



UNIVERSITAT DE
BARCELONA

Análisis dinámico de atmósferas con matrices de sensores mediante procesado de señal no-lineal

Antonio Pardo Martínez



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement 3.0. Espanya de Creative Commons.**

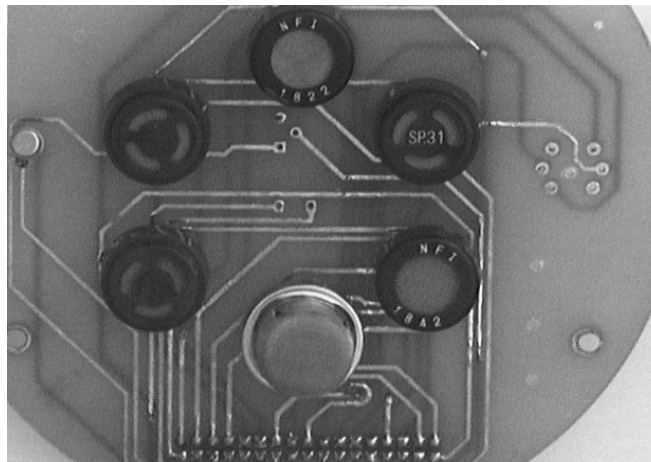
Esta tesis doctoral está sujeta a la licencia **Reconocimiento 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution 3.0. Spain License.**



Universitat de Barcelona
Departament d'Electrònica

**ANÁLISIS DINÁMICO DE ATMÓSFERAS
CON MATRICES DE SENSORES
MEDIANTE PROCESADO DE SEÑAL NO-LINEAL**



Antonio Pardo Martínez

Universitat de Barcelona
Facultat de Física
Departament d'Electrònica

**ANÁLISIS DINÁMICO DE ATMÓSFERAS CON MATRICES DE
SENSORES MEDIANTE PROCESADO DE SEÑAL NO-LINEAL**

Memòria presentada per optar al títol de doctor en Ciències Físiques

Programa de doctorat: Micro i Optoelectrònica Física
Bienni: 91-93

Barcelona, Maig de 2000

Autor: Antonio Pardo Martínez
Director de tesi: Dr. Santiago Marco Colás
Tutor: Dr. Joan Ramón Morante i Leonart

El Doctor Santiago MARCO COLÁS, Professor Titular de la Facultat de Física de la Universitat de Barcelona:

CERTIFICA:

Que la memòria “ANÁLISIS DINÁMICO DE ATMÓSFERAS CON MATRICES DE SENSORES MEDIANTE PROCESADO DE SEÑAL NO-LINEAL”, que presenta el Sr. Antonio Pardo Martínez per a optar al grau de Doctor en Ciències Físiques, s’ha realitzat sota la seva direcció,

Barcelona, 9 de Maig de 2000

Dr. Santiago MARCO COLÁS

Agradecimientos,

Por largo que se haga el texto de los agradecimientos (podría ocupar más que la propia tesis) siempre se corre el riesgo de olvidar a alguien. Vayan por delante entonces mis disculpas si alguien no se ve reflejado en ellos, no es mi intención olvidarme de nadie, pero a estas alturas del proceso, mis neuronas (que no son artificiales) ya no producen sinapsis eficientes.

En primer lugar, agradecimiento sin límites a mi mujer y mi hija que han tenido que soportar fines de semana sin salir y una invasión en casa de libros artículos, disquetes y papelotes. Es absolutamente impagable el esfuerzo que ellas han hecho por mí y sólo espero tener la oportunidad de compensarlas mínimamente. Mis primeras acciones en este sentido serán: tratar de exterminar la plaga de información que se ha extendido por casa y hacer firme propósito de dar el biberón a mi hija sin estar pensando en las últimas modificaciones de la tesis.

También agradezco a mis padres a mi hermano y a mi familia en general el soporte que me han ido dando en todo momento. Sobre todo tendré que agradecer a aquellos miembros de mi familia que asistan a mi defensa de la tesis, el tener que soportar un “rollo” que ni les va ni les viene, y sobre el que piensan (espero que sin razón) que no tiene ninguna utilidad. Aportaciones en este sentido han sido hechas por mi propio hermano al que espero convencer de lo contrario con un buen soborno a base de jamón (¿qué otra manera hay de convencerlo?).

No puedo tampoco dejar de nombrar a todos mis amigos Alberto, Carlos, Carlos, Chema, Pilar, Begoña, Helena. Alguno de ellos ya sabe lo que es esto y me comprende bien, otros están embarcados en ello y me comprenderán bien. En cualquier caso, pedirles que me perdonen si alguna vez he fallado a las citas con ellos por causa de la tesis y decirles que una buena calçotada es la mejor manera de reactivar el tema. En caso de no ser temporada, una parrillada o incluso una biquinada también servirán; el caso es hacer algo relacionado con comida y acabado en “ada”, es decir: a lo grande.

A todos mis compañeros del departamento: catedráticos, titulares, asociados, ayudantes colaboradores, becarios, ..., muchas gracias por vuestro apoyo y soporte. Para mí es un placer y un orgullo poder decir que trabajo en un departamento donde más que compañeros lo que tengo son amigos. Se suele decir que no es justo destacar a nadie en particular y probablemente es cierto, pero tampoco sería justo no nombrar a las personas que dentro del departamento más influencia han tenido para que esta tesis llegue a buen puerto. Entre estas dos aguas, me decanto por la injusticia de particularizar nombres.

En primer lugar agradecer al Dr. Joan Ramon Morante como jefe del Departament d'Electrònica y al Dr. Josep Samitier como responsable del grupo Sistemes d'Instrumentació i Comunicacions (SIC) el hecho de acogerme dentro del departamento y orientarme en el camino de los sensores de gas, que en su día comencé gracias exclusivamente a ellos.

Asimismo, tengo que agradecer también al Dr. Davide y al Dr. Hierlemann sus aportaciones a esta tesis y su generosidad a la hora de compartir sus conocimientos y sus datos experimentales.

Gracias también a mis compañeros de la Escuela Universitaria Salesiana de Sarriá y a Addlink Software científico por la versatilidad de la que he disfrutado trabajando con ellos que me ha permitido, finalmente, acabar esta tesis.

Gracias a mis amigos del grupo de tratamiento de señal, Alex, Teo, Arturo, Martí y Santi. No será posible olvidar aquellas míticas reuniones (que echo de menos) donde discutíamos las modificaciones del sistema: aquellos dilemas existenciales: “Peltier o no Peltier, esa es la cuestión”; aquellas especulaciones esquizoides: “Tenemos ruido en el

sistema, ¿Quién lo introduce?”, aquel, en fin, torrente inacabable de ideas generadas en el tortuoso, poderoso y a veces alevoso magma de los cerebros científicos discutiendo (¡Huy que miedo!). Gracias especialmente a Santi, mi director, por poner a mi disposición su tiempo y su conocimiento y por saber encauzar perfectamente mis esfuerzos.

Gracias también a Pepe, a Oscar y a Lourdes por su amistad y compañerismo, por su buen hacer y por sus interesantes aportaciones a lo largo de estos años. No sólo son buenos científicos y buenos compañeros y amigos, sino que, lo más importante, son buena gente. De hecho una parte de esta tesis la han escrito ellos: concretamente, este párrafo.

Gracias además, a mis compañeros en los equipos de basquet y fútbol sala: Manel, Oscar, Lorenzo, Jose, Jaime, Antonio, Teo, Miguel Angel, Genís, Andreu, (y seguro que me dejo gente) que han tenido que padecer mis fallos a puerta vacía y mis entradas a canasta que no tocan ni aro. A mi favor he de decir que a veces también meto goles imposibles y triples de ocho metros, o sea que una cosa va por la otra. A ver si reiniciamos estas actividades, que el deporte es salud (aunque a nuestra edad la cosa ya no está tan clara) y a ver si, sobretodo, conseguimos atraer más público del departamento (sólo con traer una persona ya generaríamos una ganancia infinita respecto al público de la pasada temporada).

Como resumen, seriamente he de decir que doy las gracias a todos en general porque, a parte de aportaciones científicas o personales, de méritos profesionales o extra-profesionales, lo mejor que puedo decir de todos es que sois buenas personas, con la que es un placer tratar.

Para acabar vuelvo al principio, a mi familia: Ester, Laura: ¡Vuelvo a estar con vosotras!

