combinan DICA con la regresión de soporte vectorial (SVR) para construir varias capas de regresión. La primera capa es la extracción de características que tiene la ventaja de un rendimiento sólido y la reducción de la complejidad del análisis. La segunda capa es la SVR que hace la estimación. Este tipo de *soft-sensor* estimador se aplica a la estimación del proceso de composición en el punto de referencia de simulación de la planta Eastman Tennessee (TE), para las variables temperatura, presión, nivel, taza de flujo. Los resultados de la simulación muestran claramente que el estimador para la extracción de características utilizando DICA puede tener un mejor desempeño que sin la extracción de características, y con otros métodos estadísticos para la extracción de características, como el PCA, ICA o DPCA.

En (Liu & et. al., 2009) se realizó una revisión en donde se mencionan los problemas de los sistemas de olfato electrónico, el estado actual y futuros trabajos; además se discute el desarrollo de nuevas tecnologías de utilización de sensores para obtener la información exacta y deseada más efectivamente, independiente de las posiciones limitadas de un sensor. En este estudio también se examinan las aplicaciones de sensores virtuales en diferentes campos, como transporte, comunicaciones inalámbricas, redes de sensores y control activo de ruido.

En (Liu , Gao , & Chen, 2013) se presenta un sistema online para predicción de la calidad de un reactor secuencial multigrado (SRMG) empleado en procesos químicos en los cuales a menudo se encuentran diferentes cambios, el cual adicionalmente está relacionado con procesos no lineales, selección y extracción de variables de entrada, relación secuencial en reactores y múltiples grados en una línea de producción. La predicción se realiza a través de un novedoso método de modelamiento secuencial no lineal *just-in-time*, propuesto por dicho autores. En donde se integra la selección y extracción de las variables de entrada en un marco unificado. En primer lugar, las variables en los reactores previos son sustituidas por variables de caridad "virtual" a través del método *least squares support vector regression (LSSVR) transform models*. Empleando el modelo secuencial global LSSVR es posible capturar la relación secuencial en un proceso de reactor secuencial usando una estrategia de entrenamiento eficiente. Adicionalmente los autores también proponen un modelo secuencial just-in-time (JS-LSSVR) el cual puede ser aplicado para la predicción online de plantas similares que trabajen con reactores secuenciales.

En (Zhang & Liu, 2013) se propone un sensor virtual adaptativo para monitoreo *online* del índice de fusión (MI), una variable importante que determina la calidad del producto en los procesos industriales de polimerización de propileno (PP). El *soft sensor* presentado obtiene el modelo empleando la técnica de aprendizaje de máquina neuro-fuzzy adaptativa (A-FNN), debido a que utilizando solamente la técnica neuro-fuzzy (FNN) en este caso conlleva a una mayor dificultad en la determinación de la estructura y la definición del número de reglas *fuzzy* no puede ser adecuada u optima posteriormente debido a la naturaleza cambiante del sistema. En vez de esto la técnica neuro-fuzzy adaptativa permite la modificación de la estructura del modelo teniendo como base uno umbrales predefinidos. Además, con el objetivo de obtener mejor generalización del sensor virtual, se empleó regresión de soporte vectorial (SVR) para la sintonización de parámetros, en donde la función de salida es transformada en un problema de optimización basado en SVR. Según los resultados mostrados en (Zhang & Liu, 2013) con las fusión de los métodos neuro-fuzzy y SVR se obtienen mejores resultados que con cada método por separado.

1.5 INFERENCIA ESTADÍSTICA

Este concepto estadístico comprende los métodos y procedimientos para deducir propiedades de una población a partir de una pequeña parte de la misma (Espejo Miranda, y otros, 2007). La inferencia estadística se basa en el conocimiento de las principales características de una determinada población (por ejemplo, media, desviación típica, su estructura probabilística, entre otras), con el objetivo de analizar y extraer conclusiones de las características y el comportamiento que afecta a todos los elementos de dicha población (Arriaza Gómez, y otros, 2008).

Los métodos estadísticos de aprendizaje de máquina van desde el cálculo de medias hasta la construcción de modelos complejos como las redes bayesianas o las redes neuronales. Tienen aplicación en informática, ingeniería, neurobiología, psicología, entre otras

ciencias. Hoy en día, el aprendizaje estadístico es un área de investigación muy activa y se han hecho enormes avances tanto en la teoría como en la práctica, hasta el punto que es posible aprender casi cualquier modelo de un variable, fenómeno o sistema y se pueden realizar tareas de inferencia aproximada o exacta (Valle Padilla, 2010).

Un modelo inferencial se construye entre las variables que son fáciles de medir en línea y una que es difícil de medir o que su medición incrementa los costos y la complejidad del sistema tal y como es el caso objeto de estudio de esta tesis de maestría, donde el valor de una variable objetivo es estimado por el modelo.

Los métodos empleados para realizar inferencia estadística están relacionados principalmente con mínimos cuadrados parciales (PLS), en particular este método fue empleado junto con el análisis de componentes independientes (ICA) para crear sensores virtuales en (Kaneko, Arakawa, & Funatsu, 2009). Otros métodos empleados para la inferencia son, componente principal de regresión (PCR), método PLS no lineal, redes neuronales artificiales (RNA) y recientemente las máquinas de vectores de soporte para regresión (SVR), sobre este último no se encontraron referencias anteriores a este trabajo, relacionadas con la modelación de sensores de gases.

Mediante el uso de VS, un valor de las variable objetivo puede estimarse con alta precisión (Kaneko, Arakawa, & Funatsu, 2009). Por ejemplo, en (Liu & et. al., 2009) se presenta una revisión muy interesante sobre los métodos existentes para inferir información con sensores virtuales.

Cabe recordar las desventajas de los VS: dificultades en la predicción precisa en ambientes o plantas cambiantes, la dependencia de los VS con respecto a los sensores físicos lo cual conlleva a que estos primeros también degraden su respuesta. Sin embargo, no dejan de ser muy buena alternativa, para reducir costos, tamaño y complejidad del hardware.

1.6 REGRESIÓN

La regresión lineal es una técnica estadística utilizada para estudiar la relación entre variables. Se puede utilizar para caracterizar la relación entre variables físicas o para calibrar medidas de sensores o instrumentos (Cea D´Ancona M. Á., 2006).

Se suele hablar de regresión simple cuando están involucradas 2 variables o regresión múltiple cuando están involucradas más de 2 variables. En ambos casos la regresión lineal se puede utilizar para establecer la relación entre una determinada variable dependiente o de salida (y) y una o más variables independientes, también conocidas como variables predictores o entradas ($x_1, x_2,\ldots, x_n$), (Cea D´Ancona M. Á., 2006). El objetivo es obtener una función que modele la relación entre las variables predictores o entradas y la variable de salida. Una función o modelo de regresión puede tener la forma mostrada en la siguiente ecuación:

$$y = B_0 + B_1 x + \delta \quad \textbf{(4)}$$

En donde, x representa las entradas o variables predictores, mientras y representa la variable de salida del predictor, los parámetros B_0 y B_1 corresponden a los términos que se deben ajustar o encontrar para obtener un modelo de regresión o predictor adecuado y δ es el error del modelo de regresión debido a la afectación del ruido o de la presencia de variables aleatorias (**Figura 13**).

Los métodos de regresión se pueden clasificar en 2 grandes clases, métodos de regresión lineal y métodos de regresión no lineal.

1.6.1 Regresión lineal

La relación más simple entre dos variables es una línea recta, el tipo de función lineal más sencillo es el mostrado en la Ecuación **(4)**. Otros tipos de funciones de tipo lineal se muestran a continuación:

$$y = B_0 + B_1 x + B_2 x^2 \quad \textbf{(5)}$$
$$y = B_0 + B_1 x + B_2 x^2 + B_3 x^3 \quad \textbf{(6)}$$
$$y = B_0 + B_1 x + B_2 z + B_3 xz \quad \textbf{(7)}$$

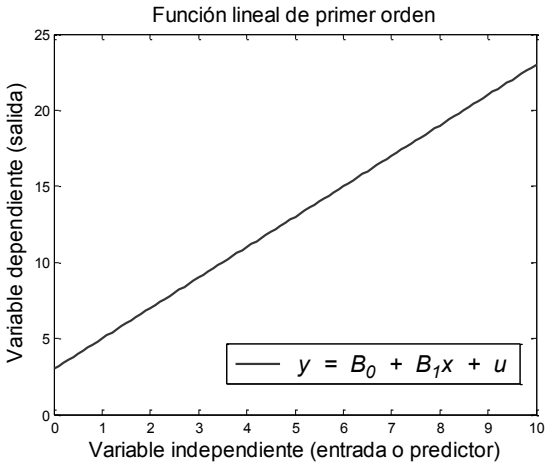


Figura 113. Gráfica de la función lineal $y = 3 + 2x$

Los modelos o funciones mostrados en las ecuaciones **(5) y (6) (**ver **Figura 14),** al igual que la ecuación **(4),** tienen una sola variable dependiente o salida y una variable predictor o entrada (pero con exponentes diferentes de uno), por lo tanto se conocen como modelos de segundo orden y tercer orden respectivamente, siendo también de tipo lineal ya que los parámetros B_0 y B_1 son lineales (exponente uno). Por su parte, el modelo mostrado en la ecuación **(7)** es lineal de primer orden pero con dos entradas o predictores (Devore, 2008).

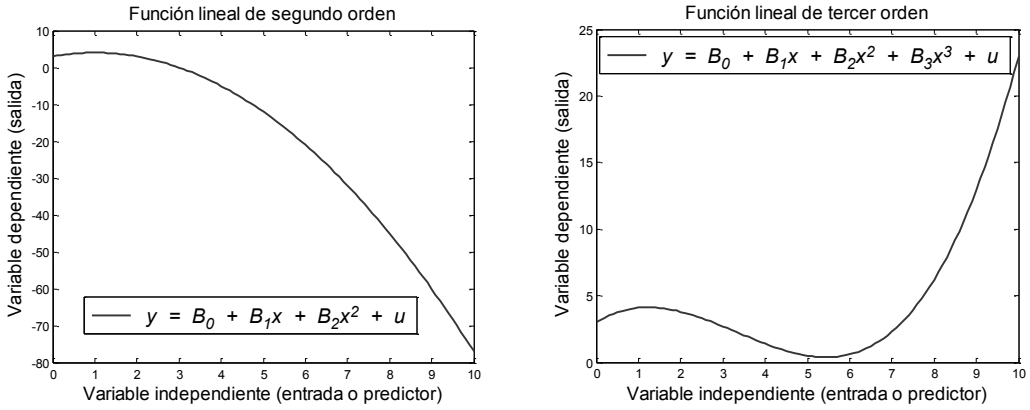


Figura 124. *Izquierda:* Gráfica de la función lineal $y = 3 + 2x - 1x^2$
Derecha: Gráfica de la función lineal $y = 3 + 2x - 1x^2 + 0.1x^3$

Como se mencionó en anteriormente, el objetivo de la regresión es encontrar alguna función que describa aproximadamente la relación entre una o más variables de entrada y otra variable de salida. Para tal efecto se debe elegir la clase de función (lineal, polinomial de orden *n*, potencial, exponencial, logarítmica, etc.), seleccionando alguna clase que permita modelar la situación o problema. Una vez elegida la clase de funciones $C = \{f(x_i, B) | B \in A \subseteq \Re^n\}$ se debe determinar alguna que describa los valores dados. Para tal propósito se necesita un criterio, por ejemplo, el método de mínimos cuadrados (Minnaard, 2010).

Para aplicar el método de mínimos cuadrados, se debe considerar que $f(x_i, B)$ es una función cualquiera de la clase *C*, entonces, para cada valor evaluado en dicha función el error δ_i, será igual a la diferencia entre el valor observado y el obtenido a través de la expresión:

$$\delta_i = Y_i - f(x_i, B) \quad \textbf{(8)}$$

Entonces se intenta buscar la función que minimice el error cuadrático, de forma que la función elegida será la que permita acercar a cero el valor de S en la ecuación **(9)**, lo cual ocurre para un determinado conjunto de parámetros B en dicha función (Freund, Miller, & Miller, 2000; Devore, 2008).

$$S = \sum_{i=1}^{n} {\delta_i}^2 \quad \textbf{(9)}$$

Figura 135. Pasos generales para realizar una regresión

1.6.2 Regresión no lineal

En muchos casos se pueden obtener modelos de las variables aplicando métodos de regresión lineal con muy buenos resultados, pero en otras ocasiones estos métodos no son muy apropiados, haciéndose necesario probar la linealidad de la curva de regresión mediante el método de análisis de la variancia, donde el resultado obtenido permitirá deducir la conveniencia de usar los métodos de regresión no lineal (Devore, 2008).

El análisis de varianza ANOVA corresponde al conjunto de situaciones experimentales y procedimientos estadísticos que permiten el análisis de respuestas cuantitativas de datos o conjuntos de datos experimentales (Devore, 2008).

Cuando se quiere establecer o cuantificar la intensidad de la relación lineal entre dos variables, conviene obtener el parámetro que nos da tal cuantificación; algunos de ellos son: la razón de correlación η^2, el coeficiente de correlación φ, el coeficiente de correlación biserial puntual r_{bp}, el coeficiente de correlación de rangos de Spearman ρ o el coeficiente de correlación tetracórica r_t, entre otros (Vila, Sedano, López, & Juan, 2006; Álvarez González, 2008). La mayoría de estos parámetros se basan en el coeficiente de correlación lineal de Pearson *R*, expresado en la ecuación **(10)**, cuyo valor oscila entre –1 y +1 (ver **Figura 16**).

$$R = \frac{\mathrm{Cov}(x,y)}{S_x S_y} \qquad \textbf{(10)}$$

Donde $\mathrm{Cov}(x,y)$ es la covarianza entre x y y, mientras que S_x y S_y son la desviación estándar de x y de y respectivamente.

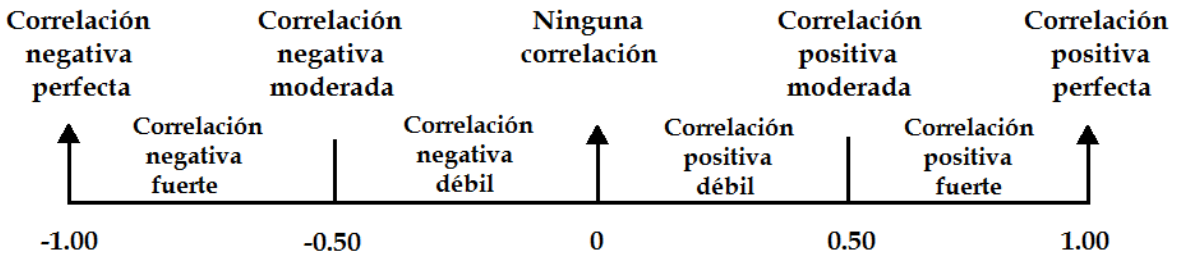


Figura 146. Diagrama de interpretación del coeficiente de correlación de Pearson. Imagen tomada de (Vila, Sedano, López, & Juan, 2006)

Si después de realizado el análisis de la varianza, se obtiene que el coeficiente de correlación *R* es 1 o cercano a 1, tanto positivo como negativo, se dice que las variables están correlacionadas o relacionadas linealmente, caso en el cual no sería necesario utilizar métodos de regresión no lineales. Pero, si el coeficiente de correlación *R* es cero o cercano a cero, se dice que entre las variables no hay correlación lineal y por lo tanto de debe aplicar otros métodos de regresión no lineales. Es importante anotar que la existencia de correlación entre variables no implica causalidad (Vila, Sedano, López, & Juan, 2006).

Cómo se mencionó en la sección 1.6.1 los parámetros B (valores poblacionales desconocidos) son lineales para los modelos de regresión lineal, pero por el contrario en los modelos no lineales estos parámetros aparecen de forma no lineal en la función, como las que se muestran en las siguientes expresiones:

$$y = B_0 e^{(-e^{(-B_0+B_1 x)})} \quad \textbf{(11)}$$

$$y = \frac{B_0}{1+B_1 e^{(-B_2 x)}} \quad \textbf{(12)}$$

$$y = \frac{B_0}{\left[1+e^{-(B_1+B_2 x)}\right]^{B_3}} \quad \textbf{(13)}$$

Las funciones mostradas en **(11), (12)** y **(13)** toman el nombre de modelo de Gompertz, modelo logístico y modelo de Richards, respectivamente, y son comúnmente usadas para describir modelos de crecimiento (Macchiavelli, 2010).

Otra función no lineal comúnmente usada en regresión es la exponencial:

$$y = B_0 + B_1 e^{-B_2 x} \quad \textbf{(14)}$$

En general, todas las funciones de tipo no lineal para regresión, también deben incluir un término δ de error debido a la afectación del ruido de la variabilidad de los datos o de efectos de tipo aleatorio, de esta manera la función de la **ecuación (14)** puede rescribirse como, $y = B_0 + B_1 e^{-B_2 x} + \delta$ para introducir el término del error o de aleatoriedad presente en la mayoría de los fenómenos físicos.

Una vez escogida la clase de función para modelar la relación entre las variables, lo siguiente es encontrar los parámetros y ajustar la función que mejor represente el modelo, para ello los métodos y procedimientos de ajuste son similares a los empleados en el caso de la regresión lineal.

1.7 MÁQUINAS DE VECTORES DE SOPORTE

Las máquinas de vectores de soporte (SVM) fueron desarrolladas por Vapnik (1995) y han ganado popularidad debido a muchas características atractivas y rendimientos prometedores (Gunn, 1998).

Las SVM inicialmente fueron desarrolladas para resolver problemas de clasificación, pero recientemente su aplicación se ha extendido a los problemas de regresión. Comúnmente el termino SVM se utiliza para referirse a los métodos para clasificación y regresión, aunque se han introducido dos siglas para diferenciar la una de la otra, la SVC para referirse a métodos basados en máquinas de vectores de soporte para clasificación y SVR para referirse a métodos basados en máquinas de vectores de soporte para regresión (Gunn, 1998).

Las SVM de tipo no lineal se basan en la construcción de un mapeo dentro de un espacio de características de alta dimensionalidad mediante el uso de funciones kernel (Gunn, 1998). En otras palabras los datos de entrada son transformados implícitamente a un nuevo espacio, generalmente de una dimensión superior, en donde se debe diferenciar entre el espacio inicial o espacio de entradas (*input space*) y el espacio que contiene los espacios transformados o espacio de características (*feature space*) (Gómez Morales & Hernández, 2009).

La idea de la función kernel es permitir que las operaciones se realicen en el espacio de entrada en lugar que en el espacio de características, el cual es potencialmente de alta dimensionalidad (Gunn, 1998). De ahí, que no sea necesario estimar el producto interno en el espacio de características, haciéndolo menos pesado en procesamiento de cómputo.
Los kernel comúnmente usados son el kernel lineal, el kernel polinomial y el kernel gaussiano, a continuación se referencia en que consiste cada una de estas funciones.

Kernel lineal.
Es la función de kernel más simple (Ecuación **15**). Los algoritmos usando una función de kernel lineal son a menudo equivalentes a su contraparte de no utilizar kernel. Esta dado por el producto punto en el espacio de entrada $<x,x'>$ mas una constante opcional C.

$$k(x,x') = \langle x,x' \rangle + C \qquad \textbf{(15)}$$

Kernel polinomial.
El kernel polinomial es un kernel no estacionario. Son muy adecuados para problemas en los que se normaliza todos los datos de entrenamiento. Representa la expansión a todas las combinaciones de monomios de orden d, y está dado por:

$$k(x,x') = \langle x,x' + C \rangle^d \qquad \textbf{(16)}$$

El parámetro constante C y el grado del polinomio d deben ser ajustados.

Kernel gaussiano

El kernel gaussiano es un ejemplo de las funciones kernel de base radial, está dado por:

$$k(x,x') = e^{\left(-\frac{\|x-x'\|^2}{2\sigma^2}+c\right)} \qquad \textbf{(17)}$$

El parámetro ajustable sigma (σ) tiene un papel importante en el rendimiento del kernel y debe ser cuidadosamente sintonizado para el problema en cuestión. Si este parámetro es sobreestimado, la exponencial se comportara de forma casi lineal y la proyección de dimensiones superiores comenzara a perder su poder no-lineal. Por otro lado, si el subestimado, la función carecerá de regularización y la frontera de decisión será muy sensible al ruido en los datos de entrenamiento (Sánchez, Osorio, & Suárez, 2008); (Álvarez, Fetecua, Orozco, & Castellanos, 2010).

Sin embargo, la complejidad computacional de las SVM también depende del número de patrones de entrenamiento, teniendo en cuenta que para proporcionar una buena distribución del conjunto de datos en un problema de alta dimensionalidad, habitualmente se requiere un conjunto grande de datos de entrenamiento (Gunn, 1998).
Las SVM implementan el principio de minimización del riesgo estructural (R_{est}), el cual busca minimizar un límite superior del error de generalización en vez de la minimización del riesgo empírico (R_{emp}), que es el que busca minimizar el error de entrenamiento, generalmente empleado en las redes neuronal (RNA). La minimización del riego estructural se basa en que el error de generalización está acotado por la suma del error de entrenamiento y un término de intervalo de confianza que depende de la dimensión de Vapnik-Chervoneskis (Fossi & D'Ambrosio, 2004).

El riesgo empírico se define como la medida del error promedio en el conjunto de datos de entrenamiento y está dada por:

$$R_{emp}(\alpha) = \frac{l}{2l}\sum_{i=1}^{l}|y_i - f(x_i,\alpha)| \qquad \textbf{(18)}$$

La probabilidad de distribución no aparece en la Ecuación **(18)** y el valor de R_{emp} está fijado para un α (etiquetas) y un conjunto de datos de entrenamiento $\{x_i,y_i\}$ en particular. Entonces el límite superior del riesgo esperado o riesgo actual establecido por Vapnik está definido por:

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h\left(\log\left(\frac{2l}{h}\right)+1\right)-\log\left(\frac{\eta}{4}\right)}{l}} \qquad \textbf{(19)}$$

Donde $\eta = P\{ y = 1 \}$, h es la dimensión VC de $\{f(x, \alpha)\}$ y l es la cantidad de datos de entrenamiento (Borovikov, 1999).

Encontrar un modelo generalizado para los problemas de reconocimiento de patrones es importante y es lo que se quiere lograr, pero se debe tener en cuenta que ello no implica que el error de entrenamiento tenga que ser cero, sin embargo en la mayoría de los casos se busca que sea cercano a cero para los datos de entrenamiento y validación (W. Duin & Pękalska, 2012). Un valor para el riesgo empírico R_{emp} pequeño (cercano a cero) (Koltchinskii, 2009), no implica necesariamente un riesgo real esperado pequeño (Schlesinger, 2012), puede suceder que los datos queden poco ajustados (sub-entrenamiento), sobre-ajustados (sobre-entrenamiento) o bien ajustados (caso ideal), tal y como se ilustra en la **Figura 17**.

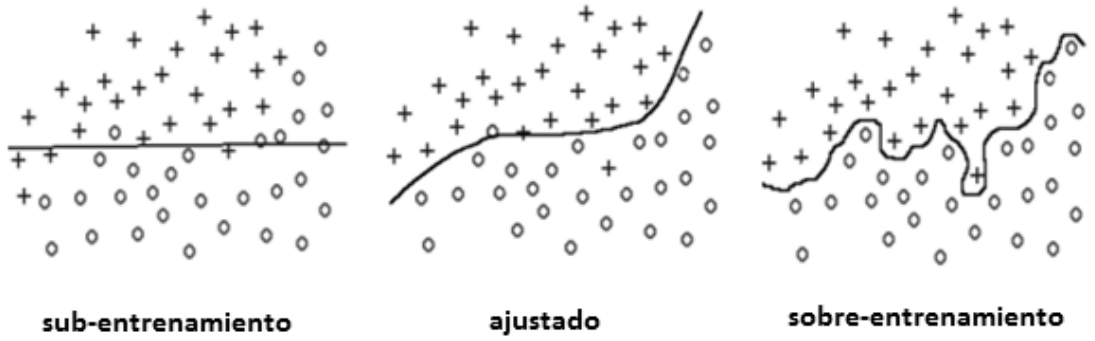


Figura 157. Diferentes tipos de ajuste de datos. Imagen tomada de (Schlesinger, 2012)

1.7.1 Máquinas de Vectores de Soporte para Clasificación

En general, los problemas de clasificación centran su interés en conocer si una determinada muestra pertenece a una u otra clase, teniendo en cuenta las características de entrada. Sin embargo, desde el punto de vista del entrenamiento de máquinas de aprendizaje, la clasificación consiste en encontrar una regla de decisión, tal que dada una muestra externa, sea asignada a su clase correspondiente. La búsqueda de una regla de decisión adecuada puede interpretarse como la estimación de una función *f*, que asigna a cada punto del espacio de observación, un punto en el espacio de alguna de las clases posibles. Esta búsqueda se lleva a cabo usando un conjunto de datos de entrenamiento formado por *l* muestras etiquetadas, distribuidas generalmente mediante una distribución de probabilidad desconocida $P(x, y)$ (Valle Padilla, 2010).