ÍNDICE

PRÓLOGO	
---------------	--

CAPÍTULO 1

Prólogo	1.2
1.1. La percepción del mundo físico.....	1.4
1.2. Los gases y su análisis: Problemática actual	1.5
1.2.1. Ahorro energético	1.6
1.2.2. Control medio ambiental	1.7
1.2.3. Seguridad	1.10
1.3. Aplicaciones de la determinación de olores y aromas.....	1.14
1.3.1. Cosmética y perfumería.....	1.15
1.3.2. Alimentación	1.16
1.3.3. Medicina	1.17
1.4. Problemática de la medida: Sensores	1.18
1.4.1. Sensores electroquímicos	1.20
1.4.2. Sensores de polímeros	1.22
1.4.3 Sensores SAW y microbalanzas de cuarzo (QMB).....	1.24
1.4.4. Sensores ópticos	1.26
1.4.5. Sensores de óxidos metálicos	1.27
1.4.6. Sensores catalíticos.....	1.31
1.5. Detección natural: La percepción olfativa.....	1.32
1.5.1. Los receptores olfativos.....	1.33
1.5.1.1. Electrofisiología de las células olfatorias	1.34
1.5.2. El bulbo olfatorio y las conexiones con los centros nerviosos	1.35
1.6. Trastornos y disfunciones del sistema olfativo.....	1.37
1.7. Cuantificación de la sensación olfativa: Psicofisiología olfativa	1.38
1.8. Subjetividad en la percepción olfativa	1.40
1.9. Percepción artificial: la nariz electrónica	1.40
1.10. Breve descripción de la fisiología de una neurona	1.41
1.11. La neurona artificial	1.43
1.12. Redes neuronales artificiales: Arquitectura.....	1.46
1.13 Redes neuronales artificiales: entrenamiento	1.47
1.14. La nariz electrónica	1.50
1.15. Limitaciones de las narices electrónicas.....	1.53
1.16. Narices electrónica existentes.....	1.54
1.17 Resumen	1.55
Bibliografía.....	1.57

CAPÍTULO 2

Prólogo	2.2
2.1. Introducción.....	2.3
2.2. Concepción del sistema	2.5
2.2.1. Sobre gases y mezclas	2.6
2.2.2. Sobre sensores	2.6
2.2.3. Sobre acondicionamiento de la señal	2.7
2.3. Sistema experimental.....	2.8

2.4. Gases.....	2.9
2.5 Controladores de flujo másico	2.11
2.6. Líquidos y gases	2.13
2.7. Matriz de sensores	2.15
2.8. Electrónica de acondicionamiento y control	2.16
2.8.1. Interfaz para Mass Flow Controlers	2.16
2.8.2. Interfaz para los sensores.....	2.17
2.9. Estimación de las prestaciones	2.19
2.10. Sistema experimental de la Universidad de Tübingen	2.21
Bibliografía.....	2.23

CAPÍTULO 3

Prólogo	3.2
3.1. Introducción a la teoría de sistemas: Conceptos básicos.....	3.4
3.2. Modelos de sistemas.....	3.8
3.3. El proceso identificativo.....	3.9
3.4. Construcción de Experimentos.....	3.13
3.5. Presentación de resultados.....	3.18
3.6. Caracterización estática de los sensores	3.20
3.7. Resumen de resultados y conclusiones del capítulo.....	3.25
Bibliografía.....	3.27

CAPÍTULO 4

Prólogo	4.2
4.1. Introducción a la descripción matemática de un sistema LTI	4.2
4.2. Estimación de un sistema LTI	4.4
4.3. Validación y análisis de residuos	4.5
4.4. Determinación de no linealidades a partir del análisis de residuos	4.7
4.5. Estimación no paramétrica mediante análisis de correlación.....	4.9
4.6. Aplicación del análisis de correlación a un problema sintético.....	4.10
4.7. Aplicación del análisis de correlación al sistema	4.14
4.7.1. Identificación de un sistema SISO	4.14
4.7.2. Identificación de un sistema MISO	4.22
4.8. Estimación paramétrica de un sistema LTI	4.24
4.9. Estimación paramétrica mediante un modelo ARX	4.27
4.10. Criterios para la estimación de los parámetros.....	4.28
4.11 Determinación del orden del modelo: métodos de regularización	4.29
4.12. Aplicación de la estimación paramétrica a un problema sintético	4.31
4.12.1. Modelo FIR	4.31
4.12.2. Modelo ARX	4.31
4.13. Aplicación de la estimación paramétrica al sistema.....	4.34
4.13.1. Identificación SISO: Estimación con estructura FIR	4.34
4.13.2. Identificación SISO: Modelo paramétrico con estructura ARX.....	4.35
4.13.3. Estimación MISO: Estructura ARX	4.41
4.14. Elección entre método paramétrico y no paramétrico	4.45
4.15. Resumen de los resultados y conclusiones	4.47
Bibliografía.....	4.48

CAPÍTULO 5

Prólogo	5.2
5.1. Desarrollo de un sistema en series funcionales	5.2
5.2. Descripción de la serie: Kernels de Volterra	5.3
5.3. De los kernels de Volterra a los kernels de Wiener	5.5
5.4. Estimación de kernels mediante análisis de correlación (Lee-Schetzen)	5.6
5.5. Aplicación de la estimación Lee-Schetzen a un problema sintético	5.7
5.6. Aplicación de la estimación Lee-Schetzen al sistema	5.13
5.6.1 Identificación de un sistema SISO	5.13
5.6.2. Estimación de un modelo MISO mediante el método Lee-Schetzen	5.15
5.7. Estimación mediante modelado por cajas	5.17
5.8. Aplicación del modelado de cajas a un problema sintético	5.19
5.9. Aplicación del modelado por cajas al sistema	5.24
5.10. Estimación mediante un método paramétrico no lineal: NARX	5.26
5.11. Aplicación del método NARX a un problema sintético	5.27
5.12. Aplicación del método NARX al sistema	5.29
5.13. Estimación mediante redes neuronales	5.32
5.14. Aplicación de las redes neuronales a un problema sintético	5.33
5.15. Aplicación de las redes neuronales a la estimación del sistema	5.35
5.16. Resumen de los resultados y conclusiones	5.37
Bibliografía	5.39

CAPÍTULO 6

Prólogo	6.2
6.1. El problema inverso	6.2
6.2. Aplicación del modelado inverso al sistema real: Experimentos	6.6
6.3. Aplicación del modelado inverso al sistema real: Determinación de los modelos	6.7
6.4. Aplicación del modelado inverso al sistema real: modelos no lineales	6.10
6.4.1. Estimación inversa mediante el método NARX	6.10
6.4.2. Estimación inversa mediante redes neuronales	6.14
6.5. Resumen de resultados y conclusiones	6.22
Bibliografía	6.23

Principales aportaciones y conclusiones

Resumen de artículos y comunicaciones a congresos

“If you are looking for a new science, measure a smell” Alexander Graham Bell

Prólogo

Reconocer un olor o un aroma es una tarea natural que ni siquiera es específica de especies altamente desarrolladas. Mamíferos, peces, aves o insectos son capaces de una u otra forma de reconocer ciertos olores y utilizan esta habilidad en su propio beneficio. A pesar de ser ésta una tarea muy común en la naturaleza, tratar de imitarla de forma artificial no es sencillo. Añádase ahora a la tarea del reconocimiento la tarea de cuantificación del olor y aparecerá un problema complejo que, sin embargo, presenta interesantes y numerosas aplicaciones prácticas.

Para poder cuantificar un olor es necesario disponer de elementos sensibles bien calibrados con los cuales desarrollar un sistema de medida. Sería deseable, a su vez, que este sistema de medida fuese lineal, fiable, con buena resolución y exactitud, a ser posible portátil, y además barato. Desgraciadamente, este tipo de sistema perfecto no existe. Sí existen grandes aparatos, como espectrógrafos o cromatógrafos, capaces de realizar tareas de medición con muy buena resolución. Pero aún y funcionando relativamente bien, distan mucho de ser ni portátiles ni baratos, y pueden resultar inadecuados para determinadas tareas de reconocimiento de olores. Se trata entonces de abordar el problema desde otra perspectiva, con un sistema relativamente económico, conformado esencialmente por una matriz de sensores para realizar tareas tanto de medición cuantitativa como de reconocimiento.

Los problemas habituales en un sistema de medida así son la no linealidad de la respuesta de los sensores, su baja selectividad, la dinámica de la respuesta del sistema y las derivas en el tiempo. Estos y otros posibles problemas pueden ser aliviados, en buena parte, a través de procesado de señal aplicado a las respuestas de la matriz de sensores de gases. Esta idea no es nueva; de hecho la esencia de lo que hoy en día se entiende por **nariz electrónica**, es en definitiva una matriz de sensores acompañada de un procesado de señal inteligente^(ver figura). Actualmente existe un reducido número de pequeñas empresas en todo el mundo tratando de explotar esta idea comercialmente. Pero para poder diseñar estos algoritmos de procesado inteligente es esencial disponer de un modelo de comportamiento del sistema de medida entendido no sólo como un cierto número de sensores agrupados en una matriz, sino como la combinación de los sensores más la cámara donde se alojan más el acondicionamiento de la señal de respuesta de todo el conjunto.

El estudio y modelización de un sistema de medida en general es un trabajo que puede abordarse desde muy diversas perspectivas, que precisa de conocimientos que competen a múltiples disciplinas científicas y de ingeniería y que, por tanto, exige el manejo de una relativamente amplia variedad de métodos matemáticos. De todas las diversas opciones, debe quedar claro que la perspectiva en la cual se está interesado en este trabajo es el estudio y modelización en base a las señales entrada-salida del sistema de medida; no es el interés de este trabajo modelizar la química del problema ni establecer fórmulas o ecuaciones que describan las relaciones físico-químicas del mismo.

Hasta este punto se ha evitado conscientemente el término “identificación de sistemas” y se ha estado hablando o se hablará más adelante de ‘desarrollar caracterizaciones’ o ‘generar modelos’ del sistema a partir de las señales de entrada y salida del mismo. Precisamente, la identificación de sistemas, término acuñado por Lafti Zadeh (considerado el padre de la lógica difusa) en 1962, trata estrictamente de la determinación del comportamiento de un sistema en base a sus entradas y sus salidas

(“Identification is the determination, on basis on input and output, of a system within in a specified class of systems, to which the system under test is equivalent” L. Zadeh (1962) From circuit theory to system theory. Proc. IRE 50 856-865). Es decir, a partir de este punto se asume pues el término de identificación como sinónimo de modelización en base a las entradas y salidas del sistema.

Establecido este enfoque y hablando de las herramientas matemáticas útiles para tratarlo, debe decirse que una gran mayoría de las mismas se basan en realizar aproximaciones lineales al problema. Son herramientas ampliamente utilizadas por diversas disciplinas y por tanto, están muy bien estudiadas, son de solvencia contrastada y están ampliamente difundidas. En este sentido, la base de la metodología de estudio de un sistema de medida, siempre que se realice desde un punto de vista estático y lineal, es bien conocida. Sin embargo, ya ha quedado expuesto que, en general, un sistema multientrada-multisalida basado en una matriz de sensores de gas no se ajusta bien a una caracterización estática y lineal y, de hecho lo habitual es que presente una cierta dinámica y un cierto grado de no linealidad.

Desarrollar métodos de modelización dinámicos supone un mayor grado de complejidad en el estudio, aunque suelen ser comunes en el ámbito de la teoría de control. Son más escasos, sin embargo, los trabajos donde en general se utiliza la modelización dinámica no lineal. En el ámbito concreto de la teoría de medida, existe un interés claro en la modelización dinámica de sistemas, pero los principales empeños siguen estando concentrados en la base estadística de las aproximaciones lineales. Aún dentro del campo de la teoría de medida y para el caso particular de los sistemas de medida basados en matrices de sensores de gases, los mayores esfuerzos puestos en juego hasta ahora se han dedicado a la modelización estática no lineal, pero la traslación hacia los modelos dinámicos no lineales es inexistente, siendo éste el primer empeño en este sentido y una de las principales aportaciones de este trabajo.

Esta línea de investigación dinámica no-lineal sobre sistemas de medida de gases y todo el procesado de señal que comporta es nueva en el departamento de electrónica y, de hecho, para poder desarrollarla en toda su potencia resulta imprescindible disponer de un sistema dedicado a la generación de atmósferas gaseosas bien controladas. Dada por tanto la inexistencia de la estructura inicial necesaria, partiendo de cero se ha debido desarrollar todo un complejo sistema de generación de atmósferas gaseosas y de test de sensores. A partir de él y de sistemas de similares será posible disponer de señales para la modelización de sistemas de análisis basados en matrices de sensores de gases.

El esfuerzo invertido en el desarrollo de los sistemas y en la aplicación de la teoría dinámica no lineal se verá recompensado con la obtención de modelos de alto rendimiento. Sin embargo, antes de llegar a ellos, habrá que pasar por el desarrollo de estudios estáticos y lineales del propio sistema. Pero aún queda por establecer una matización más; las relaciones dinámicas no lineales entre las variables del sistema pueden construirse en sentido directo: desde la entrada hacia la salida, o en sentido inverso: desde la salida hacia la entrada. Es decir, en un modelo directo se trata de predecir la salida del sistema en función de la composición de la atmósfera de entrada mientras que en un modelo inverso el interés se centra en conocer la composición de la atmósfera en la entrada en función de la salida del sistema.

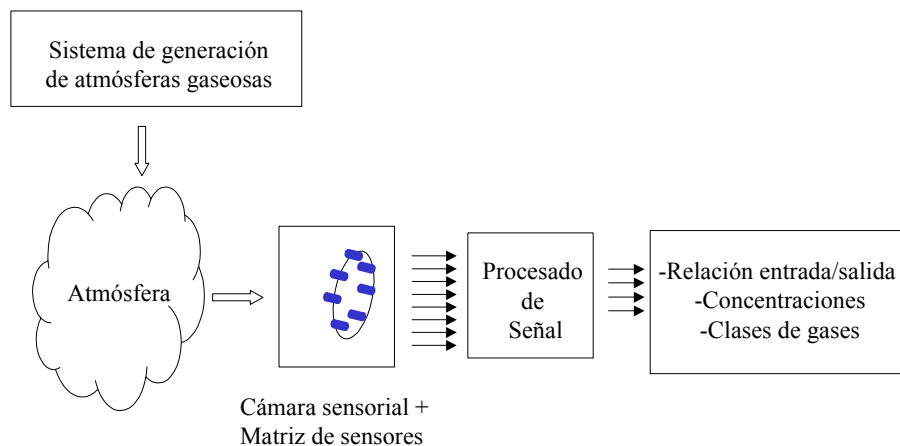
La aplicabilidad de una estimación inversa de este estilo es inmediata; si un sistema sensor es excitado por una mezcla gaseosa, bajo ciertas condiciones ha de ser posible, a partir de la señal de respuesta, establecer qué concentración de cada gas hay en la mezcla. Éste es de hecho el objetivo primordial de todo sistema de medida. En este trabajo se verá cómo el procesado de señal es capaz de modelizar un sistema inverso de

este estilo. Si se pretende que el modelo sea capaz de ofrecer resultados correctos ante cambios súbitos de la entrada será necesario también trabajar con modelos inversos dinámicos.

El interés de la identificación directa e inversa en el marco de los analizadores de gases/narices electrónica se centra en la obtención de controles óptimos (control de sistemas), simulación de respuestas del sistema (descripción del sistema), predicción de respuestas futuras (predicción) o estimación de la entrada a partir de la salida (medición). En la introducción del capítulo 3 se desarrollan con más detalle estas ideas.

Aparte de la aproximación cuantitativa, existe otro tipo de aproximación al problema: la aproximación cualitativa. Ésta, no se pregunta por el “cuánto hay”, sino por el “qué es” y no se está interesado en el valor x de las concentraciones del gas G ni del valor concreto de la respuesta y del sistema, sino en distinguir, a partir de la respuesta y si el gas es efectivamente G o es H . De alguna manera se trata de imitar el comportamiento del olfato, con el cual podemos reconocer diferentes olores, aunque no podemos precisar su concentración. Este aspecto cualitativo no se va a desarrollar en el presente estudio, dejando el tema para su completa explotación en posteriores trabajos que se presentarán a corto plazo.

Así, dentro del entorno del procesado de señal aplicado a las narices electrónicas, el resumen de las contribuciones del presente trabajo se centra en dos puntos clave: la generación de modelos dinámicos no lineales que describan el comportamiento del sistema ante entradas arbitrarias y la generación de modelos inversos, capaces de recuperar las concentraciones a partir de las señales de los sensores.



Establecidos concisamente los puntos importantes del trabajo, los diferentes capítulos del mismo tratan de dar respuesta a todos los propósitos expuestos.

El capítulo 1 está dedicado a realizar una amplia introducción al problema de la medición gaseosa desde el punto de vista de gases en general y particularizando brevemente el problema de los aromas. Se comentará brevemente el funcionamiento del sistema olfativo humano, y sus analogías con la filosofía de las narices electrónicas.

El capítulo 2 se dedica enteramente al desarrollo del equipo experimental. El desarrollo de dicho sistema ha sido uno de los objetivos que se perseguían con la realización de este trabajo. Cabe decir ahora que el sistema desarrollado es un sistema activo en la toma de muestras, en el sentido que se generan muestras gaseosas que se hacen llegar directamente a la cámara sensorial. Un sistema pasivo como una alarma, realiza la detección a partir de la señal que llega a él gracias a la difusión del gas en la atmósfera, pero no contiene ningún elemento que fuerce dicha llegada. El sistema experimental

desarrollado y las narices electrónicas son elementos activos en la toma de muestras aunque, la diferencia entre ambos sistemas está radicada en el hecho de que la muestra en el sistema experimental está controlada por un experimento y/o un experimentador. Sin embargo, la nariz electrónica tomando muestras activamente (a través de una bomba de succión, por ejemplo), lo hace a partir de la atmósfera que rodea a la propia nariz, pero no tiene porqué ser en virtud de un experimento concreto. Esta es la única diferencia entre los sistemas experimentales como el desarrollado y las narices electrónicas,

El capítulo 3 es de hecho, el capítulo donde se inician las tareas relativas a la identificación de sistemas, y es por tanto donde se inicia el proceso que ha de llevar a la resolución de los objetivos relativos a ello. En él se vuelven a plantear los objetivos dentro del campo de la identificación, se explicita la metodología identificativa a seguir, se presentan los datos experimentales que se utilizarán en el proceso de caracterización e identificación y, finalmente, se inicia dicho proceso desde el punto de vista estático, ofreciéndose así los primeros resultados.

Establecida la metodología y iniciada la identificación desde el punto de vista estático en el capítulo 3, el capítulo 4 continúa la identificación del sistema desde el punto de vista dinámico lineal y, por tanto, contiene tanto la base teórica esencial de los desarrollos que se utilizan como los resultados obtenidos. Se ha procurado no convertir el trabajo en un libro específico de teoría de la identificación de sistemas y es por ello que no se ha introducido más teoría que la necesaria para la comprensión de los resultados, dejando a la consulta bibliográfica la fundamentación estricta de dicha teoría.