En lo que respecta a las máquinas de vectores de soporte para clasificación o SVC, la regla de decisión o función de decisión permite obtener el híper-plano óptimo de separación (línea más gruesa en la **Figura 18**), para el cual es importante establecer la dimensión VC. En la teoría del aprendizaje estadístico, la dimensión VC (Vapnik-Chervoneskis), especificada como el escalar *h*, es una medida de la capacidad de un algoritmo de clasificación estadística, entendida como la cardinalidad del mayor conjunto de puntos que el algoritmo puede separar (Burges, 1998).

La capacidad de un modelo de clasificación está relacionada directamente con lo complicado que puede llegar a ser, entiéndase complicado como la cantidad de vectores de soporte calculados para logar un nivel de confiabilidad o porcentaje de acierto aceptable (Ben-Hur & Weston, 2010). En efecto, un modelo de clasificación f con algún vector de parámetros θ, se dice que separa o divide un conjunto de puntos de datos (x_1, x_2,…, x_n), si para todas las asignaciones de etiquetas en esos puntos, existe un θ tal que el modelo f hace que el error sea cero (0), cuando se evalúan dichos puntos del conjunto de datos (Burges, 1998).

Figura 168. Híper-plano óptimo de separación. Imagen de ejemplo tomada de (Gunn, 1998)

Por lo tanto, la dimensión VC es el máximo número de datos del conjunto de datos que se debe utilizar para entrenar adecuadamente un clasificador, con el fin de obtener un modelo de clasificación tendiendo al óptimo. En otras palabras, la dimensión VC establece el número de datos adecuados para obtener un conjunto de datos de entrenamiento separable (Burges, 1998).

Generalizando, la dimensión VC indica la dimensión del híper-plano separador, que para un espacio de entradas n-dimensional, el valor sería de $h=n+1$. Por lo tanto, como en el ejemplo anterior, el espacio de entrada es de $n=2$ (2 características), entonces $h=3$, por lo que existe al menos un conjunto de puntos h (3 datos) en el espacio de entrada que pueden ser separados, sin embargo, esto no quiere decir que todos los conjuntos de puntos h en el espacio de entrada, puedan ser separados por una determinada función, lo cual se puede mejorar utilizando otro tipo de función para obtener una dimensión del híper-plano separador de mayor orden, que podría llegar a ser más conveniente (Veganzones, 2009).

1.7.2 Máquinas de Vectores de Soporte para Regresión

En los problemas de regresión lo que se pretende es estimar una variable desconocida a partir de observaciones que guardan algún tipo de relación con ella (Valle Padilla, 2010). Las máquinas de vectores de soporte también pueden ser aplicadas a problemas de regresión, introduciendo una función alternativa de pérdida. La función de pérdida debe

ser modificada para incluir una medida de distancia, donde las más utilizadas son: cuadrática, *Huber*, *Least Modulus* y ε-*Insensitive* (Gunn, 1998).

Una SVM para regresión o SVR estima una función usando un conjunto de otras funciones lineales definidas en un espacio híper-dimensional (Fossi & D'Ambrosio, 2004). Por lo tanto, para un conjunto de datos $\{x_i,d_i\}$, donde x_i es el vector de entrada, d_i es la salida deseada, la SVR aproxima la función de regresión usando:

$$f(x)=w\Phi(x)+b \qquad \textbf{(20)}$$

Donde $\Phi(x)$ es el espacio de rasgos híper-dimensional al cual se proyecta (no linealmente) el espacio de entrada x y los coeficientes w y b se estiman bajo el criterio de minimización del riesgo empírico **Ecuación (21)**. El primer término de la ecuación 21 es el error empírico (riesgo) y el segundo término es la constante de regularización. En esta ecuación existe una dependencia de la constante de regularización C y del termino épsilon ε, conocido como el tamaño del cilindro híper-dimensional, el cual equivale a la exactitud de aproximación con respecto a los datos de entrenamiento (Fossi & D'Ambrosio, 2004)**.**

$$R_{SVM}(C) = C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon(d_i,y_i) + \frac{1}{2}\|w\|^2 \qquad (\mathbf{21})$$

En donde,

$$L_\varepsilon(d_i,y_i) = |d-y| - \varepsilon \; |d-y| \geq \varepsilon \text{ y } 0 \text{ en el resto de los casos}$$

Como los parámetros C y ε son de vital importancia en el diseño de la SVR, deben ser escogidos cuidadosamente para obtener un predictor de buen desempeño, pero adicionalmente, el kernel seleccionado también juega un papel fundamental a la hora obtener el modelo predictor o máquina de regresión, como también se conoce.

A fin de estimar los parámetros w y b para la SVR, se utiliza una función de pérdida con alguna medida de distancia, por ejemplo, la ε-insensible. En ese caso la **ecuación (21)** se transforma en la **ecuación (22)**, usando las variables auxiliares ξ_i y $\xi_i^{(*)}$ que representan los límites superior e inferior respectivamente, en la salida del sistema (Gunn, 1998; Fossi & D'Ambrosio, 2004), como se ilustra en la **Figura 19**.

$$R_{SVM}\left(w,\xi^{(*)}\right) = C\sum_{i=1}^{l}\left(\xi_i+\xi^{(*)}\right) + \frac{1}{2}\|w\|^2 \qquad (\mathbf{22})$$

Tomando en cuenta las siguientes restricciones:

$$d_i - w\Phi(x) - b_i \leq \varepsilon + \xi_i$$

$$w\Phi(x) + b_i - d_i \leq \varepsilon + \xi_i^{(*)} \quad , \quad \xi_i^{(*)} \geq 0$$

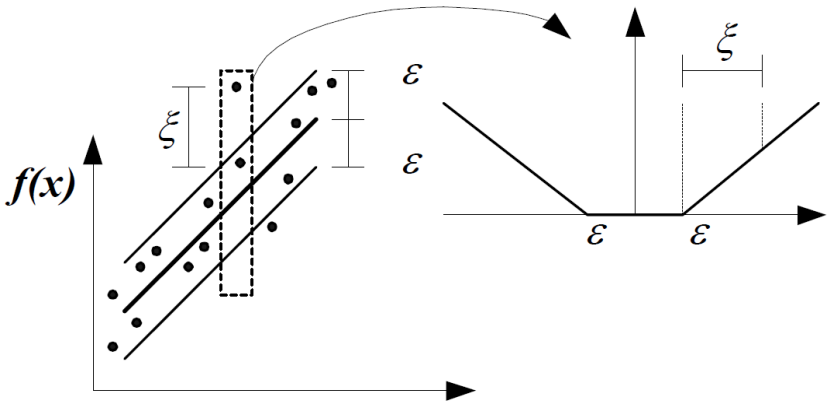


Figura 179. Error pre-establecido ε y los límites ξ de la función ε-insensible para una SVR

Introduciendo los multiplicadores de Lagrange a_i y a_i^*, la función de regresión dada en la **ecuación (20)** queda como se muestra a continuación:

$$f(x,a_i,a_i^*) = \sum_{i=1}^{n}(a_i - a_i^*)K(x_i,x_j) + b \qquad (23)$$

Donde $K(x_i,x_j)$ es la función kernel y a_i,a_i^* son los multiplicadores de Lagrange, que satisfacen las siguientes restricciones:

$a_i * a_i^* = 0\,,\ a_i \geq 0$
$a_i^* \geq 0\,,\ i = 1,2,\ldots\ldots,l$

Los cuales se calculan maximizando la función de riesgo dada en la **ecuación (21)**, resultando en la siguiente expresión:

$$R(a_i,a_i^*) = \sum_{i=1}^{l} d_i\,(a_i - a_i^*) - \varepsilon\sum_{i=1}^{l}(a_i + a_i^*) - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(a_i - a_i^*)(a_j - a_j^*)K(x_i,x_j) \qquad (24)$$

$$\sum_{i=1}^{l}(a_i - a_i^*) = 0 \qquad 0 \leq a_i \leq C, \quad 0 \leq a_i^* \leq C, \quad i = 1,2,\ldots\ldots,l$$

Solo algunos de estos multiplicadores de Lagrange (basados en las condiciones de programación cuadrática de Karush-Kuhn-Tucker) tendrán valores diferentes de cero y los puntos o vectores de datos asociados tendrán errores de aproximación iguales o mayores a ε, lo que significa que están fuera del límite ε de la función de decisión *f(x)*, y estos son los que finalmente se denominan vectores de soporte de la SVM (Fossi & D'Ambrosio, 2004; Gunn, 1998).

2. DISEÑO EXPERIMENTAL

2.1 BASES DE DATOS

Las bases de datos empleadas en el desarrollo de este trabajo de investigación se detallan en la **Tabla 2,** estas bases de datos son reales o experimentales, lo que quiere decir que fueron tomadas con sensores físicos. Dichas bases de datos fueron facilitadas por el grupo de Investigación en Sistemas Multisensoriales y Reconocimiento de Patrones de la Universidad de Pamplona - Colombia (UP).

NOMBRE	ACCESO	TIPO	TAMAÑO	NÚMERO SENSORES	UBICACIÓN/ CONTACTO
A-NOSE	Privada (Accesible con autorización)	Real	120 archivos de 600 muestras cada uno	8	Tomados en la Universidad de Pamplona (Colombia), Instituto de investigación y desarrollo de tecnología avanzadas. Juan Carlos Rodriguez Gamboa ju4n@unipamplona.edu.co
B-NOSE	Privada (Accesible con autorización)	Real	125 archivos de 630 muestras cada uno	16	Tomados en la Universidad de Pamplona (Colombia), Instituto de investigación y desarrollo de tecnología avanzadas. Cristhian Manuel Duran Acevedo cmduran@unipamplona.edu.co

Tabla 2. Bases de datos empleadas con medidas de sensores de gases

Datos importantes de las bases de datos experimentales A-NOSE y B-NOSE se encuentran en la **Tabla 3** y **Tabla 4**, respectivamente, en donde se detallan las referencias de los sensores utilizados en cada una de las correspondientes bases de datos, así como la aplicación recomendada, el tipo de gas al cual presenta mayor sensibilidad y la ubicación de los sensores dentro de cada matriz.

El equipo empleado y los procedimientos que se utilizaron para adquirir la base de datos A-NOSE se encuentran detallados en (Rodriguez Gamboa & Durán Acevedo, 2008), cabe mencionar que este equipo tiene una cámara de sensores herméticamente cerrada con capacidad para albergar 8 sensores de gases en un volumen de 0.00036 m^3 aproximadamente y una cámara de concentración con un volumen aproximado de 0,00338 m^3, las señales de sensores fueron tomadas con una tarjeta de adquisición de datos conectada a un computador a través de un puerto USB y se adquirieron a través de una aplicación desarrollada en el software Labview versión 8.2.

El equipo empleado y los procedimientos que se utilizaron para adquirir la base de datos B-NOSE se encuentran detallados en (Gualdrón G., Durán, Isaza, Carvajal F., & Uribe, 2011), cabe mencionar que este equipo tiene una cámara de sensores con capacidad para albergar 16 sensores de gases en un volumen de 0.00040 m^3 aproximadamente y una cámara de concentración con un volumen aproximado de 0,0017 m^3, las señales de sensores fueron tomadas con una tarjeta de adquisición de datos conectada a un computador a través de un puerto USB y se adquirieron a través de una aplicación desarrollada en el software Matlab versión 7.5.

Referencia del Sensor de gas	Aplicación	Tipo de gas sensible	Fabricante	Cantidad	Ubicación del sensor en la Matriz
SP-12A	Detección de gases inflamables	Metano	FIS	1	S1
SP31	Detección de disolventes orgánicos	Propósito general	FIS	1	S2
TGS-813	Detección de gases combustibles	Metano, butano y propano (hidrocarburos en general)	FIGARO	2	S3, S6
TGS-842	Detección de gases inflamables	Metano y gas natural	FIGARO	1	S4
SP-AQ3	Control de calidad del aire	Humo de cigarrillo	FIS	1	S5
ST-31	Detección de disolventes orgánicos	Propósito general	FIS	1	S7
TGS-800	Control de calidad del aire	Aire contaminado, tabaco, gasolina, (contaminantes del aire en general)	FIGARO	1	S8

Tabla 3. Configuración de los sensores de la Base de Datos A-NOSE

Cabe resaltar que la base de datos A-NOSE tiene 120 mediciones realizadas con café de diferentes tipos, algunas frutas, bebidas alcohólicas y compuestos como etanol y metanol, siendo las medidas de café las que predominan, como se puede apreciar en la **Tabla 5**. Mientras que la base de datos B-NOSE tiene 95 mediciones realizadas con diferentes compuestos químicos (Tolueno, Xileno, acetona, entre otros), como se puede apreciar en la **Tabla 6**.

Sensor de gas FIGARO	Aplicación	Tipo de gas sensible	Serie 8/ Tipo	Cantidad	Ubicación del sensor en la Matriz
TGS 826	Detección de olores	Amoníaco y aminas	Tipo C	2	S1, S4
TGS 831	Detección de gas halocarburos (gases refrigerantes)	R-22, Monoclorodifluorometano	Tipo C	2	S2,S15
TGS 821	Detección de gas combustible	Hidrógeno	Tipo C	2	S3, S12
TGS 842	Detección de gas combustible	Metano y gas natural	Tipo R	2	S5, S14
TGS 880	Control de cocción	humo de los alimentos (Alcohol, olor)	Tipo M	2	S6, S10
TGS 825	Detección de gas toxico	sulfuro de hidrógeno	Tipo C	2	S7
TGS 813	Detección de gas combustible	Hidrocarburos en general	Tipo R	1	S8
TGS 800	Control de calidad del aire	contaminantes del aire en general	Tipo R	1	S9
TGS 822	Detección de vapores de disolventes	Alcohol y disolventes orgánicos	Tipo R	1	S11
TGS 832	Detección de gas halocarburos (gases refrigerantes)	R-134A 1,1,1,2-Tetrafluoroethane	Tipo C	1	S13
TGS 830	Detección de gas halocarburos (gases refrigerantes)	R-22, Monoclorodifluorometano	Tipo C	1	S16

Tabla 4. Configuración de los sensores de la Base de Datos B-NOSE

Producto	Variedades de un mismo tipo de producto
Aguardiente	
Café con granos defectuoso	Blanqueado y vinagre, blanqueado, negro, vinagre
Café Excelso	Excelso, variedad Toledo y UGQ dos moliendas diferentes
Café pasilla	Pasilla de máquina y pasilla con químico
Durazno	
Etanol	10%, 25%, 50% y 95%
Maracuyá	
Metanol	10%, 25%, 50% y 95%
Vino	Blanco, tinto y de naranja

Tabla 5. Productos analizados con la base de datos A-NOSE y las diferentes variedades.

Producto	Variedades de un mismo tipo de producto
Acetona	
Benceno	500 ppm 1000ppm 1500ppm
Etanol	
Hidróxidodamonio	
Metanol	
Tolueno	500ppm 1000 ppm 1500 ppm
Vacio	
Xileno	500ppm 1000 ppm 1500 ppm

Tabla 6. Productos analizados con la base de datos B-NOSE y las diferentes variedades.

2.2 METODOLOGÍA GENERAL

En esta tesis de maestría se utilizaron una serie de técnicas y procedimientos que permitieron alcanzar los objetivos propuestos. El trabajo se dividió en 3 etapas en concordancia con los objetivos específicos (**Figura 20**). Las diferentes etapas están compuestas por los procedimientos que se detallan en los siguientes párrafos.

Figura 20. Etapas de la metodología general

2.2.1 Caracterización

En la etapa de caracterización, la población objeto de estudio fueron los sensores de gases de tipo MOS, utilizados en sistemas de olfato electrónico y las técnicas de análisis estadístico multivariante. Esta parte se apoyó en la revisión y observación de las hojas características de los sensores, así como el análisis de las bases de datos con las que se cuenta (Medidas realizadas con café, etanol, metanol, vinos y pacientes con EPOC, entre otras). Las variables a tener en cuenta fueron algunos parámetros estadísticos (valores máximos, mínimos, valor promedio, desviación estándar, entre otros) de las curvas características de los sensores de gases. Los resultados de esta etapa hacen parte del pre-procesamiento de la información y se recopilan en algunas variables utilizadas en el software de procesamiento. Las gráficas de la respuesta de los sensores de gases y los correspondientes datos estadísticos de las variables analizadas, permitieron obtener las mejores condiciones para el pre-procesamiento de las señales.

2.2.2 Modelo de Inferencia

En la etapa de inferencia, la población objeto de estudio fueron los métodos de regresión, específicamente la regresión de soporte vectorial SVR. Se utilizaron herramientas, *Toolbox* y aplicaciones de libre acceso, disponibilidad y licenciamiento, para: `MATLAB`, como *svm-light*. Esta parte se apoyó en la revisión exhaustiva del estado del arte respecto a las técnicas y métodos más apropiados para realizar inferencia y la experimentación con dichas técnicas. Las variables que se tuvieron en cuenta fueron la confiabilidad (Precisión en las respuestas), la repetitividad (Precisión bajo un conjunto de condiciones) y exactitud (Valor cercano al real). Se recopilaron los resultados de esta etapa usando tablas de registro de experimentos, así como: gráficas comparativas de las respuestas de los sensores de gases, obtenidas con el sensor real y a través del modelo propuesto. Adicionalmente, se realizó el registro de datos estadísticos de las variables analizadas.

2.2.3 Validación

En la etapa de validación el análisis se realizó sobre el modelo de inferencia propuesto en este trabajo de grado y la validación se realizó usando validación cruzada (CV). Esta parte se apoyó en la revisión del estado del arte respecto a las técnicas más apropiadas para realizar validación de modelos de inferencia y la aplicación de dichas técnicas. Las variables tomadas en cuenta fueron la repetitividad y las condiciones de prueba. Se recopilaron los resultados de esta etapa usando hojas de registro de experimentos, así como: gráficas comparativas de la respuesta de los sensores de gases obtenidas con el sensor real y a través del modelo propuesto, para diferentes condiciones y diferentes tipos de muestras. Adicionalmente se realizó el registro de datos estadísticos de las variables analizadas.

2.3 PRE-PROCESAMIENTO DE LA INFORMACIÓN

De las técnicas de pre-procesamiento descritas en la sección 1.3, a continuación se especifica cuáles de ellas se emplearon para manipular la información obtenida de los sensores de gases.

2.3.1 Remoción de datos anómalos

En razón de que exista algún sensor defectuoso, o que este entregando una respuesta inadecuada, como se mencionó en la sección 1.3.1, es mejor omitir el efecto de dicho sensor, para ello en la aplicación desarrollada en `Matlab` para el equipo A-NOSE (Rodríguez, Durán, & Reyes, 2010), se puede seleccionar los sensores que se tendrán en cuenta para el análisis (ver **Figura 21**), este mismo principio se utilizó para manipular la información en este caso.

Figura 21. Ventana de selección de los sensores descartados (versión mejorada)

2.3.2 Escalado

Entre los métodos de escalado y normalización mencionados en la sección 1.3.2, debido al tipo de señales que manejan los sensores de gases, en este caso se encontró mejores resultados al aplicar un método de escalado, en vez de la normalización, para lo cual se probaron los métodos de media centrada y auto-escalado. De las pruebas iniciales se obtuvieron mejores resultados con el auto-escalado, razón por la cual fue escogido este último como método de escalado[2]. Cabe destacar que esto se evidencia al aplicar otros análisis, i.e., PCA o someterlo a un entrenamiento y comparar los resultados obtenidos con los datos escalados o normalizados con uno u otro método.

2.3.3 Manipulación de la Línea Base

De los métodos comúnmente empleados para la manipulación de la línea base (diferencial, relativo y fraccional) se utilizó el método diferencial para no generar respuestas adimensionales como ocurre con los otros dos métodos. El método diferencial consiste en eliminar el mínimo a cada una de las medidas de los sensores. En la **Tabla 7** se muestran los mínimos de la señal original y los mínimos de la señal después de manipular la línea base, y cabe destacar que lo valores de la medida de los sensores están en conductancia, razón por la cual los mínimos de la señal original tiene valores tan pequeños.

1e-4*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
O	0.407	0.135	0.172	0.520	0.180	0.175	0.171	0.184	0.182	0.180	0.254	0.168	0.096	0.138	0.155	0.218
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabla 7. Valores mínimos de las señales O (originales) y la señales M (manipuladas en la línea base)

[2] *En el APÉNDICE A se encuentran los códigos escritos en* `Matlab` *para los dos métodos probados.*

2.3.4 Filtrado de la señal

Como se mencionó en la sección 1.3.4 se probaron dos filtros diferentes: el de media móvil en el tiempo y el filtro *Butterworth* para filtrar frecuencia altas, conocidos como pasa bajas. A continuación se muestran las características de cada uno y el filtro elegido para las señales.

Filtro de media móvil en el tiempo

Este es un filtro tipo ventana que se desplaza a través de los datos. Para la aplicación de este filtro se utilizó una ventana con anchura de 12 puntos, se escogió este tamaño de la ventana con base en una metodología heurística y se encontró que una ventana de un menor tamaño no causa mucho efecto en el filtro y una ventana de mayor tamaño ocasiona perdida de información y distorsión de la curva característica de la señal. Para este filtro se tiene que la **ecuación (2)** queda expresada como: $y(i) = \frac{1}{12}\sum_{j=0}^{11} s[i+j]$.

Cabe recordar que $s[i]$ es la señal de entrada (datos sin filtrar **Figura 22**), $y[i]$ es la correspondiente señal de salida (datos filtrados **Figura 23**) y M es el número de puntos escogidos para promediar (tamaño de la ventana) denominado factor de anchura del filtro.