El capítulo 5 recoge los resultados de los desarrollos lineales y trata de extenderlos hacia la identificación dinámica no lineal. De nuevo el capítulo debe compaginar de una manera equilibrada la teoría básica con los resultados. Tal compaginación se realiza en el mismo sentido que en el capítulo 4.

El capítulo 6 trata de completar el ciclo identificativo cuantitativo desarrollando la identificación inversa. En cuanto a la descripción teórica, será este capítulo más ligero que los anteriores ya que la mayoría de conceptos ya habrán sido descritos en los capítulos 4 y 5.

En definitiva, de lo que trata este trabajo es de la conjunción de un dispositivo experimental con una serie de herramientas matemáticas, que deben permitir desarrollar un sistema de medida artificial muy potente, en el que el modo de detección cuantitativa quede complementado por la aproximación cualitativa (no presentado en este estudio) y viceversa. Esta conjunción es la esencia de lo que se suele denominar nariz electrónica y es el objetivo a perseguir como continuación de este trabajo. Cabe apuntar, finalmente, que los datos experimentales usados en esta tesis no provienen tan sólo del sistema experimental desarrollado sino que también provienen de experimentos realizados en la Universidad de Tübingen en Alemania. Mi agradecimiento al grupo del profesor Göpel, desgraciadamente fallecido recientemente, por el uso que se hace de sus datos.

Capítulo 1

Hacia la percepción gaseosa inteligente

Ahora ya no olía casi nada, hacía rato que las sustancias volátiles que respiraba le habían aturdido y ni siquiera era capaz de reconocer de nuevo lo que al principio del experimento creía haber analizado sin lugar a dudas. Jamás llegaría a averiguar la composición del nuevo perfume; esta noche, no, desde luego, pero tampoco mañana, cuando con ayuda de Dios su nariz se hubiese recuperado. Nunca había conseguido aprender a utilizar el olfato para este fin. Captar los elementos de un perfume era un trabajo antipático y repugnante para él; no le interesaba dividir una fragancia más o menos buena en las partes que la componían. Lo mejor era dejarlo.

No obstante, su mano continuaba humedeciendo mecánicamente el pañuelo de encaje con delicados movimientos practicados mil veces, agitándolo y pasándolo con rapidez por delante del rostro y, también mecánicamente, inhalando una porción de aire perfumado y expulsándolo en pequeñas cantidades, tal como mandan las reglas.

Patrick Süskind: “El perfume”

-¿Y ahora dónde has puesto las cuatro monedas? - le preguntó el Hada.

-¡Las he perdido! - respondió Pinocho; pero dijo una mentira porque, por el contrario las tenía en el bolsillo.

- Apenas dijo la mentira, su nariz, que ya era larga, le creció de pronto dos dedos más.

Carlo Collodi: “Las aventuras de Pinocho”
(Nariz artificial para la detección de mentiras)

Prólogo del capítulo

En la novela *el perfume*, el protagonista Jean-Baptiste Grenouille, posee una nariz excepcional que le permite identificar a distancia, incluso atravesando paredes, cualquier tipo de olor. Desgraciadamente, este don especial va acompañado de una ausencia de olor corporal y de una personalidad supersticiosa y psicótica que le lleva a una carrera desenfrenada de homicidios y que culmina en un final alucinante. De alguna forma esta ficción novelesca perfectamente ambientada en la Francia del siglo XVIII, durante la época de la aparición la industria perfumista, hace hincapié en algunos aspectos interesantes que atañen a la elaboración de este trabajo. Ciertamente, la capacidad de discernir los olores esenciales de un aroma puede ser de gran utilidad no tan solo para la industria cosmética sino, como pretendo dejar patente en este capítulo, para otros ámbitos, industriales o no, absolutamente diferentes. Por otro lado el olfato, aun siendo un instrumento de detección química altamente evolucionado, no es lo suficientemente fiable para realizar mediciones cuantitativas, amén de otras problemáticas asociadas como insensibilidad a gases en general, influencia de factores fisiológicos o psicológicos, etc. Resultaría por tanto muy útil disponer de un instrumento artificial, capaz de realizar medidas de detección gaseosa cualitativas y/o cuantitativas on-line e inmune a la problemática asociada al olfato natural. Un dispositivo de este tipo con ciertos condicionantes adicionales (multiplicidad de sensores, no selectividad de los mismos,...) suele clasificarse comúnmente bajo un calificativo general de nariz electrónica.

En este primer capítulo se pretende introducir el concepto de nariz electrónica como conclusión de un desarrollo paralelo y diferenciado entre lo artificial y lo natural y así sentar la base de la realización experimental descrita en capítulos posteriores. Para ello el capítulo se estructura de forma general según el siguiente organigrama:

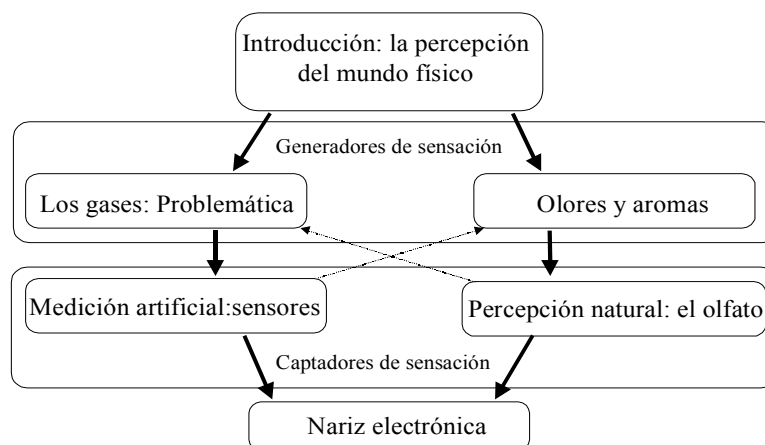


Figura 1. 1 Esquema del capítulo

Se distinguen en el organigrama dos grandes grupos. El primer grupo trata el tema de las sustancias gaseosa o volátiles generadoras de señal perceptible y su problemática actual. Dentro de este grupo, se trata de particularizar dos casos diferentes. El primero se refiere a gases, englobando en este conjunto, todas aquellas sustancias gaseosas que, independientemente de la sensación olfativa que generen, no son una mezcla heterogénea de sustancias sino que están compuestas por moléculas, más o menos complejas, pero puras (CH_4 , CO_2 , CO ,...).

El segundo hace referencia referente a olores y aromas, entendidos estos como mezclas heterogéneas de sustancias, que en su conjunto presentan un cierto grado de volatilidad y proporcionan sensación olfativa.

La idea que subyace en la división es la de permitir tratar los aspectos de detección y percepción de manera separada. En el caso de los gases, generalmente no interesan tanto sus características odoríferas sino otro tipo de características como pueden ser el rendimiento energético, el grado de explosividad, la toxicidad,... que además, casi siempre están en función de la concentración. En este sentido, interesa en estos casos una estimación cuantitativa. Por el contrario, en el caso de los olores, no interesa tanto conocer las concentraciones de tal o cual sustancia, sino extraer un patrón global que permita diferenciar cada olor. Aquí se trata, por tanto, de una estimación cualitativa. El organigrama de la figura 1.1 trata de hacer referencia precisamente a la diferenciación entre los intereses cuantitativos y cualitativos. Por un lado, el problema del análisis de gases ha de ser tratado con sensores, realizando una medición cuantitativa mientras que el problema de la diferenciación de olores se rige, en cuanto a la percepción natural, por una medición más bien cualitativa que en todo caso viene modulada por parámetros descriptivos de la intensidad global como leve, moderado, intenso,... La manera de aunar las vías analítica y perceptiva o, si se prefiere, cualitativa y cuantitativa, representadas en el organigrama de la figura 1.1 y desarrolladas a lo largo del capítulo, es a través del bloque final del organigrama: la nariz electrónica.

De una forma más concreta, a lo largo del capítulo se trata de dar una visión global de la problemática de los gases y aromas en el mundo actual. Además, se resumen algunas alternativas de detección artificial centrando sobretodo la atención en el campo de los sensores. Por otra parte, y en oposición/complemento de la medición artificial, se introducen ciertas nociones de percepción olfativa natural. Estos conceptos, en principio dispares deben ser llevados hacia un punto de convergencia común que es la nariz electrónica o nariz artificial. Para ello se introduce la noción de red neuronal artificial como algoritmo matemático capaz de imitar artificialmente los procesos que se llevan a cabo en la detección natural. Se repasará al final lo que se entiende por nariz electrónica y se realizará una breve exposición acerca de la historia de este dispositivo y de las narices electrónicas existentes en el mercado actual.

Así, a lo largo de este capítulo no sólo se presentarán los conceptos sino que se tratará de resaltar las analogías y diferencias entre la percepción natural y la artificial. De esta forma, cuando en capítulos posteriores se desarrollen modelos dinámicos para un determinado sensor, no sólo se hará pensando en el modelo en sí, sino que, de una forma más sutil, se pensará en el modelo y en el sensor como una pieza de un sistema mayor que ha de desembocar en la nariz electrónica, que como ya se ha dicho antes, trata de conjugar los aspectos relativos a la percepción y a la detección artificial de gases y aromas.

1.1. La percepción del mundo físico.

“ ... Así, fundándome en que los sentidos nos engañan algunas veces, quise suponer que no había cosa alguna que fuese tal y como ellos nos la hacen imaginar; y en vista de que hay hombres que se engañan al razonar y cometen paralogismos, aun en las más simples materias de geometría, y juzgando que yo estaba tan sujeto a equivocarme como cualquier otro, rechacé como falsas todas las razones que antes había aceptado como demostración; y, finalmente, considerando que los mismos pensamientos que tenemos estando despiertos pueden también ocurrirnos estando dormidos, sin que en este caso ninguno de ellos sea verdadero, me resolví a fingir que nada de lo que hasta entonces había entrado en mi mente era más verdadero que las ilusiones de mis sueños. Pero inmediatamente después caí en cuenta de que, mientras de esta manera intentaba pensar que todo era falso, era absolutamente necesario que yo, que lo pensaba, fuese algo; y advirtiéndome que esta verdad: pienso luego existo, era tan firme y segura que las más extravagantes suposiciones de los escépticos eran incapaces de conmovérla, pensé que podía aceptar sin escrúpulo como el primer principio de la filosofía que andaba buscando... ”

La filosofía cartesiana¹, entroncando con la filosofía clásica, se pregunta acerca de las relaciones entre el mundo real y el mundo de los sentidos. Cuestiones sobre si es real lo que se percibe o si existe aquello que no se percibe aparecen constantemente en filosofía y no es en absoluto ilógico planteárselas ya que somos, de hecho, prisioneros de nuestro propio yo y todo lo que sabemos del mundo se basa en información que recibimos mediante nuestros sentidos. Nuestra experiencia se basa en lo que vemos, lo que oímos, lo que olemos, lo que gustamos y lo que tocamos. Superar este solipsismo es, de alguna manera, una cuestión de conveniencia. Debemos aceptar como ciertas las informaciones que percibimos a través de los sentidos, no sólo en beneficio de la cordura, sino porque son similares a las experiencias del resto de personas y permiten la comunicación y el flujo de información entre las mismas. El mundo es real y en él existen objetos situados en el espacio y en el tiempo, que tiene ciertas cualidades a priori tales como color, temperatura, gusto, olor, forma, etc. La percepción de estos objetos por parte de un observador es una revelación directa del mundo físico, aunque no puede negarse que, en ocasiones, esta revelación puede resultar distorsionada. Pero los sentidos no sólo pueden provocar percepciones distorsionadas o ilusiones, además no son demasiado útiles como elementos de valoración cuantitativa de las diferentes cualidades sensibles de modo que, muchas cualidades físicas de los objetos especialmente adecuadas para caracterizarlos con exactitud, no excitan impresiones en nuestros sentidos o los excitan de forma vaga. La percepción a través de los cinco sentidos está por tanto sometida a problemas de precisión y subjetividad. La medición cuantitativa de las cualidades físicas ha de ser efectuada a través de algún artificio intermedio capaz de extraer la información requerida de manera objetiva y precisa.

Son los sensores las herramientas dedicadas a realizar la tarea de captar la información del mundo exterior de manera objetiva, mientras que la tarea del cerebro como procesador de información, verdadero integrador de las sensaciones y órgano que da sentido a la percepción, tiene que ser llevada a cabo por microprocesadores o sistemas DSP que de una manera diferente (mejores en cuanto a rapidez y objetividad y peores en cuanto a versatilidad, rendimiento y capacidad de abstracción) son capaces de realizar también el procesado de la información.

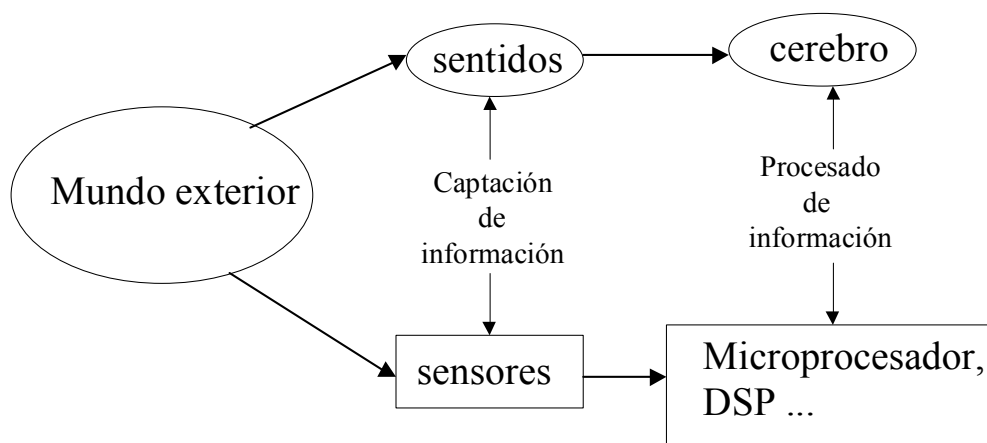


Figura 1.2 Paralelismo entre la medición y la percepción

Este trabajo concentra el objeto de interés en la medida cuantitativa y cualitativa de la presencia de gases o de aromas, y es por ello que, una vez establecida la realidad del mundo físico y la forma de conocerlo objetivamente, debe avanzarse en el entorno concreto de la medición de gases y aromas. En los siguientes apartados, tal y como se introduce en el prólogo del capítulo, se desarrollarán paralelamente aspectos relativos a la importancia e influencia de los aromas y los gases en nuestra sociedad, la detección natural de sustancias aromáticas, la medición artificial de sustancias gaseosas y la necesidad de obtener un compromiso entre la detección natural y la medición artificial a través de las llamadas narices electrónicas.

1.2. Los gases y su análisis: Problemática actual

Hasta mediados del siglo XX, el estudio de la composición de una muestra gaseosa era objeto de interés prácticamente sólo desde el punto de vista científico. Sin embargo, a partir de los años 70, la industrialización, la creciente demanda de gases para consumo industrial y doméstico, el consecuente control de calidad y los problemas de polución asociados a todo este fenómeno, generaron la necesidad en la sociedad de obtener información cuantitativa y fiable de la composición gaseosa del ambiente^{2 3}. Esta necesidad no sólo está localizada en el mundo industrial y en la generación de energía^{4 5}, también existe un especial interés en el entorno de la ecología y medio ambiente, así como en temas de seguridad y de control de la calidad del aire en ambientes cerrados, ya sea en el trabajo o en el hogar⁶. La problemática del análisis gaseoso no es sencilla. Por un lado, los aparatos de detección y medida de precisión son, en general demasiado grandes y caros como para permitir una utilización de los mismos de forma versátil para los usos comunes de una industria, siendo habitual que estos grandes aparatos estén restringidos a usos dentro del campo de trabajo de los centros de investigación. Por otro lado, los sensores de gas presentan problemas de derivas en el tiempo, no linealidades, menor precisión o inadecuación para las condiciones de medida exigidas. No obstante, la versatilidad que proporcionan aparatos desarrollados en base a sensores, su bajo costo en comparación con los grandes aparatos, su relativa simplicidad, y su portabilidad son algunas ventajas a considerar frente a

los posibles problemas que, por otra parte, de una u otra manera es posible minimizar, tal y como se verá a lo largo de este trabajo. Además, aunque actualmente todavía no es

común, la posibilidad de incorporar capacidades en el tratamiento de la señal hace de un sistema basado en sensores de gas portátil, una herramienta muy interesante y potente de análisis gaseoso. A continuación se considerarán algunas áreas de aplicación del análisis cuantitativo de gases:

- a) Generación de energía y ahorro energético
- b) Control medioambiental
- c) Seguridad

1.2.1. Ahorro energético

La necesidad de energía de nuestra sociedad es, desde luego, uno de los problemas que previsiblemente se arrastrarán sin solución hasta bien entrado el siglo XXI. Hoy por hoy, sea por intereses creados, sea por falta de alternativas eficientes, el gas natural, el carbón y el petróleo son fuentes de energía primaria, junto con la hidroeléctrica y nuclear, todavía insustituibles como base energética.