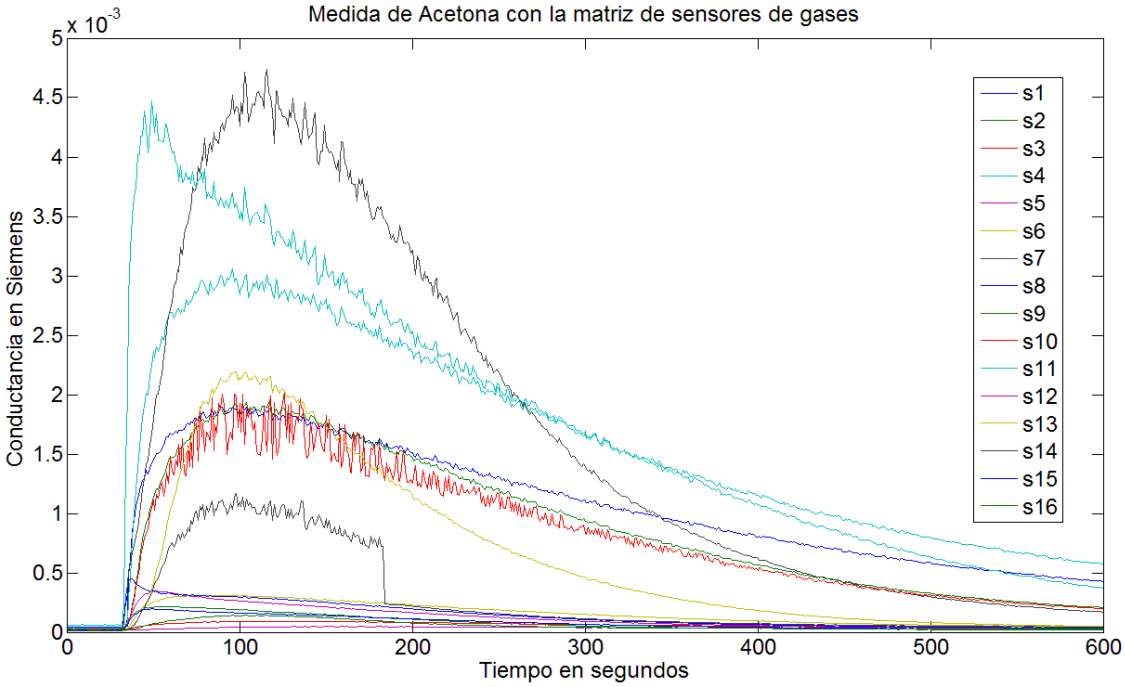


Figura 22. Medida de un arreglo de sensores sin filtrado de señal (Datos tomados de la base de datos B-NOSE)

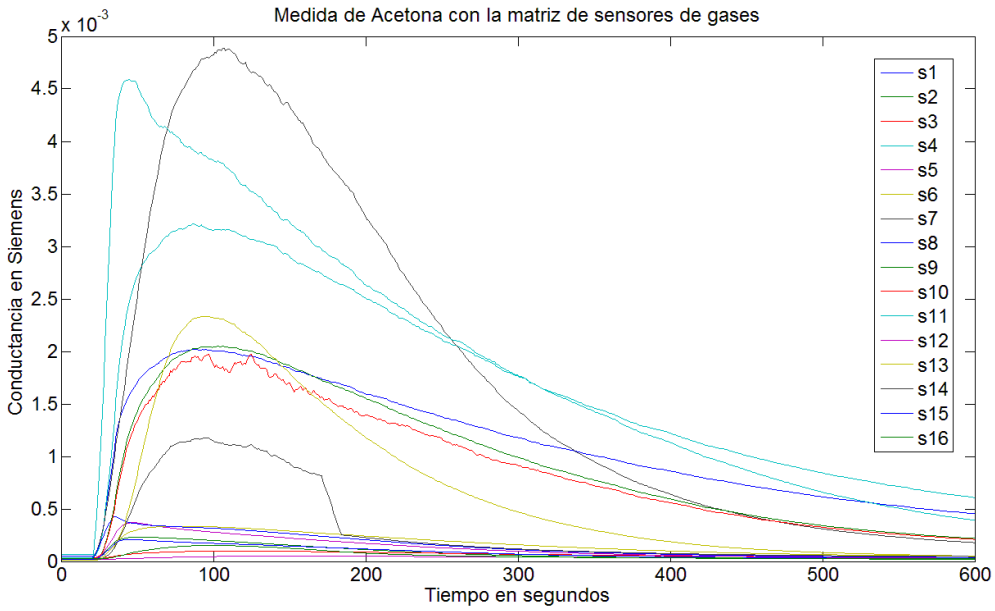


Figura 23. Resultado de aplicación del filtro de media móvil (Comparar con la **Figura 22**)

Filtro Butterworth

En `Matlab` para encontrar el orden mínimo de un filtro Butterworth analógico o digital que cumpla con las especificaciones de diseño, se tiene la función `buttord`. `[n,Wn]=buttord(Wp, Ws, Rp, Rs, 's'),`

Donde:

`Wp` => esquina superior de la frecuencia de paso.
`Ws` => esquina superior de la frecuencia de rechazo.
`Rp` => atenuación deseada en dB en la banda de paso.
`Rs` => atenuación deseada en dB en la banda de rechazo.

Esta función entrega el orden del filtro y la frecuencia natural del mismo. Una vez que se tienen todos los parámetros se puede aplicar el filtro y obtener una señal filtrada. En este caso en particular se ajustó el filtro para dejar pasar las señales de hasta 1 *Hz* aproximadamente y atenuar el resto, con el propósito de atenuar la componente oscilante indeseada que se encuentra enmascarada en la señal de los sensores, los parámetros del filtro se ajustaron con base en una metodología heurística, en el **Apéndice A** se encuentra el procedimiento aplicado y en la **Figura 24** se encuentra el resultado de aplicar el filtro *Butterworth* a una medida de la base de datos B-NOSE.

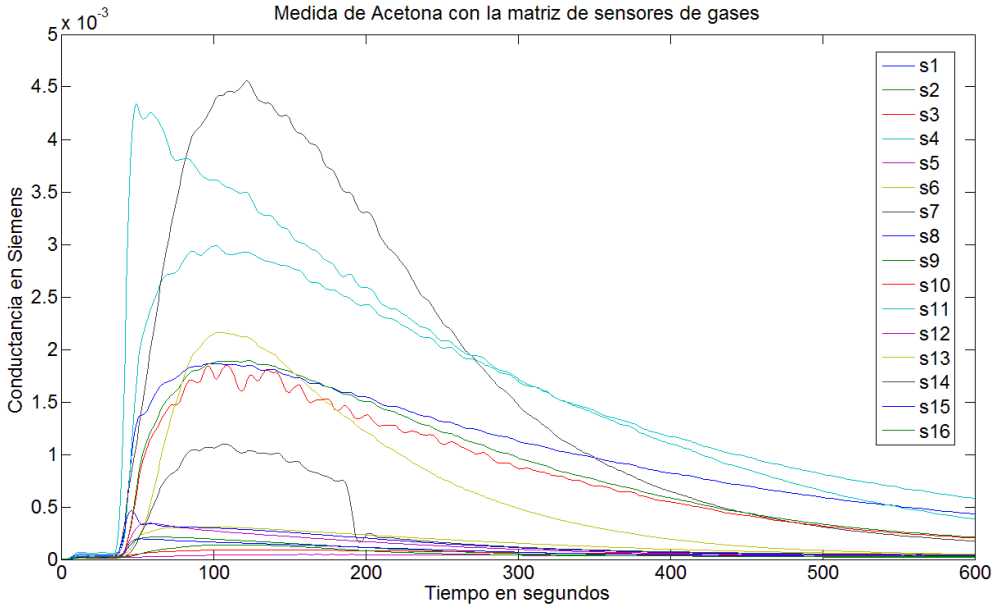


Figura 184. Resultado de aplicación de un filtro pasa bajas *Butterworth* (comparar con la **Figura 22**)

Selección del filtro

Para tener algún criterio para seleccionar el filtro adecuado se optó por calcular el error cuadrático medio y error absoluto de las señales filtradas con respecto a las señales originales, tal y como se muestra en la **Tabla 8**.

Filtro	Promedio del error absoluto	Promedio del error cuadrático medio
Media móvil	3.7805e-05	8.1731e-06
Butterworth	4.3835e-07	7.8356e-06

Tabla 8. Medida de los errores de los 2 filtros aplicados

Comparando los errores obtenidos después de aplicar los dos filtros, usando la **Tabla 8**, se evidencia un mejor comportamiento para el filtro de media móvil, por lo que se optó por escoger el filtro de media móvil y adicionalmente con este se obtiene mayor suavidad de las señales que la presentada con el filtro *Butterworth*, el cual todavía tiene asociado una pequeña componente aleatoria en algunas señales.

2.4 ENTRENAMIENTO DE LA MÁQUINA DE VECTORES DE SOPORTE PARA REGRESIÓN

El proceso de entrenamiento de la máquina de vectores de soporte para regresión SVR se explica a continuación**.** La estrategia o método utilizado en este caso se fundamenta en 6 etapas, el orden mostrado en la **Figura 26** no es del todo estricto, debido a que las etapas 2, 3 y 4 pueden intercambiarse sin llegar a afectar los resultados o el proceso de entrenamiento como tal; sin embargo, la etapas 1, 5 y 6, lógicamente deben realizarse en el orden indicado y no pueden ser intercambiadas.

Los datos de entrenamiento se organizan en una matriz de *m* filas (muestras) por *n* columnas (sensores), la cual se forma a su vez con los datos de varias matrices, cada una de las cuales corresponde a una medida especifica (i.e., una medida realizada con Etanol al 99%), partiendo de la base que las medidas debieron ser obtenidas con una misma matriz de sensores de gases (i.e., base de datos A-NOSE), en este caso la matriz con los datos de entrenamiento se nombra `MATRIXtrain`. De esta matriz se escoge una columna (sensor), el cual será el elegido como sensor virtual, en otras palabras este es el *target*, que servirá en el proceso de entrenamiento, en el cual la máquina de regresión debe intentar emular o encontrar su modelo (ver **Figura 25**).

Figura 195. Ventana de selección del sensor virtual

La selección de los parámetros del regresor lineal SVR, en este caso se hace a través de una Interfaz Gráfica de Usuario (GUI) implementada en `Matlab` (ver **Figura 27**), los parámetros que se deben configurar son *C*, ε, el tipo de kernel y los parámetros del kernel (si aplica).

Posteriormente a los datos de entrenamiento se le aplican las técnicas de pre-procesamiento descritas en la sección 2.3 de este libro.

Después de iniciar el proceso de entrenamiento, se realiza el cómputo del modelo de regresión de acuerdo a los parámetros escogidos; en las pruebas realizadas el tiempo que toma entrenar la máquina de regresión puede tomar uno cuantos minutos, aproximadamente 5 minutos para una matriz de entrenamiento de 2400x8 con datos en

coma flotante (dependiendo del computador empleado, los parámetros escogidos y de las tareas simultaneas que se estén realizando) en un computador con procesador de doble núcleo a 2.5 GHz y 4 GB de RAM. El modelo obtenido servirá posteriormente como el sensor virtual.

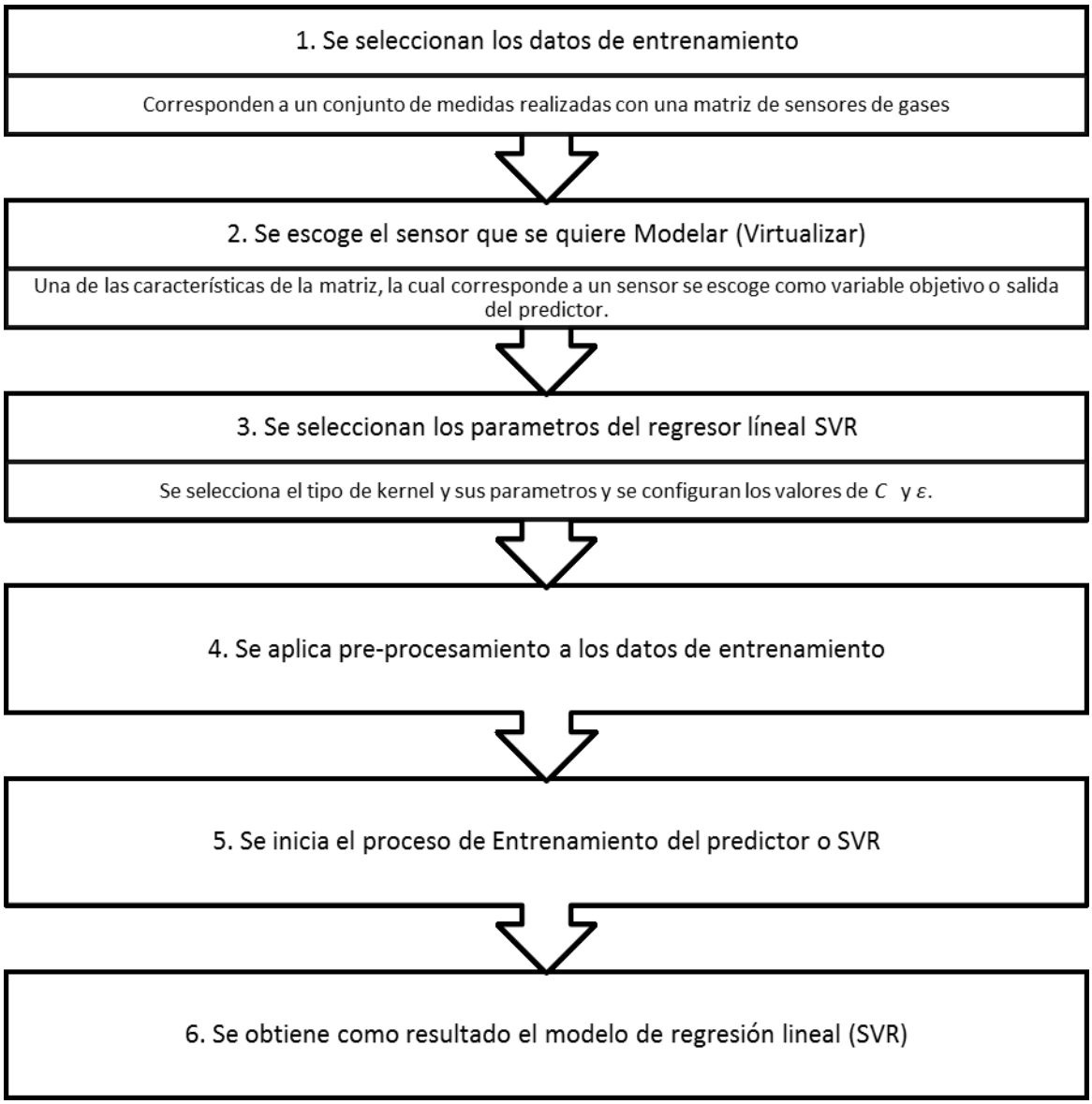


Figura 206. Proceso de entrenamiento de la máquina de vectores de soporte para regresión

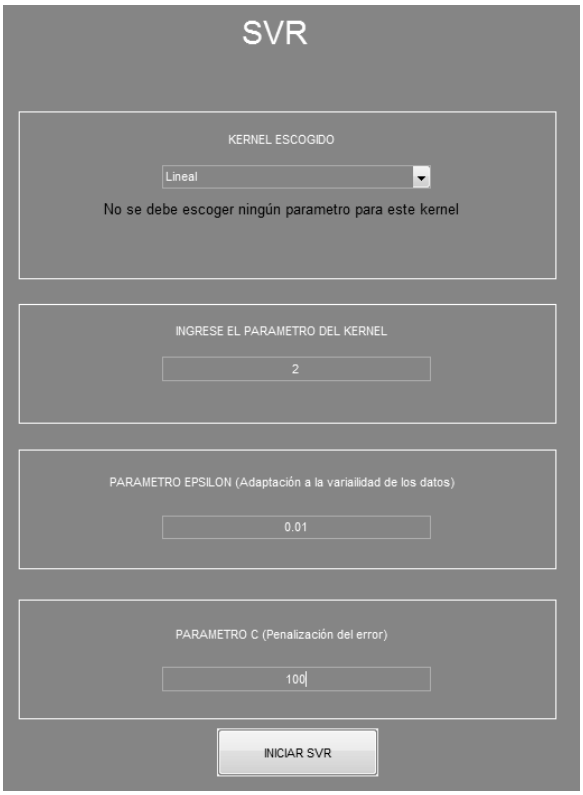


Figura 217. Interfaz para configurar los parámetros de la SVR

2.5 SINTONIZACIÓN DE LOS PARÁMETROS DE ENTRENAMIENTO

El proceso de sintonización de los parámetros de entrenamiento de la SVR, se llevó a cabo aplicando el método de búsqueda sistemática tipo rejilla, realizando un recorrido ordenado del espacio de estados con el fin de encontrar los parámetros óptimos. Comenzando con una rejilla gruesa para identificar las zonas y luego utilizando una rejilla fina. A continuación se detalla la aplicación del método para este caso.

Como primera medida se fijó un tipo de kernel, unos parámetros fijos para ese kernel (si aplican), y se comenzó a variar C y ε, los valores iniciales fueron 100 y 0,1 respectivamente y disminuyendo o aumentando el valor según se evidenciara una reducción del error cuadrático medio. El valor de 100 se escogió debido a que en el estado del arte se encontró este valor como un punto medio de partida en el proceso de sintonización *v-SVR* en (Pérez Jordán, 2011). Se debe tener en cuenta el parámetro C debe ser un escalar positivo mayor o igual a los multiplicadores de LaGrange, razón por la cual debe ser mayor a los valores de los datos de entrenamiento. Si se quiere obtener un alto grado de generalización se debe escoger un valor de C bajo para no penalizar demasiado los errores, mientras que si se requiere una modelización ajustada se debe usar valores de C altos (Gómez Pérez, 2004).

Fundamentado en lo anterior se justifica la razón de la escogencia de *C* en 100 como punto inicial, ya que corresponde a un valor que sobrepasa a cualquiera de los datos de entrenamiento sin llegar a ser extremadamente alto. También se debe tener en cuenta que valores más grandes de *C*, aumentarían el costo computacional sin generar mejores resultados. Mientras que el valor inicial de ε se escogió con base en los resultados de las pruebas iniciales, valores más pequeños de ε, aumentarían el costo computacional en algunos casos sin generar mejores resultados.

Después de realizar la búsqueda sistemática de *C* y ε, obtener el error cuadrático medio de cada entrenamiento y validación se seleccionó el valor de *C* y ε con los cuales se obtuvo el menor error de entrenamiento. Con este valor del parámetro *C* y ε se realizó la búsqueda sistemática para el parámetro(s) del kernel en caso de ser necesario. El método aplicado para la sintonización de los parámetros se muestra en la **Figura 28**.

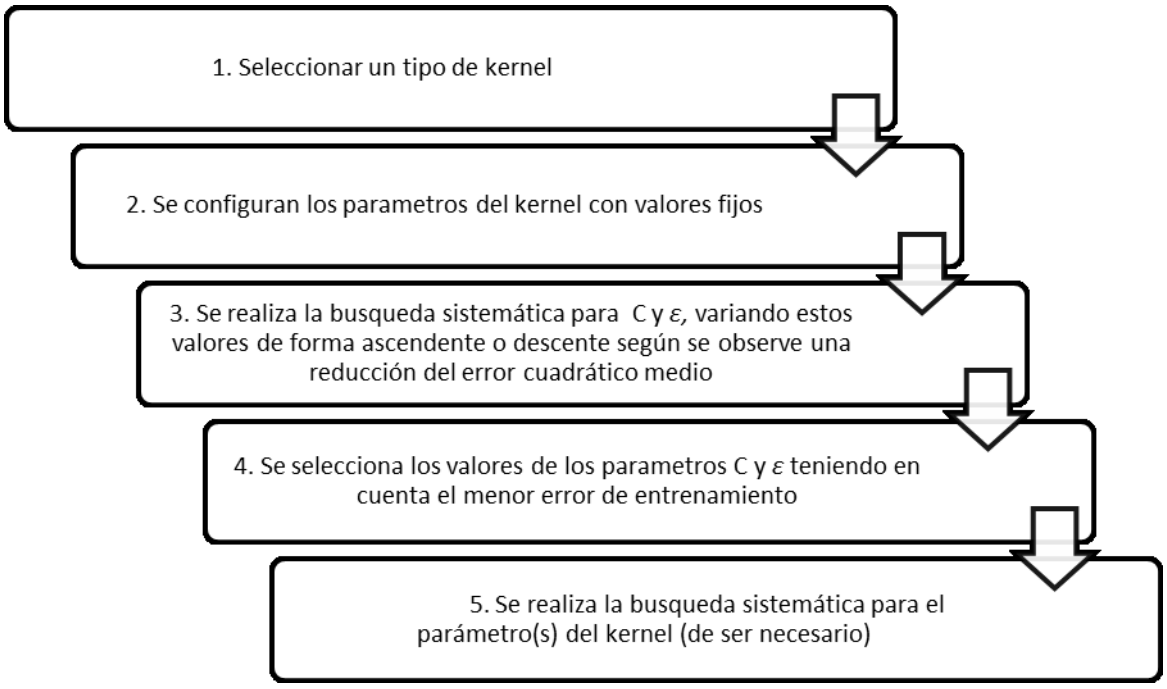


Figura 28. Método de sintonización de los parámetros de entrenamiento aplicado para la regresión con SVM

El proceso de sintonización de parámetros se debe realizar cada vez que se desee obtener un sensor virtual diferente o cuando las condiciones cambien, por ejemplo los sensores físicos deben ser remplazados o cambiados, también a medida que pase el tiempo, debido al deterioro de los sensores y las derivas sería recomendable ajustar los parámetros nuevamente con el objetivo de tener una medida confiable de los sensores virtuales.

El costo computacional para este método en la fase de entrenamiento está relacionado con: el tipo de kernel empleado, la cantidad de los datos a analizar, el método de escalado de

los datos utilizado y la cantidad de iteraciones que se deban llevar a cabo en la búsqueda sistemática, lo que se busca en cada iteración es reducir el error cuadrático medio entre la medida del sensor real y del sensor virtual, si en cada iteración existe una reducción de dicho error el proceso de sintonización es más rápido y se realizaran menos iteraciones, pero si por el contrario no se evidencia una reducción en dicho error el software deberá continuar con la búsqueda de unos mejores parámetros y por lo tanto el tiempo de ejecución aumentara.

A continuación se pueden apreciar detalles del proceso de selección de parámetros y los tiempos de ejecución.

Nombre de la prueba	**Kernel**	**Método de escalado**	**Datos de entrenamiento (matriz)**	**Datos de validación (matriz)**	**Tiempo (Horas:minutos:segundos)**
Primera	Lineal	mcx	2400x8	2400x8	00:00:16
Primera	Lineal	aux	2400x8	2400x8	00:05:02
Primera	Lineal	Datos sin escalar	2400x8	2400x8	00:00:16
Segunda	Lineal	mcx	2400x8	2400x8	00:02:53
Segunda	Lineal	aux	2400x8	2400x8	00:09:20
Segunda	Lineal	Datos sin escalar	2400x8	2400x8	00:41:05
Tercera	Lineal	mcx	2400x8	2400x8	00:00:17
Tercera	Lineal	aux	2400x8	2400x8	00:03:12
Tercera	Lineal	Datos sin escalar	2400x8	2400x8	00:00:17
Cuarta	Lineal	mcx	2400x8	2400x8	00:04:50
Cuarta	Lineal	aux	2400x8	2400x8	00:07:52
Cuarta	Lineal	Datos sin escalar	2400x8	2400x8	00:36:18

Tabla 9. Tiempos de ejecución en la primera fase de sintonización de parámetros con el kernel lineal

El equipo de cómputo utilizado para llevar a cabo este proceso cuenta con las siguientes características: procesadores de por lo menos 2,4 GHz, 4 GB de memoria RAM y sistema operativo Windows 7 de 64 bits. Características del software: `MATLAB` versión 7.14.0.739 (R2012a), `Java` versión 1.6._17-b04 con Sun Microsystems Inc. Java HotSpot(TM) 64BitServer VM mixed mode.

Según los tiempos de ejecución mostrados en la **Tabla 9** para el kernel lineal se ejecuta más rápido el proceso con los datos a los cuales se les aplicó el escalado `mcx` (media centrada), seguido de los datos a los cuales se les aplicó el escalado `aux` (auto-escalado) y por último los datos que no están escalados, siendo estos últimos tiempos de ejecución muy altos en algunos casos y con los que se obtuvieron los resultados menos favorables.

Nombre de la prueba	Kernel	Método de escalado	Datos de entrenamiento (matriz)	Datos de validación (matriz)	Tiempo (Horas:minutos:segundos)
Primera	Gaussiano	aux	2400x8	2400x8	00:00:40
Primera	Gaussiano	Datos sin escalar	2400x8	2400x8	00:00:22
Segunda	Gaussiano	aux	2400x8	2400x8	00:00:38
Segunda	Gaussiano	Datos sin escalar	2400x8	2400x8	00:00:22
Tercera	Gaussiano	mcx	2400x8	2400x8	00:00:21
Cuarta	Gaussiano	mcx	2400x8	2400x8	00:00:23

Tabla 10. Tiempos de ejecución en la primera fase de sintonización de parámetros con el kernel gaussiano

Según los tiempos de ejecución mostrados en la **Tabla 10** para el kernel gaussiano se encuentra que los tiempos de ejecución son muy similares para los datos escalados con media centrada `mcx` y los datos sin escalar, mientras que para los datos auto-escalados los tiempos de ejecución son aproximadamente del doble que en los otros dos casos.