		Ktep	%	
Consumo de energía primaria en España (1998)		Petroleo	61.671	55.8
		Carbón	17.660	16.0
		Nuclear	15.375	13.9
		Hidráulica	3.103	2.8
		Gas	11.817	10.7
		Otras	818	0.8

Figura 1. 3 Consumo de energía primaria en España (1998) ^{w11}

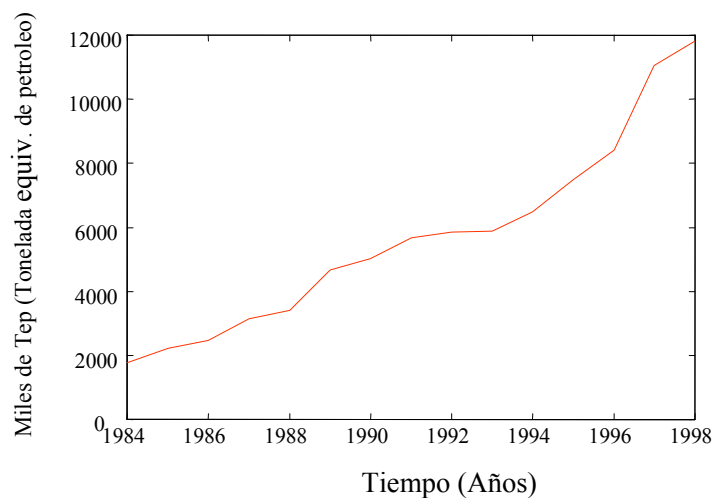


Figura 1. 4 Evolución del consumo de gas natural ^{w11}

Centrando la cuestión en los procesos de generación de energía a partir de reacciones de combustión y problemas medio ambientales aparte, estas reacciones plantean un problema de rendimiento energético. La pérdida de cantidades masivas de energía en los procesos energéticos combustibles se debe, generalmente a un mal funcionamiento de las calderas de combustión. Para un mínimo control del proceso resulta imprescindible realizar mediciones de oxígeno. Existen en el mercado dispositivos bien consolidados capaces de realizar este análisis con buena fiabilidad; por ejemplo las sondas lambda, un sensor ampliamente utilizado en el mundo de la automoción, que penetra en la corriente de los gases y genera una diferencia de potencial si entre sus extremos sensibles se detecta una diferencia en el contenido de oxígeno. Sin embargo, siendo esta medición necesaria, no suele ser suficiente. Es también importante conocer el contenido de CO de los gases que se producen para efectuar una comprobación completa del rendimiento de la caldera⁷. En este sentido, los estudios sobre el uso de sensores de CO más el procesado de señal adecuado, aplicado a la supervisión de calderas de combustión es motivo de estudio por parte del sector energético mundial.

Desde la perspectiva económica, el ahorro típico que se obtiene utilizando adecuadamente un control de combustión a través de un analizador de gases es, como mínimo de un 2% del combustible⁸. Por otra parte, la detección de un mal funcionamiento de la caldera o problemas del equipo es muy difícil si no se conocen los niveles de oxígeno y CO en los gases. En definitiva, un mínimo sistema de medida para el control de los procesos de combustión desde el punto de vista energético, debe incorporar por tanto, sensores de oxígeno y de monóxido de carbono, realizando lecturas periódicas de los gases expulsados. No debe descartarse sin embargo, la incorporación de sensores para el análisis de otros subproductos de la combustión como óxidos de nitrógeno u óxidos de azufre.

Tabla 1. 1 Emisiones de Gas Natural en cocinas domésticas (Kg/10 ⁶ m ³) ⁷²					
Partículas	CO	SO ₂	NO _x	Metano	Otros
16-80	320	9.6	1600	43	84

1.2.2. Control medio ambiental

La energía procedente de combustibles fósiles como el gas natural, carbón o petróleo genera una cierta cantidad de gases contaminantes. De hecho, la mayoría de estos contaminantes no sólo provienen de procesos industriales, las instalaciones de calefacción doméstica, la circulación de vehículos de motor, la navegación aérea y marina, contribuyen también en gran medida a la polución atmosférica, sin despreciar, por supuesto las causas naturales como volcanes, incendios o descomposición de materia orgánica en el suelo y en el mar.

Las emisiones de los vehículos motorizados por ejemplo contribuyen a la contaminación atmosférica en proporciones que se han llegado a estimar de hasta el 70% para el monóxido de carbono mientras que industrias como las energéticas, siderurgia, cementera, química e incluso alimentaria son responsables de un gran porcentaje de contaminación por óxidos de nitrógeno y azufre^{9 10 4}.

Las principales sustancias a considerar como contaminantes atmosféricos son:

Aerosoles	Compuestos halogenados
Partículas sólidas y líquidas	ácido clorhídrico y derivados del
Gases	cloro, ácido fluorhídrico y
óxidos de nitrógeno NO _x	derivados del flúor
óxidos de azufre SO _x	Compuestos orgánicos volátiles (COV)
dióxido de carbono CO ₂	Tolueno, Xileno, Octano, parafinas,
monóxido de carbono CO	naftalinas, acetonas, oleafinas y
hidrocarburos C _n H _m	demás compuestos aromáticos.
Metales pesados	Compuestos orgánicos azufrados
Plomo, Cromo, Cobre, Manganeso,	Mercaptanos
Vanadio, Níquel, Mercurio, Cadmio	Compuestos orgánicos halogenados
Sustancias minerales	PCBs, dioxinas, furanos, etc.
Asbestos, amianto	Sustancias radioactivas.

Todos ellos, en mayor o menor medida son responsables de diferentes efectos nocivos. Existen pruebas abundantes que indican que una concentración elevada de contaminantes en el aire afecta seriamente a la salud, con riesgos de intoxicación grave hasta el punto de causar la muerte en casos extremos. Aunque no es en este punto del trabajo donde se analizan los efectos en la salud de los contaminantes gaseosos, cabe comentar citándose exclusivamente a contaminantes gaseosos (como NO_x, SO_x, CO...), que puede comprobarse estadísticamente que la contaminación gaseosa produce un incremento de la mortalidad en personas de edad avanzada, aumenta los casos de bronquitis crónica, provoca trastornos relacionados con dificultades respiratorias e incrementa los problemas de cefaleas^{10 11}.

No sólo las personas reciben un fuerte impacto a causa de la contaminación, animales y plantas sufren daños significativos a concentraciones más bajas que las necesarias para provocar trastornos en un hombre, y los materiales sufren procesos de deterioro importantes o incluso irreparables, produciéndose la destrucción de monumentos y obras de incalculable valor artístico. Otros problemas asociados a la contaminación atmosférica son los que afectan a los ecosistemas en forma de lluvia ácida y efecto invernadero, causados por excesos de CO₂ y SO_x en la atmósfera¹⁰.

La mayoría de países industriales han desarrollado y siguen desarrollando reglamentaciones de lucha contra la contaminación atmosférica, adoptando estrategias en función de sus peculiaridades tanto culturales como ambientales. Cumbres internacionales como la celebrada en Río de Janeiro en 1992, y legislaciones cada vez más estrictas intentan minimizar las emisiones de contaminantes a la atmósfera³. Existen acuerdos para actuaciones concretas como el de la eliminación del uso de agentes contaminantes como los CFC o la minimización de emisiones de CO₂. Algunas normativas sobre la calidad del aire y de las emisiones de la comunidad económica europea son vigentes de ya hace tiempo, como la 80/779/CEE, directiva relativa a valores límites y guía para SO₂, en vigor desde 19/07/82, la 89/427/CEE que amplía la normativa anterior desde 11.01.91, la 85/203/CEE sobre normas para el NO₂, en vigor desde 01.07.87, la 80/372/CEE relativa a decisiones sobre los CFC y en vigor desde 31.12.81, la 77/102/CEE, relativa al control y medición de CO, CH₄, NO_x, etc. En la normativa española también existen decretos concretos relativos a emisiones de gases y control de usos de combustibles y carburantes como por ejemplo el decreto D/2204/85 sobre características, calidades y condiciones de empleo de carburantes y combustibles.

Sin embargo, estas legislaciones parecen todavía superadas por problemas de otra índole. En aras de la eficacia, toda decisión legislativa en el campo de la contaminación medioambiental debe fundamentarse en datos correctos obtenidos del conocimiento científico y técnico.

La necesidad pues, de control de los procesos industriales relacionados con la emisión de gases es un imperativo por un doble motivo: la adaptación a una normativa estricta que lo exige desde una óptica científico-técnica y que endurece cada día más las sanciones por su incumplimiento, y la posibilidad de reciclar y reutilizar ciertas sustancias con el consiguiente ahorro económico. La primera tarea a realizar pues para evaluar la viabilidad de las técnicas de control a utilizar consiste en realizar un inventario de las emisiones que se producen.

Los equipos analizadores para realizar el estudio de emisiones gaseosas son muy variados, dependen en gran medida de la aplicación concreta para la que son requeridos. Por ejemplo, para el control medio ambiental general se utiliza una red de estaciones de vigilancia fijas para suministrar información de los contaminantes atmosféricos en las zonas en las que los niveles de contaminación se prevén altos, en zonas pobladas, en zonas de cultivos y vegetación, o en zonas especialmente valiosas. Se suelen medir óxidos de nitrógeno, óxidos de azufre, partículas en suspensión, Compuestos Orgánicos Volátiles (COV), CO₂ y en general agentes causantes de lluvias ácidas, efecto invernadero y nieblas potencialmente tóxicas. El hecho de que las estaciones sean fijas, permite la utilización de equipos de análisis más precisos no portátiles. En general utilizan métodos analíticos como la espectroscopía, reflectometría o gravimetría. Sin embargo, para realizar estudios de emisión de vehículos automóviles se utilizan unas estaciones móviles con aparatos portátiles y semiportátiles (en el sentido de aparatos de un cierto tamaño que precisan de una cierta infraestructura pero no tienen que estar fijos en un local) para medir emisiones de CO₂, NO_x y SO₂, O₂, ozono, etc. Son aparatos basados en un más o menos amplio abanico de sensores entre los que destacan los sensores de infrarrojo, electroquímicos, cerámicos o de estado sólido. Cabe destacar, de entre este abanico de dispositivos de medición aplicados a la automoción, la sonda lambda, ya mencionadas en el apartado anterior. La sonda lambda suministra a la unidad de medida y control una señal relacionada con la composición momentánea de la mezcla gaseosa. Esta sonda se monta en el tubo de escape del motor, donde reina la temperatura necesaria para su adecuado funcionamiento. La efectividad de la sonda se basa en el material cerámico que la conforma, sobre la cual se insertan electrodos de platino. Este material cerámico y los electrodos son porosos, permitiendo la difusión del oxígeno del aire. La cerámica se hace conductora a temperaturas elevadas y si el contenido de oxígeno, como se mencionó ya en el apartado anterior, a ambos lados de los electrodos es diferente, aparece una tensión eléctrica entre los mismos.

Para analizar las emisiones de procesos en plantas industriales se utilizan, tal como muestra la figura 1.5, estaciones automáticas fijas para medir CO, CO₂, óxidos de nitrógeno, óxidos de azufre, oxígeno, etc¹⁰. Otras especies imprescindibles de cuantificar para la industria son los COV, no solo por su posible reciclaje y reutilización sino porque pueden ser potencialmente peligrosos a causa del bajo límite de explosividad de muchos de ellos (tolueno, octano), o sus potenciales riegos para la salud (COV cancerígenos para la fabricación de disolventes).

Existen varios tipos de sistemas para la realización de estas medidas industriales, pero el más utilizado es el llamado sistema de monitorización de emisiones continuo, basado en sensores electroquímicos, de estado sólido o de absorción de infrarrojos, que proporciona medidas directas de modo periódico (generalmente se realizan medidas

cada 15 minutos). También se suelen utilizar en la industria equipos simuladores, que ofrecen estimaciones de las emisiones a partir de los parámetros del proceso como flujo de combustible, calidad del combustible y temperatura^{12 7}.

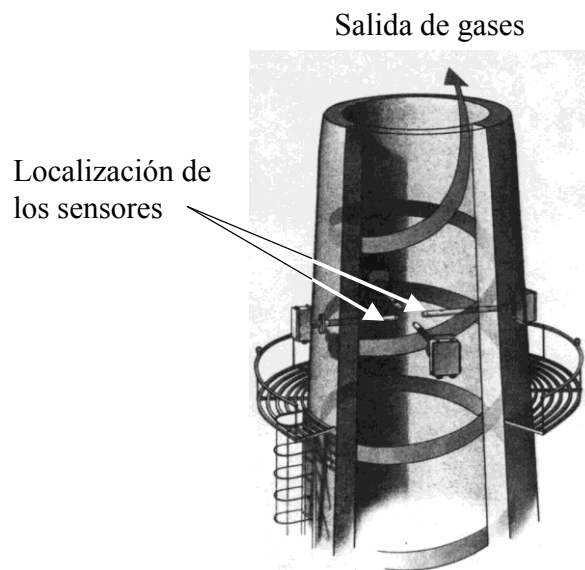


Figura 1.5 Medición gaseosa en una chimenea industrial

1.2.3. Seguridad

Si el apartado anterior aborda la incidencia de los gases desde el punto de vista de contaminación atmosférica general, este pretende ampliar los aspectos donde la contaminación gaseosa puede incidir directamente en la seguridad y la salud. La exposición continua a ambientes contaminados, u ocasional a productos gaseosos tóxicos debe ser prevenida en todo lugar de trabajo susceptible de estar sometido a tales eventualidades.

En primera aproximación, puede establecerse una clasificación de los tipos de gases que deben tenerse controlados. Así, puede hablarse de:

- gases explosivos: como metano, vapores de alcoholes o en general
COVs como tolueno, octano, xileno, ...
- gases tóxicos: como CO, sulfhídrico, óxidos de azufre, óxidos
de nitrógeno, cloro, cianuros, cloruros,
hidrógeno, ...
- gases perjudiciales: no son de gran peligrosidad pero contribuyen a
una mala calidad del aire: CO₂, ozono...

Los sistemas de alarma para los diversos gases deben estar sometidos a las normativas establecidas por los organismos nacionales e internacionales, donde se especifican los niveles de concentración recomendables y peligrosos para cada sustancia, así como las pruebas a las que estos sistemas deben superar con éxito.

Por otro lado, estos sistemas deben ser muy específicos, a fin de que se produzcan el menor número posible de falsas alarmas causadas por gases interferentes. Además deben trabajar un régimen de funcionamiento continuo, una precisión alta y un tiempo de respuesta relativamente pequeño. La gran mayoría de estos sistemas están sometidos

a una estandarización de sus salidas (es habitual la salida 4-20 mA), así como a comunicaciones típicas tipo RS-232 o RS-485.



Figura 1. 6 Sistema de alarma fija



Figura 1. 7 Sistema de alarma portátil

Los sistemas pueden ser fijos como el de la figura 1.6 para monitorizar niveles peligrosos en estancias ya de por sí susceptibles de estar expuestas a contaminación gaseosa, o portátiles como el de la figura 1.7 para que los trabajadores puedan llevarlo incorporado y avisar cuando están sometidos a ambientes peligrosos.

Pero no solo en ambientes laborales se sufren las consecuencias de la contaminación. Cualquier ambiente cerrado ya sea laboral, doméstico, edificios públicos, galerías comerciales etc. puede estar contaminado en mayor o menor medida provocando diversos efectos para la salud. La mala calidad del aire en ambientes cerrados y su problemática viene siendo conocida bajo la denominación de síndrome del edificio enfermo^{13 6}. La contaminación de estos ambientes proviene de muchas fuentes; el mero hecho de la presencia humana ya produce de forma natural CO₂, y vapor de agua; en las cocinas, por combustión pueden generarse CO₂, CO, óxidos de nitrógeno y de azufre; los aerosoles son fuentes de diferentes contaminantes y partículas; el humo del tabaco contiene mas de 3000 compuestos de los cuales una buena parte son contaminantes perjudiciales. No deben desdeñarse las fuentes provenientes de colas y pinturas, causantes de la presencia de formaldehídos y COVs, ni el ozono desprendido por fotocopiadoras o usado en procesos industriales de desinfección y limpieza, ni diferentes gases procedentes de productos de limpieza, etc.

La tabla 1.2 presenta un listado resumido de los contaminantes más frecuentes capaces de originar problemas de salud al ser inhalados así como la fuente de procedencia más común.

Todos los contaminantes listados contribuyen a originar problemas oculares como enrojecimiento, escozor o lagrimeo; problemas en vías respiratorias como goteo nasal, congestión, estornudos, sequedad de garganta, ronquera, sed, etc.; problemas pulmonares como sensación de ahogo, tos seca, opresión torácica; problemas cutáneos como eritema, sequedad cutánea, prurito generalizado, etc., o problemas generales como dolor de cabeza, somnolencia, dificultad de concentración, irritabilidad, mareos, etc.

Tabla 1. 2 Contaminantes usuales en ambientes cerrados

Productos de combustión	Materiales de construcción	Productos de consumo	Varios
NO _x	Fibra de vidrio	Pinturas	O ₃
CO ₂	Amianto	Barnices	Pb
CO	Compuestos orgánicos	Plásticos	Fe
SO _x	Disolventes	Colas	Mn
COVs	Formaldehído	Pegamentos	Al ₂ O ₃
humo de tabaco		Disolventes	Radón
aldehydos		Sellantes	
cetonas		Fibras textiles	
HCN		Insecticidas	
Nitrilos		Limpiadores	
Nitrosaminas		Cosméticos	
Nicotina			
Arsénico			
Cadmio			

De entre todos los contaminantes gaseosos, debe tenerse especial atención por su especial peligrosidad, con el dióxido de azufre y el monóxido de carbono. La presencia en el aire de elevadas concentraciones de CO representan una grave amenaza para la salud¹⁴. El CO inhalado se combina con la hemoglobina de la sangre formando la carboxihemoglobina, reduciéndose la capacidad para transportar oxígeno desde los pulmones a los tejidos. Se ha comprobado estadísticamente que una concentración por encima del 5% de carboxihemoglobina (COHb) comienza a producir ciertos leves cambios funcionales cardíacos y pulmonares, reduciéndose la capacidad de percepción visual y la habilidad manual y de lectura. A partir del 10% de concentración de COHb (nivel “normal” de un fumador) los síntomas cobran mayor fuerza. Por encima del 10% la COHb provoca efectos sobre la función psicomotora, cansancio, cefaleas, náuseas y alteraciones de la coordinación. En concentraciones entre el 30% y el 40% se producen dolores de cabeza graves, distorsiones en la visión, vértigos, fatiga y debilidad, vómito y puede llegar a provocar desmayos. Entre el 40% y el 50% arranca el rango de concentraciones que puede producir un desenlace fatal. En este rango la respiración y el ritmo cardíaco se acelera, se producen desfallecimiento y puede entrarse en coma. Si la concentración de COHb llega al 50% o es superior el individuo entra en colapso, el sistema respiratorio falla, se producen shocks y convulsiones, y finalmente el individuo muere. La relación entre concentración de carboxihemoglobina y concentración de CO depende del individuo y del tiempo de exposición. Para una persona media, en el sentido estadístico, la siguiente gráfica muestra la concentración de COHb en función del tiempo de exposición para diferentes concentraciones de CO y considerando al individuo en situación de reposo¹⁶.