En concordancia con los tiempos de ejecución mostrados en la **Tabla 10**, de la **Tabla 11** se puede evidenciar el mismo comportamiento, el tiempo de ejecución aumenta básicamente por el aumento en la cantidad de datos de entrenamiento y validación. Los tiempos de ejecución mostrados en la **Tabla 11** corresponden a otro proceso de sintonización de parámetros. Cabe aclarar que estos tiempos de ejecución son altos debido a que se realizan múltiples procesos de entrenamiento, una vez encontrados los parámetros adecuados, entrenar la SVR para obtener el sensor virtual toma unos cuantos minutos o incluso segundos dependiendo de las condiciones.

Nombre de la prueba	Kernel	Método de escalado	Datos de entrenamiento (matriz)	Datos de validación (matriz)	Tiempo (Horas:minutos:segundos)
Primera	Lineal	aux	28800x8	7200x8	01:27:06
Primera	Gaussiano	aux	28800x8	7200x8	08:53:22
Segunda	Lineal	aux	28800x8	7200x8	01:19:25
Segunda	Gaussiano	aux	28800x8	7200x8	07:28:49
Tercera	Lineal	aux	28800x8	7200x8	01:07:04
Tercera	Gaussiano	aux	28800x8	7200x8	07:23:04
Primera	Lineal	mcx	28800x8	7200x8	00:04:54
Primera	Gaussiano	mcx	28800x8	7200x8	00:40:54
Segunda	Lineal	mcx	28800x8	7200x8	00:04:26
Segunda	Gaussiano	mcx	28800x8	7200x8	00:25:04
Tercera	Lineal	mcx	28800x8	7200x8	00:04:54
Tercera	Gaussiano	mcx	28800x8	7200x8	00:41:03

Tabla 11. Tiempos de ejecución en la primera fase de sintonización de parámetros con el kernel lineal

2.6 VALIDACIÓN DEL MODELO DE INFERENCIA PROPUESTO

Para validar el modelo de inferencia propuesto se utilizó la validación cruzada. La cual sirve para verificar cómo se comporta el modelo ante muestras o especímenes de productos conocidos y productos desconocidos, de acuerdo a como se organicen los grupos de medidas de entrenamiento y validación.

Se realizaron tres validaciones cruzadas como se muestra en la sección 3.4 de resultados, con medidas de la base de datos A-NOSE, de las cuales se tomaron los grupos de medidas que poseían igual cantidad de especímenes para cada producto, lo cual resulto en un total de 120 medidas de 24 productos diferentes, con ellas se conformaron de forma aleatoria 5 paquetes con una medida de cada uno de los 24 productos. La primera validación se realizó con el objetivo de establecer el comportamiento del sensor virtual ante otras medidas diferentes a las utilizadas en el entrenamiento pero de productos ya conocidos. Mientras que la segunda validación se realizó con el objetivo de establecer el comportamiento del sensor virtual ante medidas de productos diferentes a las utilizadas en el entrenamiento. La tercera validación se realizó de forma similar a la segunda validación pero cambiando el sensor virtual, en este caso se escogió como sensor virtual el sensor S5 a diferencia de las otras dos validaciones donde se había escogido como sensor virtual el sensor S6. Adicionalmente en la tercera validación se realizaron pruebas reduciendo la cantidad de sensores.

También se realizó una validación cruzada con medidas de la base de datos B-NOSE y los resultados se muestran en la sección 3.6, de esta base de datos se tomaron los grupos de medidas que poseían igual cantidad de especímenes para cada producto, lo cual resulto en un total de 95 medidas de 19 productos diferentes, con ellas se conformaron de forma aleatoria 5 paquetes con una medida de cada uno de los 19 productos.

Se realizaron dos validaciones diferentes, la primera se realizó con el objetivo de corroborar la aplicabilidad de la sintonización de parámetros encontrada en la sección 3.3 y obtener el modelo del sensor virtual del sensor siete (S7) de la base de datos B-NOSE a partir de los quince sensores restantes de dicha base de datos. Y en la segunda validación se obtuvo el modelo del sensor virtual del sensor siete (S7) a partir de solo cuatro sensores, aquellos con los que el sensor siete mostraba la correlación más baja.

3. RESULTADOS

3.1 ORGANIZACIÓN DE LOS DATOS

En las bases de datos A-NOSE y B-NOSE cada medida realizada fue almacenada en un archivo plano, cada archivo que contiene los datos de una medida corresponde a una matriz donde las filas representan las muestras tomadas y las columnas corresponden a las medidas de cada sensor.

Para organizar los datos de entrenamiento y validación se puede realizar cargando varios archivos con diferentes medidas, con la información de cada archivo (cada matriz) se puede conformar una matriz principal que contienen todas las medidas que se desean analizar, esta matriz principal tiene el mismo número de columna que las matrices de medidas, pero mayor número de filas debido a que contiene varias medidas. A dicha matriz principal se le aplican las técnicas de pre-procesamiento descritas en la sección 1.3. Posteriormente se particiona dicha matriz en datos de entrenamiento y datos de validación. De la forma como se explica en este párrafo se realizó la organización de los datos para la sección 3.3 (Selección De Los Parámetros Adecuados)

La forma como se organizó la información para la sintonización de parámetros, validaciones y demás pruebas fue la siguiente. Primero se cargaron todos los archivos de una determinada base de datos (i.e. A-NOSE), previamente se seleccionaron y almacenaron en una carpeta los conjuntos de medidas más uniformes, en el caso de la base de datos A-NOSE todos los que poseían cinco medidas de un mismo producto, los cuales son la mayoría y corresponden a 24 productos, para un total de 120 medidas, para la base de datos B-NOSE todos los que poseían cinco medidas de un mismo producto, los cuales son la mayoría y corresponden a 19 productos, para un total de 95 medidas. Una vez cargadas todas las medidas que se quieren analizar se construye una matriz principal que contiene todas las medidas, se aplican las técnicas de procesamiento descritas en la sección 1.3. Debido a que por cada producto de las bases de datos analizadas se tienen 5 medidas, se armaron 5 paquetes para cada base de datos, cada uno de ellos contiene una medida de cada producto, los paquetes se armaron con un algoritmo propio de forma aleatoria. Para la base de datos A-NOSE se tienen 5 paquetes que contiene cada uno 24 medidas de los 24 productos y para la base de datos B-NOSE se tienen 5 paquetes que contiene cada uno 19 medidas de los 19 productos. Cabe aclarar que las bases de datos se analizan por separado y que la obtención de un determinado sensor virtual se realiza conforme la

información que entregue una determinada base de datos, debido a que cada base de datos fue toma con equipos distintos y sensores diferentes.

La cantidad de datos tomados para entrenamiento y validación depende de las pruebas que se quieran realizar, en las secciones siguientes en cada caso se explica cómo se tomaron los datos para realizar el entrenamiento y los datos para realizar la validación.

3.2 PRUEBAS INICIALES

Las pruebas iniciales fueron realizadas con el archivo EXCE1.txt de la base de datos A-NOSE, la cual corresponde a una medida realizada con un café colombiano Excelso tipo exportación (Rodríguez, Durán, & Reyes, 2010).

En estas pruebas se entrenó una SVM utilizando kernel: lineal, polinomial, gaussiano y sigmoidal, para determinar cuál presentaba menor error de entrenamiento. El sensor escogido para modelar fue el S6 cuya referencia es TGS-813 del fabricante FIGARO (**Tabla 3**).

En la **Tabla 12** se encuentran los parámetros que se emplearon para configurar la SVM, los que están resaltados son los parámetros que se configuraron con más frecuencia en las diferentes pruebas, a menos que se indique lo contrario se asumirá que el resto de parámetros son iguales a los mostrados en dicha tabla.

A continuación se muestran los mejores resultados de las pruebas iniciales para cada uno de los kernel probados.

Resultado con el kernel lineal

Con los mismos parámetros de la **Tabla 12** para el entrenamiento de la SVR se obtuvo los errores mostrados en la **Tabla 13** para este kernel. La gráfica de la predicción del S6 realizada por la SVR (kernel lineal) y la medida del sensor físico S6 se muestran en la **Figura 29**.

Resultado con el Kernel Polinomial

Con los mismos parámetros de la **Tabla 12** a excepción de: `Kernel` = 1 (Polinomial), `KernelParam` = 2, se realizó el entrenamiento de la SVR, se obtuvo los errores mostrados en la **Tabla 13** para este kernel.

La gráfica de la predicción del S6 realizada por la SVR (kernel polinomial de orden 2) y la medida del sensor físico S6 se muestran en la **Figura 30**.

Parámetro	Nombre de la variable	Valor por defecto	Valor ajustado
Nivel de detalle {0,1,2,3} (Muestra de resultados)	Verbosity	1	1
Tipo de SVM {0,1}	Regression	0	1
Constante de regularización (0…Inf)	***C***	**[avg(*x***x*)]^-1**	**100**
Tamaño del cilindro híper-dimensional ε (0…Inf)	**TubeWidth**	**0.1**	**0.1**
Factor de costo (0…Inf)	CostFactor	1	1
Híper-plano sesgado o parcializado {0,1}	Biased	1	1
Eliminación de errores de entrenamiento y re-entrenamiento {0,1}	RemoveIncons	0	0
Con los datos calcula las estimaciones leave-one-out {0,1}	ComputeLOO	0	1
Valor de rho para XiAlpha-estimador para la poda y cálculo de leave-one-out (0…2)	XialphaRho	1	1
Búsqueda a profundidad para largos XiAlpha-estimador {0..100}	XialphaDepth	0	0
Fracción de ejemplos no etiquetados que se clasifican en la clase positiva (0…1)	TransPosFrac	0	0
Tipo de kernel {0…4}	**Kernel**	**0**	**0**
Parámetro del kernel (Depende del Kernel)	**KernelParam**	**-**	**-**
Máximo tamaño de los sub-problemas {2…}	MaximumQP	10	10
Número de nuevas variables de entrada al conjunto de trabajo en cada iteración {2…}	NewVariables	*Q*	5
Tamaño de la memoria cache par las evaluaciones del kernel en MB (5…Inf)	CacheSize	40	400
Error en el criterio de terminación [*y* [*w***x*+*b*] - *1*] = *eps* (0…Inf)	EpsTermin	0.001	0.001
Número de iteraciones, una variable necesita considerarse óptima antes de la reducción {5…Inf}	ShrinkIter	100	100
Verificación final de optimización para las variables eliminadas por reducción o contracción {0,1}	ShrinkCheck	1	1
Archivo para escribir las etiquetas predichas de los ejemplos no etiquetados después de aplicar el método transductivo (String)	TransLabelFile		
Los alphas son almacenados en este archivo después de aprendizaje, guardando el orden de los datos de entrenamiento. (String)	AlphaFile		

Tabla 12. Parámetros de configuración de la SVM

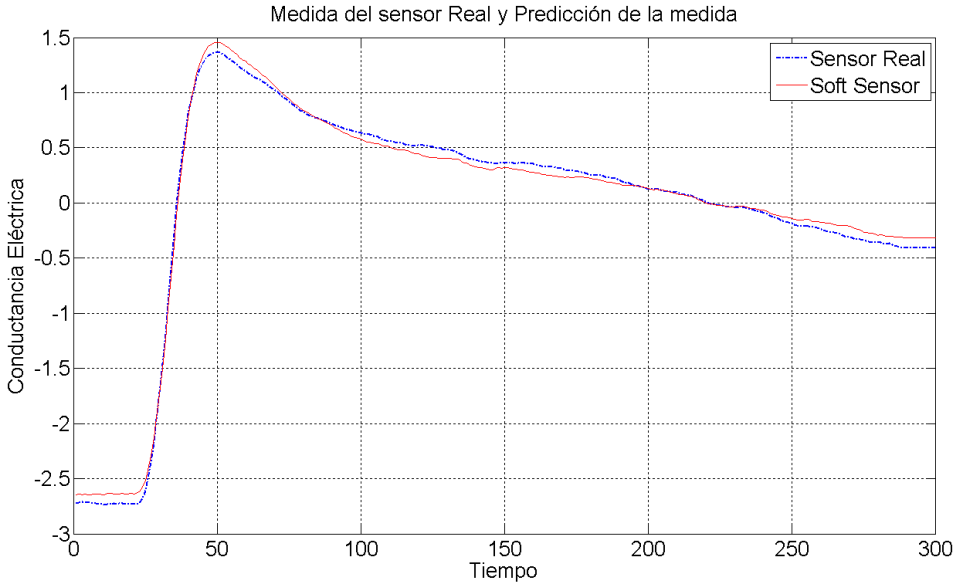


Figura 29. Gráfica del sensor físico (azul) y la predicción de la SVR para el mismo sensor (rojo) con un kernel lineal

Resultado con el Kernel gaussiano

Con los mismos parámetros de la **Tabla 12** a excepción de: `Kernel` = 2 (gaussiano), `KernelParam` = 1, se realizó el entrenamiento de la SVR, se obtuvo los errores mostrados en la **Tabla 13** para este kernel.

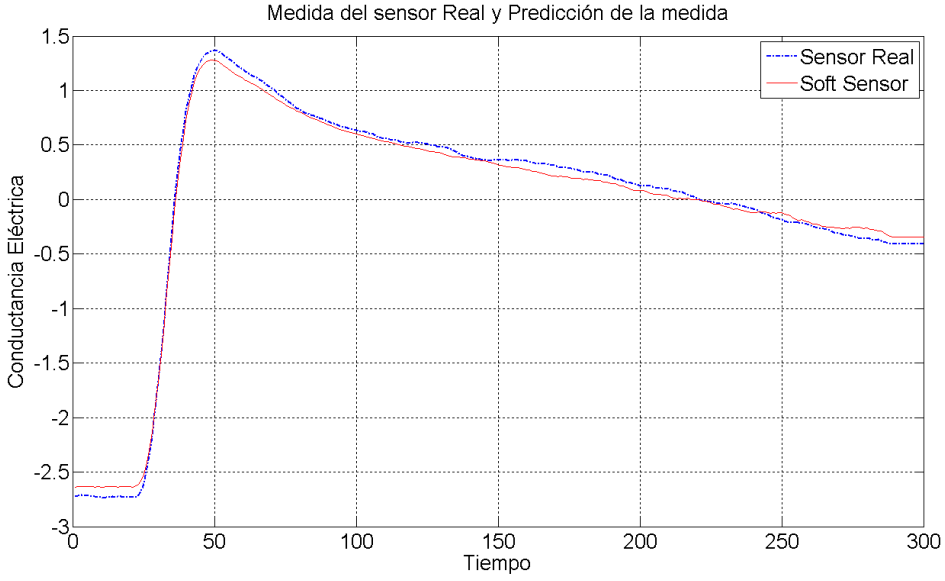


Figura 30. Gráfica del sensor físico (azul) y la predicción de la SVR para el mismo sensor (rojo) con un kernel polinomial de orden 2

En la gráfica de la predicción del S6 realizada por la SVR (kernel gaussiano) y la medida del sensor físico S6 se muestran en la **Figura 31**.

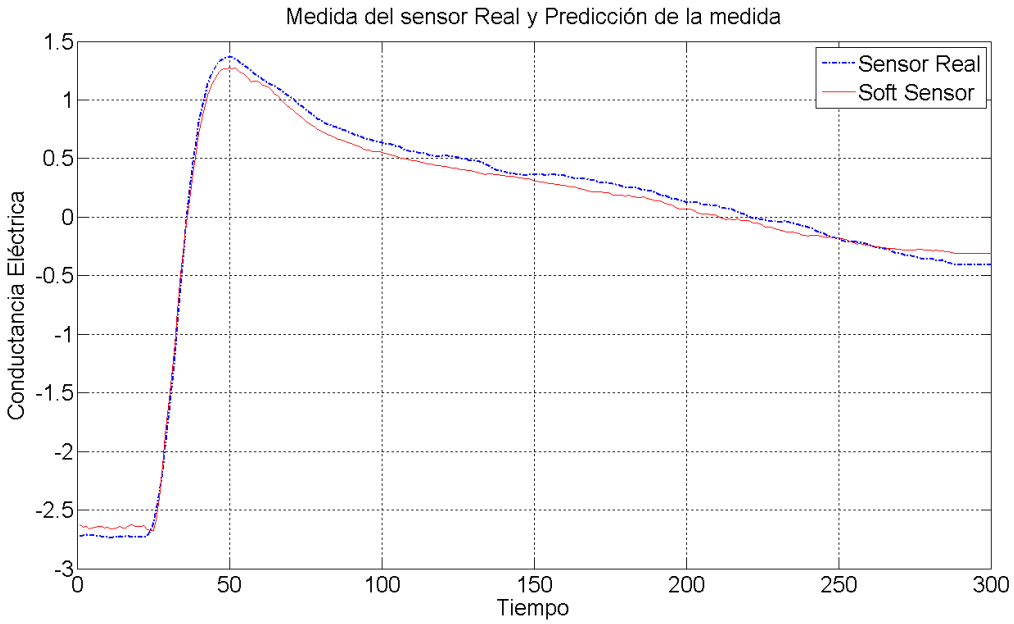


Figura 31. Gráfica del sensor físico (azul) y la predicción de la SVR para el mismo sensor (rojo) con un kernel gaussiano con gama 1

Resultado con kernel sigmoidal

Con los mismos parámetros de la **Tabla 12** a excepción de: `Kernel` = 3 (sigmoidal), `KernelParam` = [1, 0.01], se realizó el entrenamiento de la SVR, se obtuvo los errores mostrados en la **Tabla 13** para este kernel.

En la gráfica de la predicción del S6 realizada por la SVR (kernel sigmoidal) y la medida del sensor físico S6 se muestran en la **Figura 32**.

Cabe aclara que los resultados mostrados anteriormente obtenidos en las pruebas iniciales, hacen parte de una serie de resultados en donde se realizaron múltiples pruebas a ensayo y error, y los mismos corresponden a los mejores resultados en cada caso.

En la **Tabla 13** se muestra una comparación de los resultados más relevantes de la aplicación de 4 diferentes kernel, según estos resultados la SVR con el kernel lineal y polinomial son los más adecuados para realizar la regresión de este tipo de señales, mientras que el kernel sigmoidal es el menos apropiado para este problema en particular. Además el kernel gaussiano y sigmoidal son más complejos y requieren más costos computacional, teniendo en cuenta los tiempos de ejecución, la cantidad de vectores de soporte y el número de evaluaciones del kernel.

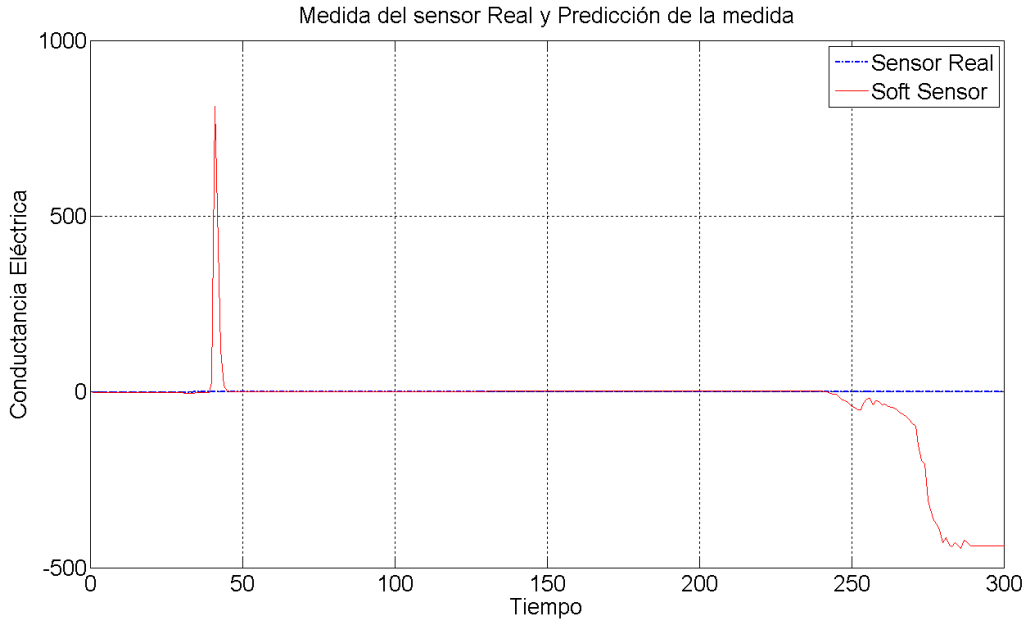


Figura 32. Gráfica del sensor físico (azul) y la predicción de la SVR para el mismo sensor (rojo) con un kernel sigmoidal

Kernel	Error Cuadrático medio	Tiempo de ejecución	Vectores de soporte	Número de evaluaciones de kernel
Lineal	0.0039	0,01 s	5	5185
Polinomial	0.0039	0,02 s	6	24550
Gaussiano	0.0059	0,04 s	25	35054
Sigmoidal	1.8723e+004	90,28 s	22 (incluyendo 13 en cota superior)	1184606

Tabla 13. Comparación de resultados del entrenamiento con diferentes kernel

Como resultado de esta sección hay que destacar que el kernel sigmoidal no mostró ser adecuado, razón por la cual en las secciones sucesivas no se tiene en cuenta; se experimentó con los 3 kernel restantes lineales, polinomial y gaussiano con los cuales se obtuvieron buenos resultados.

3.3 SELECCIÓN DE LOS PARÁMETROS ADECUADOS

Cuando se comenzó a validar la información con los resultados obtenidos en las pruebas iniciales se obtuvo sobre entrenamiento de la máquina, adicionalmente no se estaba aplicando ningún método de pre-procesado (escalado y normalización) y esto trajo consigo que no se realizaba ninguna predicción para otras medidas diferentes a las de entrenamiento.

Para solventar este error se empezaron a emplear algunas técnicas de pre-procesamiento (descritas en la sección 2.3) y posteriormente se comenzaron a realizar pruebas de entrenamiento con varias medidas para obtener una mejor generalización de la SVR, estas pruebas se muestran a continuación.

La sintonización de los parámetros se realizó aplicando el método de sintonización de parámetros descrito en la sección 2.5, realizando una búsqueda sistemática tipo rejilla a un conjunto de datos de entrenamiento de la base de datos A-NOSE.

Dentro de este grupo de entrenamiento había 16 medidas de 14 productos diferentes, 8 medidas de productos diferentes se utilizaron para entrenar la SVR y las otras 8 para validar, en la **Tabla 14** se puede apreciar el conjunto de datos de entrenamiento y el conjunto de datos de validación.

CONJUNTO DE MEDIDAS DE ENTRENAMIENTO		CONJUNTO DE MEDIDAS DE VALIDACIÓN	
Espécimen	**Nombre del fichero**	**Espécimen**	**Nombre del fichero**
Durazno	1DUR-.txt	Aguardiente	AGUX2.txt
Aguardiente	AGUX4.txt	Café Excelso UGQ	EUGQ5.txt
Etanol 99%	E99X4.txt	Manzana chilena	MCH-1.txt
Café Excelso UGQ	EUGQ3.txt	Pasilla con fermento	Pa3F4.txt
Café Excelso	EXCE3.txt	Pasilla con químico	PQ1T4.txt
Metanol 95%	M95X3.txt	Vino blanco	VINBL-02.txt
Maracuyá	MAR-3.txt	Vino de naranja	VINOR-02.txt
Pasilla con fermento	Pa3F2.txt	Vino tinto	VINTI-05.txt

Tabla 14. Medidas de entrenamiento y validación

El procedimiento mencionando anteriormente se realizó para la función kernel lineal. En el **APÉNDICE B** se muestran algunas tablas de la variación de los parámetros y de los errores obtenidos en el proceso de sintonización. Mientras que en la **Tabla 15** se muestra un resumen con los mejores resultados obtenidos en diferentes procesos de entrenamiento para los datos pre-procesados pero sin aplicar escalado.

Kernel	Parámetro(s) del kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Lineal	No aplica	100	0,1	0,00098	-116,17%	0,00081	-126,81%
Lineal	No aplica	96	0,02	0,00023	5,55%	0,00022	34,57%
Lineal	No aplica	51	0,09	0,00097	-115,35%	0,00083	-131,66%
Lineal	No aplica	49	0,01	0,00011	-5,81%	0,00008	4,70%

Tabla 15. Mejores resultados de entrenamiento aplicado al conjunto de datos pre-procesados sin escalar en la etapa de selección de parámetros.

En la **Tabla 16** se muestra un resumen con los mejores resultados obtenidos en diferentes procesos de entrenamiento para los datos pre-procesados y escalados con la técnica de auto-escalado.

Kernel	Parámetro(s) del kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Lineal	No aplica	102	0,07	0,00106	-1,11%	0,00201	-0,92%
Lineal	No aplica	98	0,02	0,00086	-2,65%	0,00166	-2,20%
Lineal	No aplica	52	0,07	0,00106	-1,11%	0,00201	-0,92%
Lineal	No aplica	50	0,01	0,00084	-2,36%	0,00147	-1,96%

Tabla 16. Mejores resultados de entrenamiento aplicado al conjunto de datos pre-procesados y auto-escalados en la etapa de selección de parámetros.

En la **Tabla 17** se muestra un resumen con los mejores resultados obtenidos en diferentes procesos de entrenamiento para los datos pre-procesados y escalados con la técnica de centrado.

Kernel	Parámetro(s) del kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Lineal	No aplica	103	0,1	0,00097	-115,53%	0,00097	-159,94%
Lineal	No aplica	100	0,01	0,00017	8,62%	0,00009	11,93%
Lineal	No aplica	47	0,07	0,00061	-71,27%	0,00063	-98,67
Lineal	No aplica	52	0,01	0,00012	2,64%	0,00007	3,65%

Tabla 17. Mejores resultados de entrenamiento aplicado al conjunto de datos pre-procesados y centrados en la etapa de selección de parámetros.

Con los resultados obtenidos como se muestran en las tres tablas anteriores se concluye que se obtienen mejores resultados para los datos auto-escalados, en segundo lugar se encuentran los datos centrados y en el último lugar los datos sin escalar. Además, se evidencia en las **Tablas 18, 19** y **20** (resaltado) que el parámetro *C* da muy buenos

resultados cuando está cercano a 52, mientras que el parámetro épsilon, en el caso de los datos auto-escalados, estuvo en 0,07 y para los datos centrados y sin escalar en 0,01, aunque con un valor de 0,01 en los datos auto-escalado, también da un muy buen resultado.

El procedimiento aplicado anteriormente para la función kernel lineal se aplicó a las funciones kernel polinomial y gaussiano. En las **Tablas 18**, **19** y **20** se encuentran los mejores resultados para el kernel polinomial aplicado a los datos de entrenamiento reseñados en la **Tabla 11** para los datos pre-procesados con y sin escalado.

Parámetro(s) del kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
2	45	0,01	0,00007	4,33%	0,00007	10,81%
3	48	0,01	0,00013	13,83%	0,00017	28,64%

Tabla 18. Mejores resultados de entrenamiento con kernel polinomial aplicado al conjunto de datos pre-procesados sin escalar en la etapa de selección de parámetros.

Parámetro(s) del kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
2	49	0,05	0,00050	-3,00%	0,00417	-31,75%
3	48	0,06	0,00050	-3,10%	0,00344	-25,14%

Tabla 19. Mejores resultados de entrenamiento con kernel polinomial aplicado al conjunto de datos pre-procesados y auto-escalados en la etapa de selección de parámetros.

Parámetro(s) del kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
2	53	0,03	0,00010	-6,47%	0,00019	-23,67
3	52	0,02	0,00010	-2,28%	0,00030	-40,30

Tabla 20. Mejores resultados de entrenamiento con kernel polinomial aplicado al conjunto de datos pre-procesados y centrados en la etapa de selección de parámetros.

En los resultados mostrados en las **Tablas 18**, **19** y **20** se puede evidenciar que para el kernel polinomial el parámetro 2 (orden del polinomio) es el que da los mejores resultados con el conjunto de entrenamiento para los datos sin escalar y para los datos auto-escalados, con los datos centrados se obtiene un mejor desempeño con los datos de entrenamiento para el parámetro 3 pero se obtuvo un desempeño muy regular con los datos de validación. En términos generales, los mejores resultados continúan siendo con los datos auto-escalados, aunque en esta ocasión con los datos sin escalar los resultados fueron buenos. Con respecto a los parámetros de configuración de la SVR, en los mejores resultados para épsilon oscilaron entre 0,01 y 0,05 y los de C estuvieron entre 45 y 52.

En las **Tablas 21**, **22** y **23** se encuentran los mejores resultados para el kernel gaussiano aplicado a los datos de entrenamiento reseñados en la **Tabla 14** para los datos pre-procesados con y sin escalado.

Parámetro(s) del kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
0,012	47	0,02	0,00008	0,44%	0,00008	13,86%
0,0012	47	0,04	0,00012	-2,82%	0,00011	19,30%

Tabla 21. Mejores resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos pre-procesados sin escalar en la etapa de selección de parámetros.

Parámetro(s) del kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
0,009	50	0,02	0,00018	-0,17%	0,00261	0,81%
0,0006	53	0,01	0,00037	-0,23%	0,00230	-1,66%

Tabla 22. Mejores resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos pre-procesados y auto-escalados en la etapa de selección de parámetros.

Parámetro(s) del kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
0,009	50	0,02	0,00009	0,34%	0,00004	4,17%
0,0012	47	0,04	0,00012	-2,49%	0,00015	-10,30%

Tabla 23. Mejores resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos pre-procesados y centrados en la etapa de selección de parámetros.

En los resultados mostrados en las **Tablas 21**, **22** y **23,** se puede evidenciar en el kernel gaussiano, que el parámetro gama toma valores cercanos a 0,009. En términos generales, los mejores resultados siguen dando con los datos auto-escalados, los datos centrados volvieron a ocupar el segundo lugar y de último lugar los datos sin escalar. Con respecto a los parámetros de configuración de la SVR, épsilon coincidió en los mejores resultados con valores de 0,02 y el parámetro *C* estuvo cercano a 50 (resaltado en las **Tablas 21, 22** y **23**).

Como resultado de esta sección hay que destacar que los mejores resultados se obtuvieron con el kernel gaussiano y kernel lineal, además con los 3 kernel probados los mejores resultados dieron con los datos auto-escalados, esto es un nuevo punto de partida para las secciones siguientes.

3.4 VALIDACIÓN DEL MODELO DE INFERENCIA PROPUESTO

En la sección anterior, donde se realizó la sintonización de parámetros para los kernel lineal, polinomial y gaussiano, se tuvieron en cuenta un grupo de medidas reducido para poder obtener unos estimados de los parámetros óptimos y que los resultados se dieran en un menor tiempo.

Para las pruebas realizadas en esta sección solo se utilizaron el kernel lineal y el kernel gaussiano, ya que fueron los que arrojaron los mejores resultados. Además se realizaron los procesos de entrenamiento con los conjuntos de medidas de la base de datos A-NOSE más uniformes, en este caso todos los que poseían 5 medidas, los cuales son la mayoría y corresponden a 24 productos, para un total de 120 medidas (**Tabla 24**).

Aunque ya se había realizado un proceso de sintonización de parámetros, para los nuevos datos de entrenamiento se realizó nuevamente este proceso para corroborar que los parámetros escogidos fueran los óptimos y para obtener unos parámetros generalizados conforme a la base de datos, pero tomando como punto inicial los parámetros con los cuales se obtuvo los mejores resultados en la sintonización de parámetros (sección 3.3).

3.4.1 Primera Validación

Cabe recordar que la primera validación se realizó con el objetivo de establecer el comportamiento del sensor virtual ante otras medidas diferentes a las utilizadas en el entrenamiento pero de productos conocidos.

Con las 120 medidas de la base de datos A-NOSE se construyeron 5 paquetes cada uno con una muestra de cada producto, los cuales se constituyeron de forma aleatoria, utilizando un algoritmo propio para ello. Por lo tanto cada paquete contiene 24 medidas de cada uno de los productos reseñadas en la **Tabla 24**.

Se aplicó validación cruzada de 5 particiones de la siguiente forma: se realizó el entrenamiento con 4 paquetes y se validó con el paquete restante, esto se repitió 5 veces. A continuación se muestran los resultados obtenidos.

En la **Tabla 25** y **Tabla 26** se encuentran los resultados de la validación para el kernel lineal y el kernel gaussiano respetivamente, con el conjunto de datos reseñados en la **Tabla 24** pre-procesados y auto-escalados.

Cabe aclarar que los parámetros de configuración de la SVR son diferentes debido a que se realizó una sintonización de dichos parámetros en cada proceso de validación con el propósito de encontrar los parámetros óptimos para cada caso.

Producto	Nombre de los ficheros (Especímenes de cada producto)
Aguardiente	AGUAR-01.txt, AGUAR-02.txt, AGUAR-03.txt, AGUAR-04.txt, AGUAR-05.txt
Café (blanqueado y vinagre)	BL&V1.txt, BL&V2.txt, BL&V3.txt, BL&V4.txt, BL&V5.txt
Café (blanqueado)	BLAN1.txt, BLAN2.txt, BLAN3.txt, BLAN4.txt, BLAN5.txt
Durazno	DUR-1.txt, DUR-2.txt, DUR-3.txt, DUR-4.txt, DUR-5.txt
Etanol al 10%	ETA10-01.txt, ETA10-02.txt, ETA10-03.txt, ETA10-04.txt, ETA10-05.txt
Etanol al 25%	ETA25-01.txt, ETA25-02.txt, ETA25-03.txt, ETA25-04.txt, ETA25-05.txt
Etanol al 50%	ETA50-01.txt, ETA50-02.txt, ETA50-03.txt, ETA50-04.txt, ETA50-05.txt
Etanol al 95%	ETA95-04.txt, ETA95-05.txt, ETA95-07.txt, ETA95-08.txt, ETA95-09.txt
Café Excelso UGQ	EUGQ1.txt, EUGQ2.txt, EUGQ3.txt, EUGQ4.txt, EUGQ5.txt
Café Excelso UGQ (otra molienda)	EUGQ6.txt, EUGQ7.txt, EUGQ8.txt, EUGQ9.txt, EUGQX.txt
Café Excelso	EXCE1.txt, EXCE2.txt, EXCE3.txt, EXCE4.txt, EXCE5.txt
Maracuyá	MAR-1.txt, MAR-2.txt, MAR-3.txt, MAR-4.txt, MAR-5.txt
Metanol al 10%	MET10-01.txt, MET10-02.txt, MET10-03.txt, MET10-04.txt, MET10-05.txt
Metanol al 25%	MET25-01.txt, MET25-02.txt, MET25-03.txt, MET25-04.txt, MET25-05.txt
Metanol al 50%	MET50-01.txt, MET50-02.txt, MET50-03.txt, MET50-04.txt, MET50-05.txt
Metanol al 95%	MET95-03.txt, MET95-04.txt, MET95-06.txt, MET95-07.txt, MET95-08.txt
Café (negro)	NEGR1.txt, NEGR2.txt, NEGR3.txt, NEGR4.txt, NEGR5.txt
Café (pasilla de máquina)	PMAQ1.txt,PMAQ2.txt,PMAQ3.txt,PMAQ4.txt, PMAQ5.txt
Café (pasilla con químico)	PQ1T1.txt, PQ1T2.txt, PQ1T3.txt, PQ1T4.txt, PQ1T5.txt
Café (Toledo)	TEMQ1.txt, TEMQ2.txt, TEMQ3.txt, TEMQ4.txt, TEMQ5.txt
Café (vinagre)	VINA1.txt, VINA2.txt, VINA3.txt, VINA4.txt, VINA5.txt
Vino blanco	VINBL-01.txt, VINBL-02.txt, VINBL-03.txt, VINBL-04.txt, VINBL-05.txt
Vino de naranja	VINOR-01.txt, VINOR-02.txt, VINOR-03.txt, VINOR-04.txt, VINOR-05.txt
Vino tinto	VINTI-02.txt, VINTI-03.txt, VINTI-04.txt, VINTI-05.txt, VINTI-06.txt

Tabla 24. Medidas de la base de datos A-NOSE empleadas en la validación

Validación	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	50	0,06	0,00033	0,94%	0,00063	0,95%
Validación2	50	0,09	0,00035	-3,61%	0,00071	-3,89%
Validación3	56	0,05	0,00031	0,17%	0,00063	0,18%
Validación4	51	0,08	0,00033	1,40%	0,00069	1,41%
Validación5	54	0,04	0,00029	-0,53%	0,00076	-0,53%
			Promedio del |Error relativo|	1,33%	Promedio del |Error relativo|	1,39%

Tabla 25. Mejores resultados de entrenamiento con kernel lineal aplicado al conjunto de datos auto-escalados

Mientras que en la **Tabla 27** y **Tabla 28** se encuentran los resultados de la validación para el kernel lineal y el kernel gaussiano respectivamente, con el conjunto de datos reseñados en la **Tabla 24** pre-procesados y centrados.

Validación	Parámetro del Kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	0,009	50	0,01	0,00010	0,94%	0,00031	1,09%
Validación2	0,009	50	0,02	0,00010	0,26%	0,00080	-3,33%
Validación3	0,0095	51	0,01	0,00011	1,12%	0,00025	0,39%
Validación4	0,009	50	0,02	0,00011	0,27%	0,00030	0,14%
Validación5	0,009	50	0,02	0,00010	0,11%	0,00032	0,88%
				Promedio del |Error relativo|	0,81%	Promedio del |Error relativo|	1,17%

Tabla 26. Mejores resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados

Validación	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	55	0,07	0,00018	-63,09%	0,00035	-74,99%
Validación2	50	0,06	0,00018	-74,77%	0,00040	-65,58%
Validación3	56	0,06	0,00017	-59,44%	0,00033	-71,57
Validación4	56	0,06	0,00018	-61,94%	0,00035	-71,25%
Validación5	50	0,06	0,00017	-60,94%	0,00035	-69,48%
			Promedio del |Error relativo|	64,04%	Promedio del |Error relativo|	56,32%

Tabla 27. Mejores resultados de entrenamiento con kernel lineal aplicado al conjunto de datos centrados

Validación	Parámetro del Kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	0,008	48	0,01	0,00002	-0,18%	0,00003	-1,25%
Validación2	0,009	50	0,01	0,00002	-2,30%	0,00008	0,01%
Validación3	0,008	48	0,01	0,00002	-0,23%	0,00004	-0,92%
Validación4	0,0105	53	0,01	000002	2,58%	0,00005	3,23%
Validación5	0,0095	51	0,01	0,00002	1,23%	0,00005	1,14%
				Promedio del |Error relativo|	1,30%	Promedio del |Error relativo|	1,31%

Tabla 28. Mejores resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos centrados

Como resultado de la primera validación se puede concluir que los mejores resultados se obtuvieron con los datos auto-escalados para el kernel gaussiano y los segundos mejores resultados se dieron para los datos centrados para el kernel gaussiano. En ambos casos coincide el kernel gaussiano para los mejores resultados, con una ligera diferencia en el error relativo entre para los datos auto-escalados y centrados, con valores de 0,81% y 1,30% respectivamente para el error de entrenamiento y de 1,17% y 1,31% respectivamente para el error de validación. Aunque el error relativo para el kernel lineal y los datos auto-escalados es 1,33% para el entrenamiento y 1,395 para la validación, muy similar al kernel gaussiano con los datos centrados.

3.4.2 Segunda Validación

Cabe recordar que la segunda validación se realizó con el objetivo de establecer el comportamiento del sensor virtual ante medidas de productos diferentes a las utilizadas en el entrenamiento.

De los paquetes que se armaron para la primera validación (sección 3.4.1) se escogió uno, cabe recordar que cada paquete contiene 24 medidas de productos diferentes como se reseña en **Tabla 24**.

Se aplicó validación cruzada de 6 particiones de la siguiente forma: se realizó el entrenamiento con 20 medidas y se validó con las 4 restantes, esto se repitió 6 veces. A continuación se muestra los resultados obtenidos.
En la **Tabla 29** y **Tabla 30** se encuentran los resultados de la validación para el kernel lineal y el kernel gaussiano respetivamente, con el conjunto de datos reseñados en la **Tabla 24** pre-procesados y auto-escalados.

Los parámetros de configuración de la SVR que se tomaron fueron los mismos en cada validación con el objetivo de comparar.

Validación	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	56	0,05	0,00081	3,31%	0,00143	6,70%
Validación2	56	0,05	0,00080	1,03%	0,00158	4,91%
Validación3	56	0,05	0,00072	-2,04%	0,00854	-32,43%
Validación4	56	0,05	0,00085	1,02%	0,00047	-11,39%
Validación5	56	0,05	0,00049	5,02%	0,01029	-31,28%
Validación6	56	0,05	0,00084	1,34%	0,00041	2,85%
			Promedio del |Error relativo|	2,29%	Promedio del |Error relativo|	14,93%

Tabla 29. Mejores resultados de entrenamiento con kernel lineal aplicado al conjunto de datos auto-escalados

De acuerdo a los resultados mostrados en la **Tabla 29** y **Tabla 30** se puede observar que los mejores resultados se obtuvieron con el kernel gaussiano. Se puede observar que en estas validaciones solo se experimentó con los datos auto-escalados debido a que en las validaciones anteriores también concordó con los mejores resultados.

Validación	Parámetro del Kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	0,009	50	0,02	0,00017	-0,35%	0,00224	-13,53%
Validación2	0,009	50	0,02	0,00020	0,07%	0,00096	5,82%
Validación3	0,009	50	0,02	0,00017	-0,62%	0,00660	-23,62%
Validación4	0,009	50	0,02	0,00022	-0,15%	0,00046	-12,83%
Validación5	0,009	50	0,02	0,00015	-0,99	0,00623	14,49%
Validación6	0,009	50	0,02	0,00022	-0,07%	0,00040	-4,50%
				Promedio del |Error relativo|	0,37%	Promedio del |Error relativo|	12,46%

Tabla 30. Mejores resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados

3.4.3 Tercera Validación

En esta validación se escogió para obtener el modelo del sensor virtual el sensor cinco (S5) a diferencia de la primera y segunda validación donde se escogió el sensor seis (S6) y se utilizaron medidas de la base de datos A-NOSE.

Al igual que en la segunda validación de los paquetes que se armaron (sección 3.4.1) se escogió uno (en este caso el PAQUETE1aux), cabe recordar que cada paquete contiene 24 medidas de productos diferentes como se reseña en **Tabla 24**.

Analizando la covarianza de la columna cinco la cual corresponde a los datos del sensor cinco (S5) con cada una de las columnas del PAQUETE1aux que corresponden a los demás sensores, se obtuvieron los datos de correlación lineal que se muestran en la **Tabla 31**.

S1 & S5	S2 & S5	S3 & S5	S4 & S5	S5 & S5	S6 & S5	S7 & S5	S8 & S5
-0,1240	0,8202	0,2677	0,9736	1,00	0,8891	0,9014	0,9588

Tabla 31. Correlación de Pearson para una medida de la base de datos A-NOSE

Analizando la correlación de la **Tabla 31** se puede encontrar que solo dos de los siete sensores diferentes al sensor cinco (S5) tienen una correlación menor comparada con los demás, el sensor uno (S1) tiene una correlación negativa débil y el sensor tres (S3) tiene una correlación positiva débil.

En esta validación al igual que en la segunda validación (sección 3.4.2) también se aplicó validación cruzada de 6 particiones de la siguiente forma: se realizó el entrenamiento con 20 medidas y se validó con las 4 restantes, esto se repitió 6 veces.

A continuación se presentan un resumen de los resultados obtenidos en cada caso para el kernel gaussiano (**Tabla 32**) y para el kernel lineal (**Tabla 33**), en este caso se utilizaron los siete sensores restantes para obtener el modelo del sensor virtual del sensor cinco (S5).

Validación	Parámetro del Kernel	C	Épsilon	Error relativo (Entrenamiento)	Error relativo (Validación)
Validación1	0,009	50	0,02	0,07%	-29,44%
Validación2	0,009	50	0,02	-0,08%	62,63%
Validación3	0,009	50	0,02	0,28%	31,55%
Validación4	0,009	50	0,02	-0,17%	-12,08%
Validación5	0,009	50	0,02	-1,02%	-25,38%
Validación6	0,009	50	0,02	-0,13%	0,34%
Promedio del |Error relativo|	-	-	-	0,29%	26,90%

Tabla 32. Resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados

De acuerdo a las **Tablas 32** y **33** se puede apreciar que tanto para el kernel gaussiano y el kernel lineal, el mejor resultado fue para la Validación6 teniendo en cuenta el error relativo de validación 0,34% y 0,92% respectivamente. Teniendo en cuenta el error relativo de entrenamiento se observa que el mejor resultado es para la Validación1 con el kernel gaussiano y la Validación6 con el kernel lineal con valores de 0,07% y 0,02% respectivamente. Analizando todos los resultados de estas validaciones se vuelve a

constatar que el kernel gaussiano tiene el mejor comportamiento en la validación, en los casos donde el error relativo es muy alto puede estar sucediendo que los datos de entrenamiento no son representativos frente a los posibles datos de entrada y se presenta el error de falta de generalización.

Validación	Parámetro del Kernel	C	Épsilon	Error relativo (Entrenamiento)	Error relativo (Validación)
Validación1	No aplica	56	0,05	-2,55%	-7,39%
Validación2	No aplica	56	0,05	-3,02%	-116,21%
Validación3	No aplica	56	0,05	-6,017%	32,26%
Validación4	No aplica	56	0,05	-0,98%	-96,29%
Validación5	No aplica	56	0,05	-2,01%	-30,84%
Validación6	No aplica	56	0,05	-0,02%	0,92%
Promedio del |Error relativo|	-	-	-	2,43%	47,32%

Tabla 33. Resultados de entrenamiento con kernel lineal aplicado al conjunto de datos auto-escalados

En la **Figura 33** se puede observar la medida del sensor real (línea azul) y medida obtenida con el sensor virtual (línea roja) para el conjunto de datos de entrenamiento, en la gráfica se ve como la medida del sensor virtual prácticamente se sobrepone sobre la medida del sensor real. La grafica para el conjunto de datos de validación se puede observar en la Figura 34, en este caso se puede observar que aunque la medida del sensor virtual no se sobrepone completamente a la medida del sensor real, si logra seguir u obtener el comportamiento del sensor. Ambas graficas fueron obtenidas con los datos de la Validación6 donde se empleó el kernel gaussiano.

En la **Figura 35** se puede observar la medida del sensor real y medida obtenida con el sensor virtual para el conjunto de datos de entrenamiento. La grafica para el conjunto de datos de validación se puede observar en la **Figura 36**. Ambas graficas fueron obtenidas con los datos de la Validación6 donde se empleó el lineal.

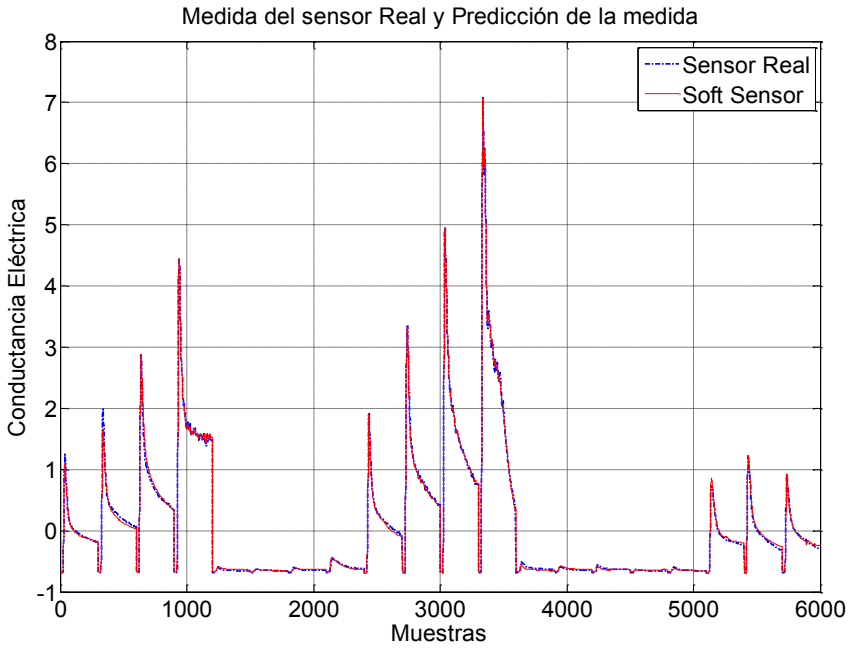


Figura 33. Señal del señor real y del sensor virtual (kernel gaussiano) para los datos de entrenamiento

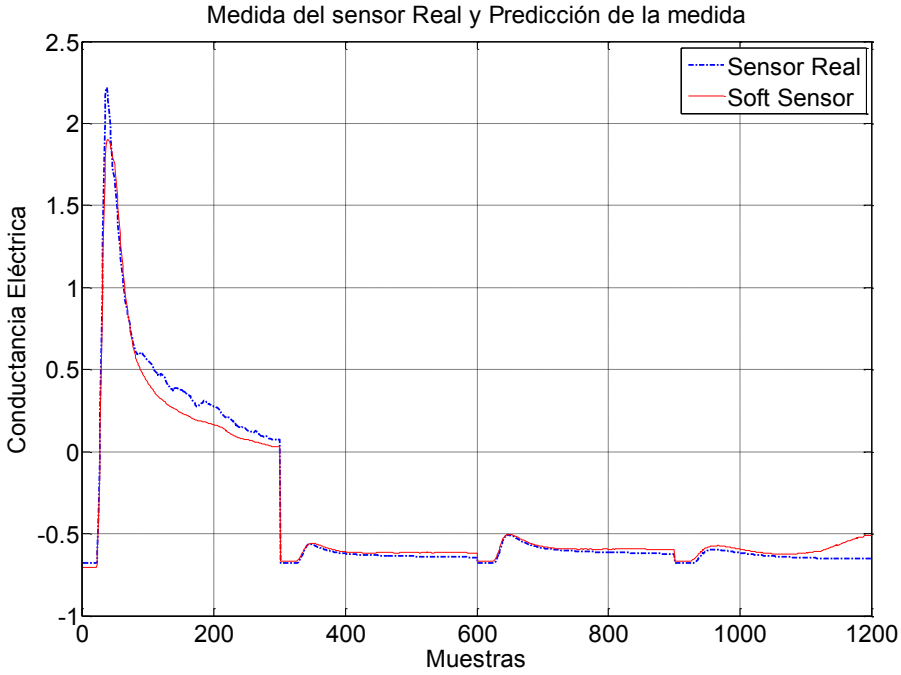


Figura 34. Señal del señor real y del sensor virtual (kernel gaussiano) para los datos de validación

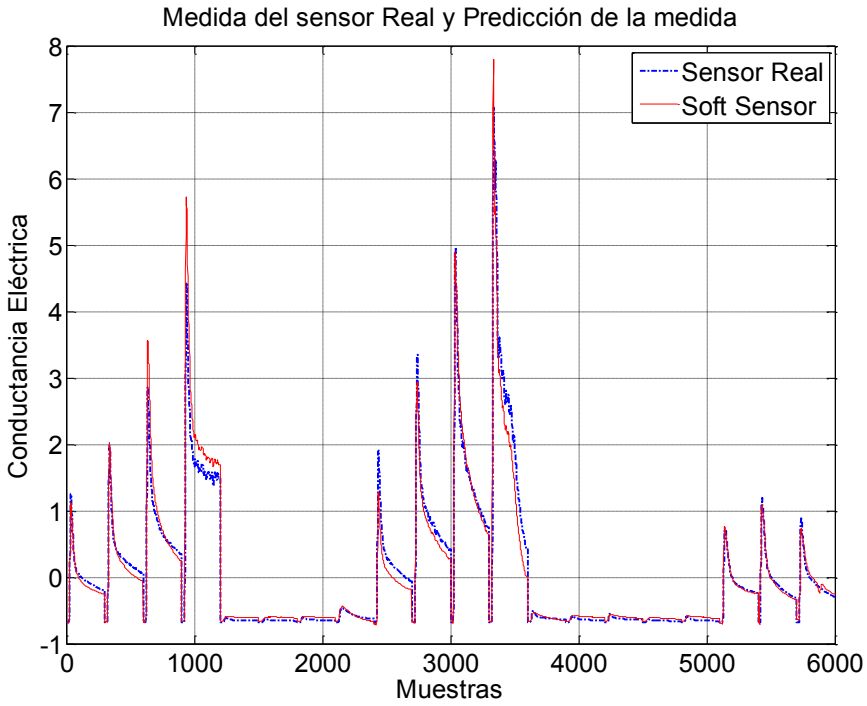


Figura 35. Señal del señor real y del sensor virtual (kernel lineal) para los datos de entrenamiento

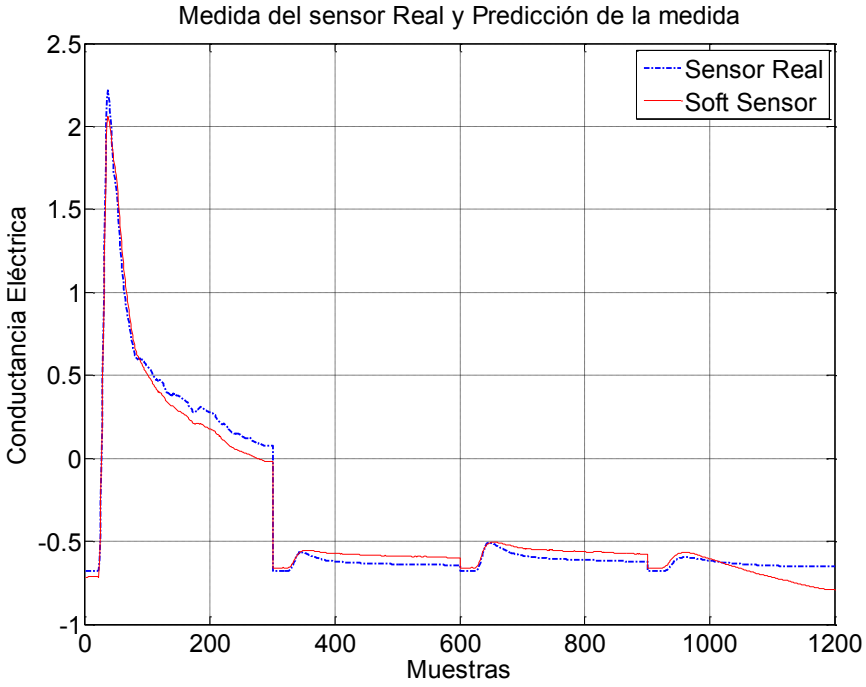


Figura 36. Señal del señor real y del sensor virtual (kernel lineal) para los datos de validación

Continuando con los mismos datos y procedimientos empleados anteriormente, se realizaron otros entrenamientos y validaciones omitiendo el aporte desde uno hasta cinco sensores, lo que significa que se obtuvo el modelo del sensor virtual a partir de seis sensores, cinco sensores, cuatro sensores, hasta un mínimo de dos sensores. A continuación se muestran los resultados con el kernel gaussiano para cada caso.

Validación	Parámetro del Kernel	C	Épsilon	Error relativo (Entrenamiento)	Error relativo (Validación)
Validación1	0,009	50	0,02	0,001%	-25,72%
Validación2	0,009	50	0,02	0,22%	49,94%
Validación3	0,009	50	0,02	0,13%	34,21%
Validación4	0,009	50	0,02	0,07%	-8,59%
Validación5	0,009	50	0,02	-0,89%	-28,97%
Validación6	0,009	50	0,02	-0,05%	2,66%
Promedio del |Error relativo|	-	-	-	0,23%	25,01%

Tabla 34. Resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados, omitiendo sensor uno (S1)

Validación	Parámetro del Kernel	C	Épsilon	Error relativo (Entrenamiento)	Error relativo (Validación)
Validación1	0,009	50	0,02	0,23%	-21,29%
Validación2	0,009	50	0,02	-0,02%	71,29%
Validación3	0,009	50	0,02	-0,33%	32,65%
Validación4	0,009	50	0,02	-0,01%	-33,35%
Validación5	0,009	50	0,02	-0,68%	-25,56%
Validación6	0,009	50	0,02	-0,03%	1,92%
Promedio del |Error relativo|	-	-	-	0,22%	31,01%

Tabla 35. Resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados, omitiendo sensor tres (S3)

Validación	Parámetro del Kernel	C	Épsilon	Error relativo (Entrenamiento)	Error relativo (Validación)
Validación1	0,009	50	0,02	0,89%	-20,92%
Validación2	0,009	50	0,02	0,98%	46,31%
Validación3	0,009	50	0,02	-0,62%	32,06%
Validación4	0,009	50	0,02	1,03%	-18,36%
Validación5	0,009	50	0,02	-0,86%	-28,10%
Validación6	0,009	50	0,02	0,88%	5,32%
Promedio del |Error relativo|	-	-	-	0,88%	28,23%

Tabla 36. Resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados, omitiendo los sensores uno y tres (S1 y S3)

Validación	Parámetro del Kernel	C	Épsilon	Error relativo (Entrenamiento)	Error relativo (Validación)
Validación1	0,009	50	0,02	0,94%	-15,40%
Validación2	0,009	50	0,02	0,77%	57,17%
Validación3	0,009	50	0,02	-0,81%	30,89
Validación4	0,009	50	0,02	0,85%	-13,14
Validación5	0,009	50	0,02	-0,71%	-28,45
Validación6	0,009	50	0,02	0,47%	4,50
Promedio del |Error relativo|	-	-	-	0,76%	24,92%

Tabla 37. Resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados, omitiendo los sensores uno, tres y dos (S1, S3 y S2)

Validación	Parámetro del Kernel	C	Épsilon	Error relativo (Entrenamiento)	Error relativo (Validación)
Validación1	0,009	50	0,02	2,21%	-14,87%
Validación2	0,009	50	0,02	2,47%	32,92%
Validación3	0,009	50	0,02	-0,80%	34,32%
Validación4	0,009	50	0,02	2,39%	-9,22%
Validación5	0,009	50	0,02	0,14%	-29,10%
Validación6	0,009	50	0,02	2,54%	0,92%
Promedio del |Error relativo|	-	-	-	1,76%	20,22%

Tabla 38. Resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados, omitiendo los sensores uno, tres, dos y seis (S1, S3, S2 y S6).

Según los resultados mostrados en las **Tablas** desde la **34** hasta la **39** se encuentra que en cuatro de las seis pruebas realizadas se obtuvieron mejores resultados con menor cantidad de sensores para obtener el sensor virtual que cuando se empleó el mayor número de sensores posibles, para el caso de la base de datos A-NOSE corresponde a siete sensores, especialmente en las últimas pruebas (**Tabla 38** y **39**) en donde el modelo del sensor virtual se obtuvo a partir de solamente tres y dos sensores respectivamente. Con lo anterior evidencia que si la correlación es fuerte entre el sensor que se desea modelar y los

demás sensores los resultados son aún mejores que cuando la correlación es débil, incluso si se obtiene el modelo a partir de dos o tres sensores, lo anterior no quiere decir que no se pueda obtener un modelo si la correlación es baja, simplemente la obtención del modelo es más demorada, ya que requerirá mayores ajustes de los parámetros y los resultados no serán tan buenos.

Validación	Parámetro del Kernel	C	Épsilon	Error relativo (Entrenamiento)	Error relativo (Validación)
Validación1	0,009	50	0,02	0,91%	-21,70%
Validación2	0,009	50	0,02	0,01%	-5,43%
Validación3	0,009	50	0,02	-0,57%	28,71%
Validación4	0,009	50	0,02	-0,23%	-22,06%
Validación5	0,009	50	0,02	-3,03%	-31,03%
Validación6	0,009	50	0,02	0,03%	-5,22%
Promedio del |Error relativo|	-	-	-	0,79%	19,02%

Tabla 39. Resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados, omitiendo los sensores uno, tres, dos, seis y siete (S1, S3, S2, S6 y S7)

3.5. SELECCIÓN DE LA CANTIDAD DE SENSORES

En las pruebas realizadas en la sección 3.4.3 se encontró que emplear un número reducido de sensores para obtener el modelo del sensor virtual, puede funcionar mucho mejor que cuando se utilizan un mayor número de sensores en donde existan algunos en donde la correlación no sea muy fuerte. Dos o tres sensores pueden ser suficientes cuando se desea obtener un soft sensor, siempre y cuando la correlación entre dichos sensores y el sensor que se desea modelar sea fuerte.

Para determinar la cantidad de sensores de gases necesarios para modelar un sensor virtual, se debe tener en cuenta, que más importante que la cantidad de sensores es la correlación existente entre los sensores escogidos y el que se desea modelar.

De la base de datos A-NOSE se tomó una medida al azar y se analizó la correlación de Pearson entre cada sensor y el sensor seis el cual ha sido el que se ha empleado para modelar en todas las pruebas, los resultados de la correlación se muestran en la **Tabla 40**.

S1 & S6	S2 & S6	S3 & S6	S4 & S6	S5 & S6	S6 & S6	S7 & S6	S8 & S6
0,48	0,98	-0,31	0,94	0,94	1,00	0,99	0,93

Tabla 40. Correlación de Pearson para una medida de la base de datos A-NOSE

Analizando la correlación de la **Tabla 40** se puede encontrar que solo dos (2) de los siete (7) sensores diferentes al sensor seis (S6) tienen una correlación menor comparada con los demás, el sensor uno (S1) tiene una correlación positiva moderada y el sensor tres (S3) tiene una correlación negativa débil. Partiendo de este hecho se repitió la segunda validación de la sección 3.4.2**,** pero sin tener en cuenta a los sensores (S1 y S3). Los resultados de estas validaciones se muestran en la **Tabla 41** para el kernel lineal y en la **Tabla 42** para el kernel gaussiano.

Allí se evidencia que para el kernel lineal hubo una reducción en el error de entrenamiento pasando de 2,27% a1,52%, pero aumento el error de validación pasando de 14,93% a 17,92%, mientras que para el kernel gaussiano aumento un poco el error de entrenamiento pasando de 0,37% a 0,73% y el error de validación pasando de 12,46% a 12,97%. Lo anterior era de esperarse debido a que suprimió información y los parámetros de configuración de la SVR se habían estimado inicialmente para los ocho sensores y no para seis, esto supondría realizar una nueva sintonización de parámetros ya que las condiciones han cambiado. Adicionalmente si te tiene en cuenta que los errores de validación están un poco altos, se estaría cayendo en un problema de falta de generalización.

Sin embargo cabe recordar que en esta validación solo se utilizó una medida de cada producto, lo recomendable es utilizar varias medidas de un mismo producto como se realizó para la primera validación de la sección 3.4.1, con el propósito de obtener una mayor generalización y mejores resultados.

Validación	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	56	0,05	0,00092	-1,00%	0,00131	1,44%
Validación2	56	0,05	0,00094	-0,59	0,00082	-26,40%
Validación3	56	0,05	0,00043	-2,58%	0,00883	-24,99%
Validación4	56	0,05	0,00089	0,30%	0,00069	-18,03%
Validación5	56	0,05	0,00072	-4,19%	0,00813	-30,23%
Validación6	56	0,05	0,00083	0,49%	0,00151	7,00%
			Promedio del |Error relativo|	1,52%	Promedio del |Error relativo|	17,92%

Tabla 41. Mejores resultados de entrenamiento con kernel lineal aplicado al conjunto de datos auto-escalados

Validación	Parámetro del Kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	0,009	50	0,02	0,00033	-0,85%	0,00136	2,08%
Validación2	0,009	50	0,02	0,00036	0,80%	0,00033	5,47%
Validación3	0,009	50	0,02	0,00020	-0,79%	0,00922	-28,27%
Validación4	0,009	50	0,02	0,00037	0,80%	0,00023	-4,93%
Validación5	0,009	50	0,02	0,00024	0,55%	0,00812	29,96%
Validación6	0,009	50	0,02	0,00033	0,61%	0,00096	7,13%
				Promedio del |Error relativo|	0,73%	Promedio del |Error relativo|	12,97%

Tabla 42. Mejores resultados de entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados

3.6 OTRAS PRUEBAS Y RESULTADOS

Para las pruebas realizadas en esta sección se utilizó solamente el kernel gaussiano, ya que es el que ha arrojado los mejores resultados. Además se realizaron los procesos de entrenamiento con los conjuntos de medidas de la base de datos B-NOSE más uniformes, en este caso todos los que poseían 5 medidas, los cuales son la mayoría y corresponden a 19 productos, para un total de 95 medidas (**Tabla 43**).

De la misma forma que se hizo con la base de datos A-NOSE, con la base de datos B-NOSE también se armaron cinco paquetes, cabe recordar que cada paquete contiene 19 medidas de productos diferentes como se reseña en **Tabla 43**.

3.6.1 Primera validación base de datos B-NOSE

Esta validación con la base de datos B-NOSE se realizó con el objetivo de obtener un sensor virtual para esta base de datos, una vez hecho esto el objetivo es establecer el comportamiento de este sensor virtual ante otras medidas diferentes a las utilizadas en el entrenamiento pero de productos conocidos.

Los parámetros que se emplearon para la configuración de la SVR son los mismo que se obtuvieron en el proceso de sintonización de parámetros que se realizó con las medidas de la base de datos A-NOSE y que se emplearon en la secciones 3.4.2 y 3.4.3 (segunda validación y tercera validación, respectivamente). No se realizó una nueva sintonización con los datos de entrenamiento para la base de datos B-NOSE, con el objetivo de corroborar que los parámetros escogidos son adecuados y para establecer unos parámetros generalizados independientes de la base de datos y de la naturaleza de los mismos.

Producto	Nombre de los ficheros (Especímenes de cada producto)
Acetona	Acetona2_1.txt, Acetona2_2.txt, Acetona2_3.txt, Acetona2_4.txt, Acetona2_5.txt, acetona_1.txt, acetona_2.txt, acetona_3.txt, 'acetona_4.txt, 'acetona_5.txt'
Benceno	Benceno1000ppm_1.txt, Benceno1000ppm_2.txt, Benceno1000ppm_3.txt, Benceno1000ppm_4.txt, Benceno1000ppm_5.txt, Benceno1500ppm_1.txt, Benceno1500ppm_2.txt, Benceno1500ppm_3.txt, Benceno1500ppm_4.txt, Benceno1500ppm_5.txt, Benceno500ppm_1.txt, Benceno500ppm_2.txt, Benceno500ppm_3.txt, Benceno500ppm_4.txt, Benceno500ppm_5.txt, Benceno_1.txt, Benceno_2.txt, Benceno_3.txt, Benceno_4.txt, Benceno_5.txt
Etanol	Etanol_1.txt, Etanol_2.txt, Etanol_3.txt, Etanol_4.txt, Etanol_5.txt
Hidróxidodamonio	hidr¢xidodamonio_1.txt, hidr¢xidodamonio_2.txt, hidr¢xidodamonio_3.txt, hidr¢xidodamonio_4.txt, hidr¢xidodamonio_5.txt, idroxidodamonio_1.txt, idroxidodamonio_2.txt, idroxidodamonio_3.txt, idroxidodamonio_4.txt, idroxidodamonio_5.txt
Metanol	metanol_1.txt, metanol_2.txt, metanol_3.txt, metanol_4.txt, metanol_5.txt
Tolueno	Tolueno1000ppm_1.txt, Tolueno1000ppm_2.txt, Tolueno1000ppm_3.txt, Tolueno1000ppm_4.txt, Tolueno1000ppm_5.txt, Tolueno1500ppm_1.txt, Tolueno1500ppm_2.txt, Tolueno1500ppm_3.txt, Tolueno1500ppm_4.txt, Tolueno1500ppm_5.txt Tolueno500ppm_1.txt, Tolueno500ppm_2.txt, Tolueno500ppm_3.txt, Tolueno500ppm_4.txt, Tolueno500ppm_5.txt, Tolueno_1.txt, Tolueno_2.txt, Tolueno_3.txt, 'Tolueno_4.txt, Tolueno_5.txt
Vacio	Vacio_1.txt, Vacio_2.txt, Vacio_3.txt, Vacio_4.txt, Vacio_5.txt
Xileno	Xileno_1.txt, Xileno_2.txt, Xileno_3.txt, Xileno_4.txt, Xileno_5.txt, xileno1000ppm_1.txt, xileno1000ppm_2.txt, xileno1000ppm_3.txt, xileno1000ppm_4.txt, xileno1000ppm_5.txt, xileno1500ppm_1.txt, xileno1500ppm_2.txt, xileno1500ppm_3.txt, xileno1500ppm_4.txt, xileno1500ppm_5.txt, xileno500ppm_1.txt, xileno500ppm_2.txt, xileno500ppm_3.txt, xileno500ppm_4.txt, xileno500ppm_5.txt

Tabla 43. Medidas de la base de datos B-NOSE empleadas en las pruebas de esta sección

Se tomó una medida al azar de la base de datos B-NOSE y se obtuvo la correlación de Pearson para cada uno de los sensores con respecto a los demás (**Tabla 44**), en la misma tabla se encuentran los valores promedio para la correlación de cada sensor con respecto a los otros y allí se puede apreciar que los sensores que tienen la correlación más baja con respecto a los demás sensores son el sensor siete (S7) y el sensor once (S11), sin embargo estos valores de correlación pertenecen a la correlación positiva fuerte ya que son de 0,75 en los dos casos. Analizando los valores de correlación para los dos sensores mencionados se encontró que la mayor correlación de estos se da con los sensores uno, cuatro, nueve y diez (S1, S4, S9 y S10) en ambos casos y la menor correlación se da con los sensores dos, doce, catorce y dieciséis (S2, S12, S14 y S16) en ambos casos también. Se escogió el sensor siete (S7) para obtener el modelo del sensor virtual y los demás sensores son los que servirán para la obtención de dicho modelo.

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**	**S13**	**S14**	**S15**	**S16**	$\overline{x}$
S1	1.00	0.73	0.88	0.98	0.84	0.99	0.91	0.83	0.96	0.96	0.91	0.78	0.97	0.79	0.83	0.72	0.88
S2	0.73	1.00	0.92	0.61	0.96	0.81	0.49	0.98	0.53	0.53	0.49	0.77	0.74	0.97	0.96	0.97	0.78
S3	0.88	0.92	1.00	0.81	0.94	0.93	0.64	0.95	0.72	0.72	0.64	0.93	0.87	0.97	0.99	0.94	0.87
S4	0.98	0.61	0.81	1.00	0.74	0.95	0.92	0.72	0.98	0.98	0.92	0.74	0.94	0.70	0.75	0.63	0.84
S5	0.84	0.96	0.94	0.74	1.00	0.90	0.66	0.99	0.69	0.69	0.65	0.81	0.87	0.94	0.96	0.91	0.85
S6	0.99	0.81	0.93	0.95	0.90	1.00	0.85	0.89	0.91	0.91	0.85	0.82	0.98	0.87	0.91	0.80	0.90
S7	0.91	0.49	0.64	0.92	0.66	0.85	1.00	0.64	0.97	0.97	1.00	0.56	0.89	0.52	0.58	0.43	0.75
S8	0.83	0.98	0.95	0.72	0.99	0.89	0.64	1.00	0.67	0.67	0.64	0.81	0.84	0.96	0.96	0.94	0.84
S9	0.96	0.53	0.72	0.98	0.69	0.91	0.97	0.67	1.00	1.00	0.97	0.65	0.93	0.60	0.66	0.52	0.80
S10	0.96	0.53	0.72	0.98	0.69	0.91	0.97	0.67	1.00	1.00	0.97	0.65	0.93	0.60	0.66	0.52	0.80
S11	0.91	0.49	0.64	0.92	0.65	0.85	1.00	0.64	0.97	0.97	1.00	0.55	0.89	0.52	0.57	0.43	0.75
S12	0.78	0.77	0.93	0.74	0.81	0.82	0.56	0.81	0.65	0.65	0.55	1.00	0.77	0.84	0.87	0.82	0.77
S13	0.97	0.74	0.87	0.94	0.87	0.98	0.89	0.84	0.93	0.93	0.89	0.77	1.00	0.78	0.84	0.70	0.87
S14	0.79	0.97	0.97	0.70	0.94	0.87	0.52	0.96	0.60	0.60	0.52	0.84	0.78	1.00	0.99	0.99	0.81
S15	0.83	0.96	0.99	0.75	0.96	0.91	0.58	0.96	0.66	0.66	0.57	0.87	0.84	0.99	1.00	0.96	0.84
S16	0.72	0.97	0.94	0.63	0.91	0.80	0.43	0.94	0.52	0.52	0.43	0.82	0.70	0.99	0.96	1.00	0.77
$\overline{x}$	0.88	0.78	0.87	0.84	0.85	0.90	0.75	0.84	0.80	0.80	0.75	0.77	0.87	0.81	0.84	0.77	

Tabla 44. Correlación de Pearson para una medida de la base de datos B-NOSE

Cabe recordar que con las 95 medidas de la base de datos A-NOSE se construyeron 5 paquetes cada uno con una muestra de cada producto, los cuales se constituyeron de forma aleatoria, utilizando un algoritmo propio para ello. Por lo tanto cada paquete contiene 24 medidas de cada uno de los productos reseñadas en la **Tabla 43**.

Se aplicó validación cruzada de 5 particiones de la siguiente forma: se realizó el entrenamiento con 4 paquetes y se validó con el paquete restante, esto se repitió 5 veces. A continuación se muestran los resultados obtenidos.

En la **Tabla 45** se encuentran los resultados de la validación para el kernel gaussiano, con el conjunto de datos reseñados en la **Tabla 43** pre-procesados y auto-escalados.

Validación	Parámetro del Kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	0,009	50	0,02	2.22E-05	-1.94%	9.40E-05	-3.83
Validación2	0,009	50	0,02	1.29E-05	-0.82%	7.02E-05	-3.51
Validación3	0,009	50	0,02	1.08E-05	-0.35%	4.09E-05	0.16
Validación4	0,009	50	0,02	1.25E-05	-0.46%	7.40E-05	1.06
Validación5	0,009	50	0,02	1.03E-05	-0.32%	3.11E-05	-0.64
				Promedio del |Error relativo|	0.78%	Promedio del |Error relativo|	1.84%

Tabla 45. Resultados del entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados.

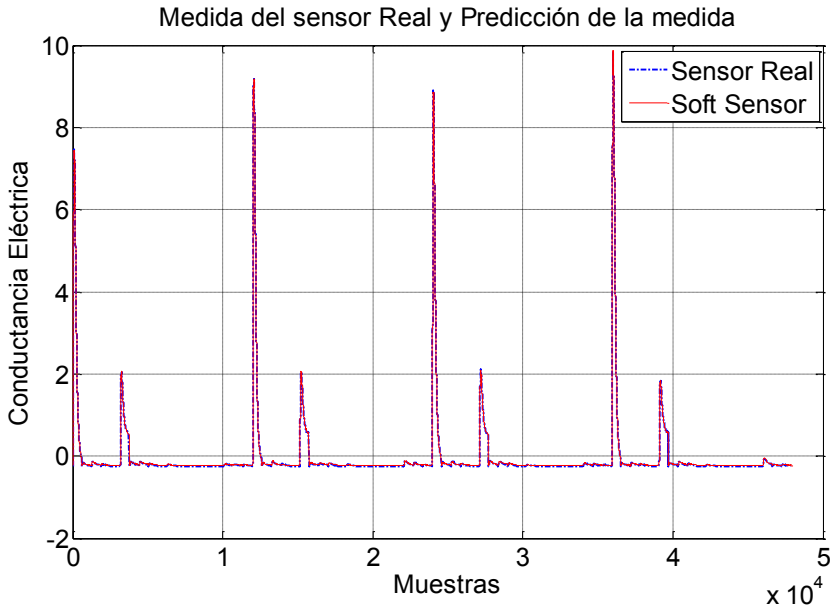


Figura 37. Señal del sensor real (S7) y del sensor virtual (kernel gaussiano) para los datos de entrenamiento

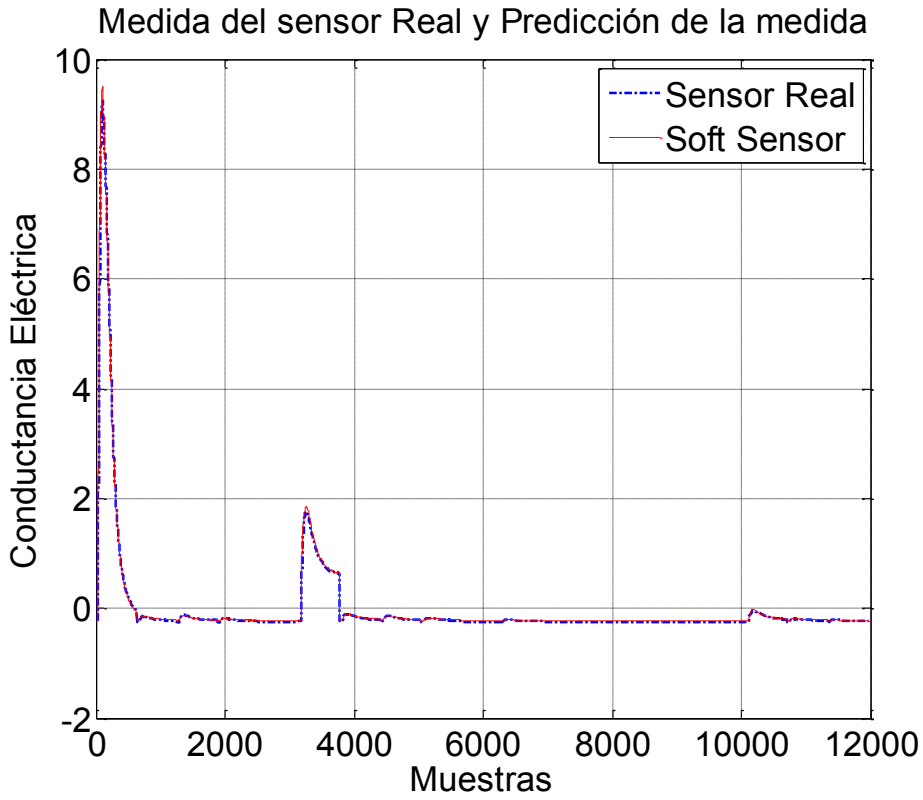


Figura 38. Señal del sensor real (S7) y del sensor virtual (kernel gaussiano) para los datos de entrenamiento

En las **Figuras 37** y **Figura 38** se pueden apreciar los resultados del entrenamiento y la validación para la Validación1 referida en la **Tabla 45**, aunque el sensor escogido para modelar es uno de los dos que tiene la correlación más baja con el resto de los sensores los resultados de estas pruebas tienen un error bajo, claro está que la cantidad de sensores empleada para la obtención del modelo es considerable (quince sensores).

3.6.2 Segunda validación base de datos B-NOSE

De la misma forma que en la sección anterior 3.6.2 se escogió al sensor siete (S7) para obtener el sensor virtual por ser el que tiene la menor correlación con respecto a los demás sensores, en este caso también se escogió dicho sensor pero la diferencia radica en que en este caso se obtuvo el modelo para este sensor virtual a partir de solo cuatro sensores y se escogieron aquellos con los cuales dicho sensor tiene la correlación más baja, como se mencionó en párrafos anteriores son los sensores dos, doce, catorce y dieciséis (S2, S12, S14 y S16).

Es importante aclarar que los parámetros de configuración de la SVR se ajustaron, es decir se realizó un proceso de sintonización previa de dichos parámetros partiendo de los valores base que se han utilizado en las anteriores validaciones, debido a que en las primeras pruebas realizadas en estas condiciones, dichos parámetros no mostraron unos resultados muy adecuados. Los parámetros encontrados en el proceso de sincronización fueron los siguientes: 0,0095 como parámetro del kernel, 51 para *C* y 0,01 para épsilon.

De la misma forma que en la sección 3.6.1, también se aplicó validación cruzada de 5 particiones, se realizó el entrenamiento con 4 paquetes y se validó con el paquete restante, esto se repitió 5 veces. A continuación se muestran los resultados obtenidos.

En la **Tabla 46** se encuentran los resultados de la validación para el kernel gaussiano, con el conjunto de datos reseñados en la **Tabla 43** pre-procesados y auto-escalados.

Validación	Parámetro del Kernel	C	Épsilon	Error cuadrático medio (Entrenamiento)	Error relativo (Entrenamiento)	Error cuadrático medio (Validación)	Error relativo (Validación)
Validación1	0,009	50	0,02	8.38E-05	4.62%	2.16E-04	7.20%
Validación2	0,009	50	0,02	8.42E-05	4.04%	2.25E-04	8.48%
Validación3	0,009	50	0,02	8.99E-05	2.86%	2.52E-04	4.03%
Validación4	0,009	50	0,02	7.66E-05	3.89%	3.01E-04	7.48%
Validación5	0,009	50	0,02	7.76E-05	2.93%	2.75E-04	-3.09%
				Promedio del |Error relativo|	3.67%	Promedio del |Error relativo|	6.06%

Tabla 46. Resultados del entrenamiento con kernel gaussiano aplicado al conjunto de datos auto-escalados.

En las **Figuras 39** y **Figura 40** se pueden apreciar los resultados del entrenamiento y la validación para la Validación2 referida en la **Tabla 46**, aunque el sensor escogido para modelar es uno de los dos que tiene la correlación más baja con el resto de los sensores los resultados de estas pruebas tienen un error bajo, máxime si se tiene en cuenta que se obtuvo el modelo a partir de los cuatro sensores con los cuales se tenía la correlación más baja en la base de datos B-NOSE.

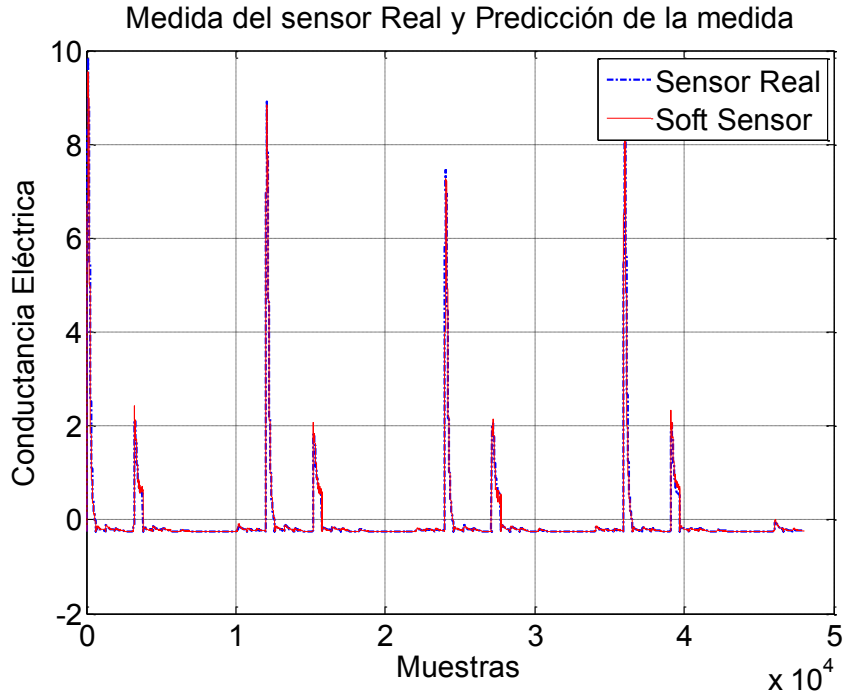


Figura 39. Señal del sensor real (S7) y del sensor virtual (kernel gaussiano) para los datos de entrenamiento

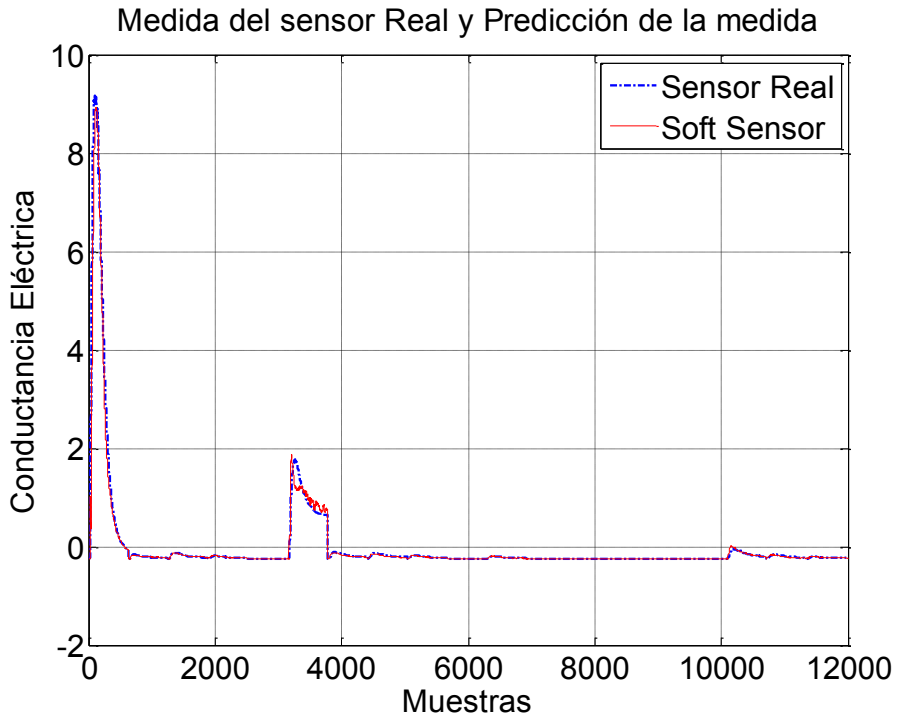


Figura 40. Señal del sensor real (S7) y del sensor virtual (kernel gaussiano) para los datos de entrenamiento

CONCLUSIONES

En este trabajo de grado de maestría se propone una metodología para obtener sensores de gases virtuales, basados en modelos de inferencia a partir de la respuesta de otros sensores de gases del mismo tipo de diferentes referencias, utilizando el método de regresión de soporte vectorial SVR.

A través del análisis de los casos de estudio se evidenció que por medio de la utilización de métodos de regresión basados en aprendizaje de máquina, se puede llegar a caracterizar y modelar la respuesta de los sensores de gases de tipo MOS de forma confiable, permitiendo obtener sensores virtuales que puedan ser utilizados como complemento de los sensores físicos, buscando reducir los problemas ocasionados por la utilización de múltiples sensores de gases en sistemas de olfato electrónico.

Las desventajas o limitantes de esta metodología radican principalmente en la dependencia de los sensores virtuales con respecto a los sensores físicos, razón por la cual estos últimos deben estar funcionando adecuadamente y en buenas condiciones, aunque a decir verdad si los sensores no funcionan adecuadamente en un sistema físico también existen problemas. Además, los sensores virtuales tienen dificultad en la predicción precisa en ambientes o plantas cambiantes, esto último supondría re-calibración o entrenamiento de los mismos si las condiciones varían drásticamente o a medida que los sensores se degraden, sin embargo, no dejan de ser una muy buena alternativa, para reducir costos, tamaño y complejidad del hardware.

La caracterización de sensores de gases permite obtener los rasgos distintivos de la señal entregada, importantes a la hora de realizar el pre-procesamiento de las distintas señales obtenidas con la matriz de sensores de gases, el cual es muy importante ya que un buen pre-procesamiento de la información conduce a obtener buenos resultados posteriormente. En este tipo de señales es conveniente aplicar un buen filtrado de la señal, alguna técnica para mitigar o remover las derivas y un escalado o normalización apropiados. Las técnicas de pre-procesamiento empleadas, tal como, la manipulación de la línea base, el filtrado de la señal y el escalado mostraron ser apropiadas como etapa previa de la regresión, debido a que con ellas se mitigan parte de las derivas, se filtra el ruido y se adecuan los datos para su posterior procesamiento. Además, se pudo establecer que debido a la característica de las señales entregadas por los sensores de gases y la correlación existente entre las mismas, posibilito que se pudiera utilizar regresión lineal múltiple para obtener el modelo de sensor virtual, aunque el método de aprendizaje de máquina SVR realiza la regresión lineal a través de un mapeo no-lineal.

El entrenamiento de la máquina de soporte vectorial para regresión SVR, con el propósito de obtener el modelo de regresión de los sensores de gases, mostró los menores errores de entrenamiento para el kernel gaussiano, como se puede apreciar en los resultados mostrados en la secciones del capítulo 3, y los valores escogidos para los parámetros C y ε

estuvieron en valores 50 y 0,02 respectivamente, aunque estos valores no suponen los valores óptimos para todos los casos, si pueden ser tomados como valores de referencia y punto de partida para la sintonización o escogencia de los mismos.

Con las validaciones realizadas a la metodología propuesta en este trabajo de grado de maestría se logró comprobar que los sensores de gases pueden ser modelados utilizando la técnica de regresión SVR, con el propósito de reemplazar algunos sensores físicos en una matriz de sensores de gases, trayendo consigo la reducción en el hardware y todos los beneficios asociados a ello. Se debe tener en cuenta que deberá existir una base de datos que posibilite el proceso de entrenamiento del sensor virtual.

Es importante tener en cuenta que el proceso de sintonización de parámetros se debe realizar cada vez que se desee obtener un sensor virtual diferente o cuando las condiciones cambien, por ejemplo, cuando los sensores físicos deben ser remplazados o cambiados por otros. También se debe tener en cuenta que a medida que pase el tiempo, debido al deterioro de los sensores y las derivas, sería recomendable ajustar los parámetros nuevamente con el objetivo de tener una medida confiable de los sensores virtuales o emplear técnicas de reducción de derivas que mitiguen el problema.

En las pruebas realizadas en la sección 3.4.3 se encontró que emplear un número reducido de sensores para obtener el modelo del sensor virtual, tres o incluso dos, puede funcionar para obtener un buen modelo, siempre y cuando la correlación entre dichos sensores y el sensor que se desea modelar sea fuerte. Sin embargo, lo anterior no quiere decir que sea estrictamente necesario que se cumpla esto, por ejemplo, cuando la correlación es moderada también se puede obtener un modelo confiable como se mostró en la sección 3.6.2. Pero como se puede evidenciar en los resultados mostrados en la sección 3.6.1. es común encontrar sensores de gases que estén correlacionados con otros de diferentes referencias, facilitando el trabajo y reduciendo los tiempos de sintonización y entrenamiento.

En las diferentes pruebas realizadas de sintonización de parámetros se evidencio que el costo computacional está relacionado con el tipo de kernel empleado, la cantidad de los datos a analizar, el método de escalado de los datos utilizado y la cantidad de iteraciones que se deban llevar a cabo en la búsqueda sistemática, debido a que en cada iteración lo que se busca es reducir el error cuadrático medio entre la medida del sensor real y del sensor virtual, si en cada iteración existe una reducción de dicho error el proceso de sintonización es más rápido y se realizaran menos iteraciones, pero si por el contrario no se evidencia una reducción en dicho error el software deberá continuar con la búsqueda de unos mejores parámetros y por lo tanto el tiempo de ejecución aumentara. Por otro lado, una vez obtenidos los parámetros de configuración, los procesos de entrenamiento no suelen tomar tanto tiempo unos cuantos minutos por lo general, dependiendo también de las variables mencionadas anteriormente.

Cómo trabajos futuros se plantea la aplicación de esta metodología con miras a ser empleada en la calibración de sensores de gases, permitiendo que a medida que pasa el tiempo se repitan algunos experimentos en condiciones controladas y en los cuales se pueda evidenciar a través de las respuestas de los sensores virtuales que tanta desviación existe en las respuestas, en caso de que existiere una desviación significativa se vería afectada repetitibilidad de las medidas y con base en ello se podrían calcular multiplicadores u operadores con el fin ajustar la repuesta de los sensores y prolongar la vida útil de los sensores. Adicionalmente existe la aplicabilidad que se puede aprovechar de este tipo de sensores virtuales en aplicaciones de olfato electrónico portables o móviles que requieran menor número de sensores, optimización del consumo de potencia y de espacio, entre otras.

Otro trabajo futuro que se propone, es la exploración de otros métodos de aprendizaje de máquina como el presentado en (Zhang & Liu, 2013), métodos adaptativos y que tengan en cuenta la presencia de derivas en las medidas, sin dejar de lado la generalización y la obtención de un modelo de sensor virtual confiable.

REFERENCIAS

Álvarez González, F. (14 de Marzo de 2008). *Universidad de Cádiz.* Recuperado el 27 de Noviembre de 2012, de REGRESIÓN Y CORRELACIÓN Métodos Estadísticos Aplicados a las Auditorías Sociolaborales: http://www.uca.es/uca/dpto/C146/pag_personal/f_alvarez/documentos/CC%20Trabajo%20Tema%202.pdf

Álvarez, D. A., Fetecua, J. G., Orozco, Á. Á., & Castellanos, C. G. (2010). Caracterización de unidades de acción facial combinando métodos kernel y análisis de componentes independientes. *Revista Facultad de Ingeniería Universidad de Antioquia*.

Arriaza Gómez, A. J., Fernández Palacín, F., López Sánchez, M. A., Muñoz Márquez, M., Pérez Plaza, S., & Sánchez Navas, A. (2008). *Estadística Básica con R y R-Commander.* Cádiz, España: Servicio de Publicaciones de la Universidad de Cádiz.

Ben-Hur , A., & Weston, J. (2010). A User's Guide to Support Vector Machines. En *Data Mining Techniques for the Life Sciences* (págs. 223-239). Humana Press.

Berna, A. (2010). Metal Oxide Sensors for Electronic Noses and Their Application to Food Analysis. *Sensors, Vol.10*, 3882-3910.

Berrueta, L. A., Alonso-Salces, R. M., & Héberger , K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, 196–214.

Bógalo, J., & Quilis, E. M. (2003). *Estimación del ciclo económico mediante filtros de Butterworth.* España: Instituto Nacional de Estadística.

Borovikov, E. A. (13 de Marzo de 1999). An Evaluation of Support Vector Machines as a Pattern Recognition Tool. Maryland, U.S.A.: University of Maryland at College Park.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, Volume 2, Issue 2*, 121-167.

Cea D´Ancona, M. Á. (11 de Febrero de 2006). *INTRODUCCIÓN AL ANÁLISIS DE DATOS ANÁLISIS MULTIVARIANTE, Universidad Complutense de Madrid.* Recuperado el 10 de Noviembre de 2012, de Capítulo 18. Análisis de Regresión Lineal: El procedimiento de Regresión Lineal.: http://www.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/18reglin_SPSS.pdf

Chen , A.-j., Song , Z.-h., & Li, P. (2005). Soft Sensor Modeling Based on DICA-SVR. *International Conference on Intelligent Computing, ICIC 2005, Proceedings, Part I* (págs. 868-877). Hefei, China: Springer Berlin Heidelberg.

Chen, X., Mingfu, C., Yan, H., Yi, L., & Ping, W. (2005). A Non-invasive Detection of Lung Cancer Combined Virtual Gas Sensors Array with Imaging Recognition Technique.

Engineering in Medicine and Biology 27th Annual Conference. (págs. 5873-5876). Shanghai, China.: IEEE.

Devore, J. L. (2008). *Probabilidad y estadística para ingeniería y ciencias.* México D.F.: Cengage Learning Editores S.A.

Durán Acevedo, C. M. (2005). Capítulo 4. Aumento en la selectividad mediante modulación de flujo. En C. M. Durán Acevedo, *Diseño y optimización de los subsistemas de un sistema de olfato electrónico para aplicaciones agroalimentarias e industriales* (pág. 206). Tarragona, España: Universitat Rovira I Virgili.

Durán, C., & Baldovino, D. (2009). Monitoring System to Detect the Maturity of Agro-industrial Products Through of an Electronic Nose. *Revista Colombiana de Tecnologías de Avanzada. Vol.1, No.13.*, 1-8.

Espejo Miranda, I., Fernández Palacín, F., López Sánchez, M. A., Muñoz Máquez, M., Rodríguez Chía, A. M., Sánchez Navas, A., & Valero Franco, C. (2007). La Inferencia Estadística. En I. Espejo Miranda, F. Fernández Palacín, M. A. López Sánchez, M. Muñoz Máquez, A. M. Rodríguez Chía, A. Sánchez Navas, & C. Valero Franco, *Inferencia Estad´ıstica* (págs. 1-10). Cádiz (España): Servicio de Publicaciones de la Universidad de C´adiz.

Fossi, C. S., & D'Ambrosio, A. C. (2004). Uso de las máquinas de vectores de soporte para la estimación del potencial de acción celular. *Revista de Ingeniería UC*, 56-61.

Freund, J. E., Miller, I., & Miller, M. (2000). *Estadística matemática con aplicaciones.* Naucalpán de Juarez (México): Prentice Hall.

Fundación Wikimedia, Inc. (5 de Noviembre de 2012). *Wikipedia - La Enciclopedia Libre*. Obtenido de Estadística Inferencial: http://es.wikipedia.org/wiki/Estad%C3%ADstica_inferencial

GHASEMI-VARNAMKHASTI, M., MOHTASEBI, S. S., RAZAVI, S. H., SIADAT, M., AHMADI, H., & DICKO, A. (2012). Discriminatory Power Assessment of the Sensor Array of an Electronic Nose System for the Detection of Non Alcoholic Beer Aging. *Czech J. Food Sci.*, 236–240.

Gómez Morales, A., & Hernández, G. (2009). Utilización de las máquinas con vectores de soporte para regresión: m2 de construcción en Bogotá. *Revista Avances en Sistemas e Informática*, 21-28.

Gómez Pérez, G. (2004). *Compresión de imágenes mediante SVM adaptativa perceptual.* Valéncia, España: Universidad de Valéncia.

Gonzalez-Jimenez, J., Monroy, J. G., & Blanco, J. L. (2011). The Multi-Chamber Electronic Nose-An Improved Olfaction Sensor for Mobile Robotics. *Sensors, 11*, 6145-6164.

Grupo E-Nose. (15 de Marzo de 2011). Obtenido de ¿Qué es una Nariz Electrónica?: http://www.e-nose.com.ar/paginas/funcionamiento.htm

Gu, D., & Wang, Z. (2008). Distributed Regression over Sensor Networks: An Support Vector Machine Approach . *International Conference on Intelligent Robots and Systems* (págs. 22-26). Nice, France : IEEE.

Gualdrón G., O. E., Durán, C. M., Isaza, C. V., Carvajal F., A., & Uribe, C. (2011). Sistema de olfato electrónico de bajo costo para la detección de diferentes compuestos químicos contaminantes. *Revista Colombiana de tecnologías de avanzada*, 121-126.

Gualdron Guerrero, O. E. (2006). *Desarrollo de diferentes métodos de selección de variables para sistemas multisensoriales.* Tarragona, España: Universitat Rovira I Virgili.

Guiñón, J. L., Ortega, E., García-Antón, J., & Pérez-Herranz, V. (2007). Implementación y análisis del filtro de media móvil. *Filtrado de señales (I)*, 220-227.

Gunn, S. (1998). *Technical Report Support Vector Machines for Classification and Regression.* Southampton: UNIVERSITY OF SOUTHAMPTON.

Gutierrez-Osuna, R. (2002). Pattern analysis for machine olfaction: a review. *Sensors Journal, IEEE*, 189-202.

Ibargüengoytia , P. H., & Reyes, A. (2006). Constructing Virtual Sensors Using Probabilistic Reasoning. *5th Mexican International Conference on Artificial Intelligence* (págs. 218-226). Apizaco, Mexico: Springer Berlin Heidelberg.

Kadleca , P., Gabrys , B., & Strandtb, S. (2009). Data-driven Soft Sensors in the process industry. *Computers and Chemical Engineering*, 795–814.

Kaneko, H., Arakawa, M., & Funatsu, K. (2009). Development of a New Soft Sensor Method Using Independent Component Analysis and Partial Least Squares. *AIChE Journal*, 87-98.

Kato, Y., & Mukai, T. (2007). A real-time intelligent gas sensor system using a nonlinear dynamic response dynamic response. *Sensors and Actuators B: Chemical*, 514–520.

Kecman, V. (2001). *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models. .* Boston - Massachusetts (USA): MIT press.

Koltchinskii, V. (2009). Sparsity in penalized empirical risk minimization. En A. d. Poincaré, *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* (págs. 7–57). Francia: Association des Publications de l'Institut Henri Poincaré.

Krutzler, C., Unger, A., Marhol, H., Fricke, T., Conrad, T., & Schütze, A. (2012). Influence of MOS Gas-Sensor Production Tolerances on Pattern Recognition Techniques in Electronic Noses. *IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT*, 276-283.

Liu , L., Kuo, S., & Zhou, M. (2009). Virtual Sensing Techniques and Their Applications. *IEEE International Conference on Networking, Sensing and Control* (págs. 31-36). Okayama, Japan: IEEE.

Liu , Y., Gao , Z., & Chen, J. (2013). Development of Soft-Sensors for Online Quality Prediction of Sequential-Reactor-Multi-Grade Industrial Processes. En *Chemical Engineering Science.* ELSEVIER.

Macchiavelli, R. E. (26 de Abril de 2010). Modelos no lineales. 1-8. Mayagüez (Universidad de Puerto Rico), Puerto Rico.

Mielle, P., & Marquis, F. (2001). One-sensor electronic olfactometer for rapid sorting of fresh fruit juices. *Sensors and Actuators B*, 470-476.

Mielle, P., Marquis, F., & Latrasse, C. (2000). Electronic noses: specify or disappear . *Sensors and Actuators B, 69*, 287-294.

Minnaard, C. (2010). Modelos de regresión lineales y no lineales: su aplicación en problemas de ingeniería. *Segundo Congreso Argentino de Ingeniería Mecánica IICaim.* San Juan (Argentina): Sociedad Argentina de Educaciòn Matematica.

Pérez Jordán, J. R. (2011). *Reconstrucción de cavidades cardíacas en estudios de navegación electrofisiológica.* Madrid, España: Universidad Rey Juan Carlos.

Peris, M., & Escuder-Gilabert, L. (2009). A 21st century technique for food control: Electronic noses. *Analytica Chimica Acta*, 1–15.

Rodriguez Gamboa, J. C., & Durán Acevedo, C. M. (2008). Sistema de olfato electrónico para la detección de compuestos volátiles. *Revista Colombiana de tecnologías de avanzada*, 20-26.

Rodríguez, J., Durán, C., & Reyes, A. (2010). Electronic Nose for Quality Control of Colombian Coffee through the Detection of Defects in "Cup Tests". *Sensors*, 36-46.

Rodríguez-Gamboa, J. C., Albarracín-Estrada, E. S., & Delgado-Trejos, E. (2011). Quality Control Through Electronic Nose System. En E. b. Eldin, *Modern Approaches To Quality Control* (págs. 505-522). Rijeka, Croatia: Intech Europe.

Romain, A.-C., Delva, J., & Nicolas, J. (2008). Complementary approaches to measure environmental odours emitted by landfill areas . *Sensors and Actuators B: Chemical*, 18–23.

Runu , B., Bipan , T., Laxmi , S., Arun , J., Nabarun , B., & Rajib , B. (2012). Instrumental testing of tea by combining the responses of electronic nose and tongue. *Journal of Food Engineering*, 356–363.

Sánchez, L. G., Osorio, G. A., & Suárez, J. F. (2008). Introducción a kernel ACP y otros métodos espectrales aplicados al aprendizaje no supervisado. *Revista Colombiana de Estadística*, 19-40.

Schlesinger, M. I. (2012). *Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Cybernetics - Center for Machine Perception.* Obtenido de Vapnik-Chervonenkis learning theory: http://cmp.felk.cvut.cz/~hlavac/Pub/KAUST/lectures/27VapnikChervonenkis.pdf

Schütze, A., Gramm, A., & Rühl, T. (2004). Identification of Organic Solvents by a Virtual Multisensor System With Hierarchical Classification . *IEEE Sensors Journal, Vol. 4, Nº. 6*, 857-863.

Tian, F., Yang, S., & Dong, K. (2005). Circuit and Noise Analysis of Odorant Gas Sensors in an E-Nose. *Sensors, Vol.5*, 85-96.

Ulivieri, N., Distante, C., Luca, T., Rocchi, S., & Siciliano, P. (2006). IEEE1451.4: A way to standardize gas sensor. *Sensors and Actuators B: Chemical*, 141–151.

Universidad Complutense de Madrid. (11 de Febrero de 2006). *INTRODUCCIÓN AL ANÁLISIS DE DATOS ANÁLISIS MULTIVARIANTE, Universidad Complutense de Madrid.* Recuperado el 10 de Noviembre de 2012, de Capítulo 18. Análisis de Regresión Lineal: El procedimiento de Regresión Lineal.: http://www.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/18reglin_SPSS.pdf

Valle Padilla, F. (2010). *Implementación eficiente de clasificadores PRIOR-SVM para Matlab.* Madrid, España: Universidad Carlos III de Madrid.

Veganzones, M. A. (2009). *Grupo de Inteligencia Computacional de la Universidad del País Vasco (UPV/EHU).* Obtenido de Support Vector Machines: www.ehu.es/ccwintco

Vergara, A., Vembua, S., Ayhan, T., Ryan, M. A., Homer, M. L., & Huerta, R. (2012). Chemical gas sensor drift compensation using classifier ensembles. *Sensors and Actuators B: Chemical*, 320–329.

Vila, A., Sedano, M., López, A., & Juan, Á. A. (2006). *Análisis de regresión y correlación lineal (Proyecto e-Math).* Cataluña (España): Universidad Abierta de Cataluña.

W. Duin, R. P., & Pękalska, E. (2012). The dissimilarity space: Bridging structural and statistical pattern recognition. *Pattern Recognition Letters*, 826–832.

Wilson, A., & Baietto, M. (2009). Applications and Advances in Electronic-Nose Technologies. *Sensors, Vol. 9*, 5099-5148.

Zhang, M., & Liu, X. (2013). A soft sensor based on adaptive fuzzy neural network and support vector regression for industrial melt index prediction. *Chemometrics and Intelligent Laboratory Systems*, 83–90.

Zhou, H., Homer, M., Shevade, A., & Ryan, M. (2005). Nonlinear Least-Squares Based Method for Identifying and Quantifying Single and Mixed Contaminants in Air with an Electronic Nose. *Sensors, Vol. 6*, 1-18.

APÉNDICE A
CÓDIGOS DE MATLAB

- ✓ A continuación se encuentran los códigos escritos en `Matlab` para los dos métodos probados de escalado:

```
%Media Centrada
[m,n]=size(x);
mx=mean(x);
mcx=(x-mx(ones(m,1),:));

%Autoescalado
[m,n]=size(x);
mx=mean(x);
stdx=std(x);
ax=(x-mx(ones(m,1),:))./stdx(ones(m,1),:);
```

En los códigos anteriores *x* corresponde a la matriz de datos, *mx* es la media de los datos, *stdx* es la desviación estándar, *mcx* son los datos centrados y *ax* los datos auto-escalados.

- ✓ Las siguientes líneas de código muestran la implementación para el filtro Butterworth

```
%Para una frecuencia de paso fp=0.1Hz y una frecuencia de corte fs=1Hz
%Wp=2*π*fp  y  Ws=2*π*fs
%Se tiene entonces un Filtro Butterworth con Wp=0.628 rad/s
%y Ws=6.28 rad/s
%Estas frecuencias se normalizaron dividiendo por la mayor resultando
%un Wp=0.1  y Ws=1 normalizados los otros 2 términos corresponden a la
%atenuación en decibeles
[n,Wn]=buttord(0.1, 1, -1, -80, 's') %Da el orden y el Wn del filtro
[NUM,DEN] = butter(n,Wn,'low'); %Da el numerador y denominador de G(w)
MATRIXFBT1 = filter(NUM,DEN,MATRIX); %Aplica el filtro
```

APÉNDICE B
ALGUNAS TABLAS DE VARIACIÓN DE LOS PARÁMETROS EN EL PROCESO DE SINTONIZACIÓN DE LA SVR

Kernel	Parámetro del Kernel	C	Épsilon
0	1	100,000	0,100
0	1	100,000	0,120
0	1	98,000	0,120
0	1	98,000	0,110
0	1	99,000	0,110
0	1	99,000	0,100
0	1	100,000	0,100
0	1	100,000	0,090
0	1	101,000	0,090
0	1	101,000	0,080
0	1	102,000	0,080
0	1	102,000	0,070

Tabla B1. Parámetros de la SVR para el kernel lineal en uno de los procesos de sintonización de parámetros

Promedio Error Absoluto	Error cuadrático Medio	Promedio error	Error mínimo	Error máximo
0,013	0,042	0,001	0,000	0,037
0,023	0,036	0,001	0,000	0,034
0,025	0,036	0,001	0,000	0,034
0,016	0,041	0,001	0,000	0,037
0,016	0,044	0,001	0,000	0,038
0,011	0,042	0,001	0,000	0,038
0,013	0,042	0,001	0,000	0,037
0,009	0,046	0,001	0,000	0,040
0,007	0,045	0,001	0,000	0,040
0,001	0,048	0,001	0,000	0,042
0,000	0,049	0,001	0,000	0,042
-0,007	0,051	0,001	0,000	0,044

Tabla B2. Errores de entrenamiento de la SVR para el kernel lineal en uno de los procesos de sintonización de parámetros

Promedio Error Absoluto	Error cuadrático Medio	Promedio error	Error mínimo	Error máximo
0,013	0,032	0,002	0,000	0,018
0,023	0,032	0,002	0,000	0,016
0,025	0,032	0,002	0,000	0,017
0,016	0,033	0,002	0,000	0,018
0,016	0,033	0,002	0,000	0,018
0,011	0,032	0,002	0,000	0,018
0,013	0,032	0,002	0,000	0,018
0,009	0,034	0,002	0,000	0,020
0,007	0,034	0,002	0,000	0,019
0,001	0,034	0,002	0,000	0,020
0,000	0,033	0,002	0,000	0,020
-0,007	0,034	0,002	0,000	0,021

Tabla B3. Errores de validación de la SVR para el kernel lineal en uno de los procesos de sintonización de parámetros

Kernel	Parámetro del Kernel	C	Épsilon
2	0	49,000	0,030
2	0	49,000	0,050
2	0	47,000	0,050
2	0	47,000	0,040
2	0	45,000	0,040
2	0	45,000	0,030
2	0	46,000	0,030
2	0	46,000	0,020
2	0	44,000	0,020
2	0	44,000	0,040
2	0	45,000	0,040
2	0	45,000	0,030
2	0	46,000	0,030
2	0	46,000	0,020
2	0	47,000	0,020

Tabla B4. Parámetros de la SVR para el kernel gaussiano en uno de los procesos de sintonización de parámetros

Promedio Error Absoluto	Error cuadrático Medio	Promedio error	Error mínimo	Error máximo
0,013	0,000	0,000	0,000	0,001
-0,007	0,000	0,000	0,000	0,001
-0,011	0,000	0,000	0,000	0,001
-0,001	0,000	0,000	0,000	0,001
-0,001	0,000	0,000	0,000	0,001
0,012	0,000	0,000	0,000	0,001
0,012	0,000	0,000	0,000	0,001
0,004	0,000	0,000	0,000	0,001
0,004	0,000	0,000	0,000	0,001
0,000	0,000	0,000	0,000	0,001
-0,001	0,000	0,000	0,000	0,001
0,012	0,000	0,000	0,000	0,001
0,012	0,000	0,000	0,000	0,001
0,004	0,000	0,000	0,000	0,001
0,004	0,000	0,000	0,000	0,001

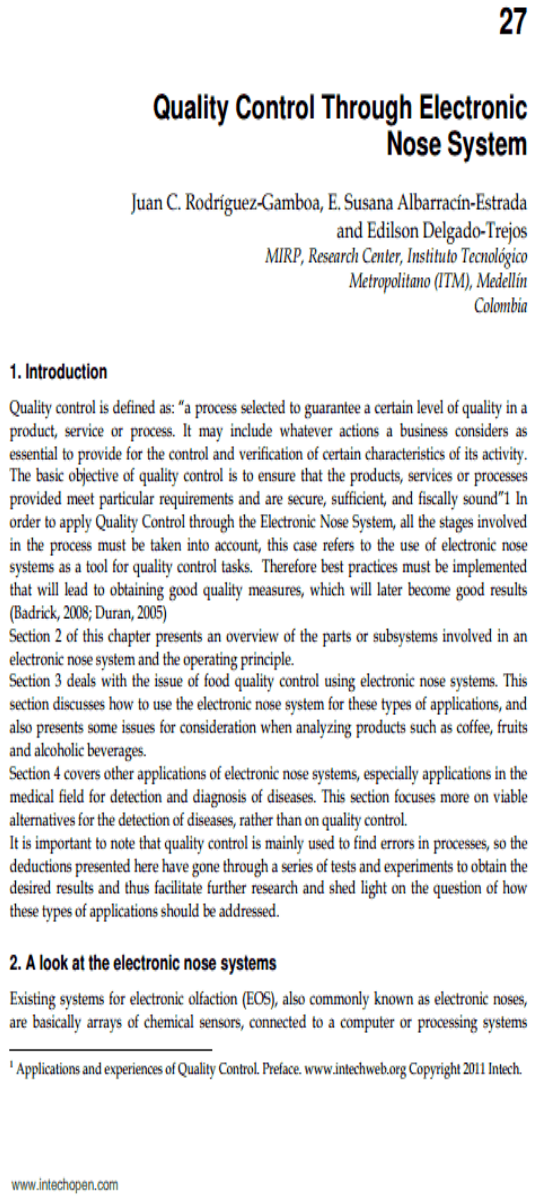
Tabla B5. Errores de entrenamiento de la SVR para el kernel gaussiano en uno de los procesos de sintonización de parámetros

Promedio Error Absoluto	Error cuadrático Medio	Promedio error	Error mínimo	Error máximo
0,009	0,000	0,000	0,000	0,001
-0,009	0,000	0,000	0,000	0,002
-0,013	0,000	0,000	0,000	0,001
-0,004	0,000	0,000	0,000	0,001
-0,004	0,000	0,000	0,000	0,001
0,008	0,000	0,000	0,000	0,001
0,008	0,000	0,000	0,000	0,001
0,000	0,000	0,000	0,000	0,001
0,000	0,000	0,000	0,000	0,001
-0,003	0,000	0,000	0,000	0,001
-0,004	0,000	0,000	0,000	0,001
0,008	0,000	0,000	0,000	0,001
0,008	0,000	0,000	0,000	0,001
0,000	0,000	0,000	0,000	0,001
-0,001	0,000	0,000	0,000	0,001

Tabla B6. Errores de validación de la SVR para el kernel gaussiano en uno de los procesos de sintonización de parámetros

APENDICE C
PRODUCTOS DERIVADOS DE ESTE TRABAJO DE GRADO

- ✓ Durante la ejecución de este proyecto de grado se elaboró y publico un capítulo de libro titulado Quality Control Through Electronic Nose System en el libro Modern Approaches To Quality Control con ISBN 978-953- 307-971-4 y DOI: 10.5772/22217.



27

Quality Control Through Electronic Nose System

Juan C. Rodríguez-Gamboa, E. Susana Albarracín-Estrada and Edilson Delgado-Trejos
MIRP, Research Center, Instituto Tecnológico Metropolitano (ITM), Medellín Colombia

1. Introduction

Quality control is defined as: "a process selected to guarantee a certain level of quality in a product, service or process. It may include whatever actions a business considers as essential to provide for the control and verification of certain characteristics of its activity. The basic objective of quality control is to ensure that the products, services or processes provided meet particular requirements and are secure, sufficient, and fiscally sound"1 In order to apply Quality Control through the Electronic Nose System, all the stages involved in the process must be taken into account, this case refers to the use of electronic nose systems as a tool for quality control tasks. Therefore best practices must be implemented that will lead to obtaining good quality measures, which will later become good results (Badrick, 2008; Duran, 2005)

Section 2 of this chapter presents an overview of the parts or subsystems involved in an electronic nose system and the operating principle.

Section 3 deals with the issue of food quality control using electronic nose systems. This section discusses how to use the electronic nose system for these types of applications, and also presents some issues for consideration when analyzing products such as coffee, fruits and alcoholic beverages.

Section 4 covers other applications of electronic nose systems, especially applications in the medical field for detection and diagnosis of diseases. This section focuses more on viable alternatives for the detection of diseases, rather than on quality control.

It is important to note that quality control is mainly used to find errors in processes, so the deductions presented here have gone through a series of tests and experiments to obtain the desired results and thus facilitate further research and shed light on the question of how these types of applications should be addressed.

2. A look at the electronic nose systems

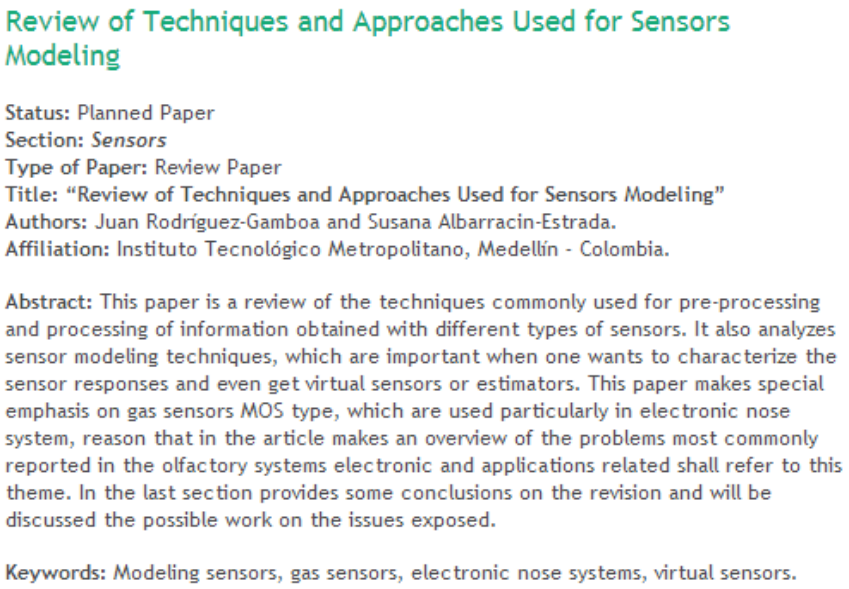
Existing systems for electronic olfaction (EOS), also commonly known as electronic noses, are basically arrays of chemical sensors, connected to a computer or processing systems

[1] Applications and experiences of Quality Control. Preface. www.intechweb.org Copyright 2011 Intech.

www.intechopen.com

Figura C1. Pantallazo de la primera página del capítulo de libro

- ✓ A la fecha, también está en proceso de producción un artículo titulado Review of Techniques and Approaches Used for Sensors Modeling para ser publicado en Science Journal.



Review of Techniques and Approaches Used for Sensors Modeling

Status: Planned Paper
Section: *Sensors*
Type of Paper: Review Paper
Title: "Review of Techniques and Approaches Used for Sensors Modeling"
Authors: Juan Rodríguez-Gamboa and Susana Albarracin-Estrada.
Affiliation: Instituto Tecnológico Metropolitano, Medellín - Colombia.

Abstract: This paper is a review of the techniques commonly used for pre-processing and processing of information obtained with different types of sensors. It also analyzes sensor modeling techniques, which are important when one wants to characterize the sensor responses and even get virtual sensors or estimators. This paper makes special emphasis on gas sensors MOS type, which are used particularly in electronic nose system, reason that in the article makes an overview of the problems most commonly reported in the olfactory systems electronic and applications related shall refer to this theme. In the last section provides some conclusions on the revision and will be discussed the possible work on the issues exposed.

Keywords: Modeling sensors, gas sensors, electronic nose systems, virtual sensors.

Figura C2. Pantallazo tomado del sitio web de Science Journal bajo la url: http://journal.insciences.org/review-of-techniques-and-approaches-used-for-sensors-modeling/

- ✓ Adicionalmente, se espera elaborar dos artículos con los resultados y aportes de este trabajo de grado y publicarlos en revistas indexadas.