Las líneas sólidas representan un ajuste matemático basado en modelos fisiológicos. Los marcadores representan mediciones reales de COHb en sangre. El departamento de electrónica de la Universitat de Barcelona ha trabajado y trabaja en esta problemática dentro de un proyecto para desarrollar un detector selectivo de CO/CH₄. En este proyecto, el problema de la integración del CO ha de tenerse bien presente y en este sentido y como fruto del trabajo desarrollado se han concebido modelos no lineales de integración del CO con el ánimo de cumplir las especificaciones de las normativas internacionales referentes a alarmas de CO domésticas¹⁵.

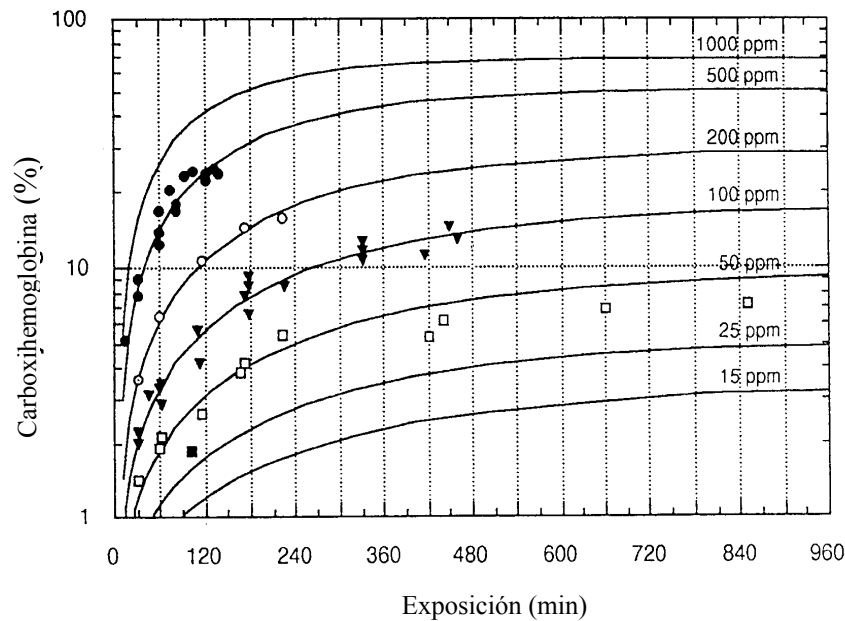


Figura 1.8 Relación entre el tiempo de exposición a un ambiente de CO y tanto por ciento de carboxihemoglobina en una persona ¹⁶

Por otra parte, el dióxido de azufre tiene una comprobada correlación con la aparición de bronquitis crónicas, provoca el incremento y la exacerbación de catarras así como dificultades respiratorias¹⁰. Cuando la concentración de SO₂ supera los 500 µg/m³ como promedio en 24 horas, se produce un aumento de la mortalidad en la población en general, siendo más afectados los individuos con procesos cardíacos o pulmonares. Los efectos del dióxido de azufre sobre la salud humana se pueden ver en la figura 1.8. Además de esta evidente peligrosidad para la salud humana, el SO₂ es un contaminante con alto índice de impacto en el mundo vegetal ya sea de forma directa, ya sea combinado con agua en forma de lluvia ácida, etc. No menos importantes son los óxidos de nitrógeno como el NO₂ cuyos efectos perniciosos para la salud se centran fundamentalmente en las vías respiratorias.

Comentario especial merecen las cocinas domésticas, ya sean a gas natural o a gas butano, que deben estar sometidas a controles periódicos rigurosos, no solo por la emisión de gases contaminantes propiciados por mala combustión sino por el evidente peligro de explosión ante posibles fugas. Aunque el gas distribuido lleva incorporado una serie de sustancias odorizantes para detectar el problema de las fugas, el usuario no puede estar permanentemente presente o puede padecer de anosmia frente al mencionado odorizante (cosa habitual en personas ancianas). Actualmente las revisiones para garantizar el correcto funcionamiento de las cocinas y calentadores son realizados por las propias empresas distribuidoras de gas, y para detectar fugas existen sistemas de alarma que, aunque no son de obligada instalación, si son recomendables. Sin embargo, y mirando la cuestión desde una óptica crítica, no debe dejar de comentarse que la utilidad de estas alarmas es cuestionable atendiendo a la gran cantidad de falsas alarmas que genera. Parece difícil, a corto plazo, que los sensores mejoren tanto como para evitar este problema, La solución hoy por hoy reside en un tratamiento inteligente de los datos de tal manera que el conjunto del sensor más el tratamiento sea capaz de discriminar entre verdaderas y falsas alarmas.

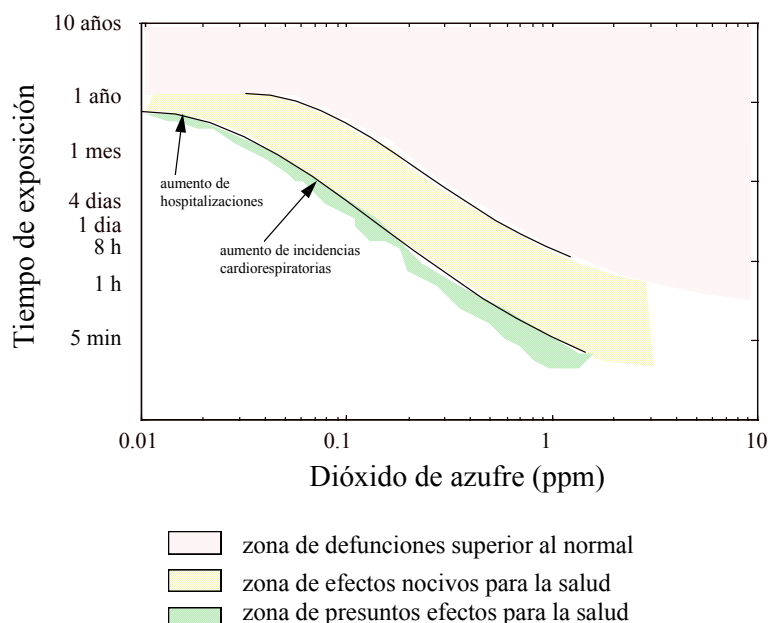


Figura 1.9 Efectos del SO₂ sobre la salud

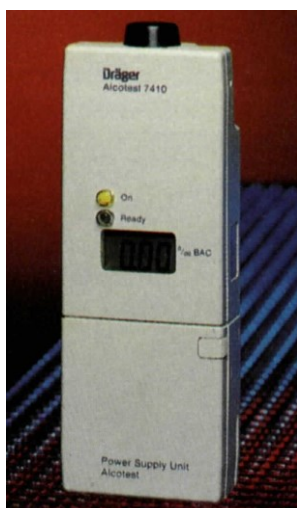


Figura 1.10 Alcoholímetro para el control de etanol en sangre a través del aire expirado

Concerniente a temas de seguridad aunque desde otra vertiente, conviene señalar, para concluir, el uso de detectores usados en controles de seguridad ciudadana.

Se utilizan en este campo sistemas sensores como los etilómetros¹⁸, encargados de detectar niveles de alcohol en sangre de los conductores a través del aire expirado basándose generalmente para ello en sensores de tres tipos: electroquímicos, de estado sólido o de infrarrojo. También se engloban en este entorno otro tipo de aparatos encaminados en la vía de seguridad ciudadana como los detectores de narcóticos (cocaína, heroína, ...),^{20 21} o detectores de explosivos²², que tratan de sustituir el uso de animales para este tipo de trabajos.

Si hasta este punto se han expuesto algunas facetas relativas a la aplicación de la detección y análisis de gases, a continuación se presentan de forma resumida, las aplicaciones más típicas del análisis de olores y aromas, siendo estas la extensión natural y complementaria de las anteriores.

1.3. Aplicaciones de la determinación de olores y aromas

La percepción de olores mediante el sentido del olfato consiste en detectar las moléculas volátiles del entorno y reconocerlas. Es por tanto, junto con el sentido del gusto, un sentido que funciona como un sistema de detección química muy evolucionado. La percepción de un sabor o de un olor no tan sólo nos proporciona información, sino que se utiliza como método de relación con el entorno tanto desde el punto de vista social, como cultural como incluso defensiva. En cualquier caso, la sensación olorosa, en general, no suele estar asociada a un solo componente gaseoso, sino que en dicha

sensación suelen participar una mezcla heterogénea de componentes volátiles. Por ejemplo, el aroma a café es producido por una combinación de numerosas sustancias volátiles, procedentes de otras no volátiles que existen en la semilla original y acaban derivando mediante el proceso de tostado. En las semillas verdes, se pueden encontrar hidrocarburos, ésteres, compuestos sulfurados, alcoholes, furanos, aldehídos, cetonas, etc. Así, un olor o un aroma puede considerarse como una mezcla gaseosa multicomponente cuya discriminación puede llegar a ser muy compleja. De hecho, la concentración de las diferentes especies que componen el aroma no tiene porqué estar directamente relacionada con la sensación aromática. Al contrario, en muchos casos, pequeñas trazas de una especie de la mezcla determinan completamente el perfil aromático de la sustancia; trazas que, por otro lado, pueden llegar prácticamente indetectables en un análisis químico convencional y cuya presencia sólo puede ser evidenciada a través de aparatos muy específicos y de alta resolución.

Evidentemente, no es siempre factible disponer de estos mencionados aparatos y sin embargo, las implicaciones del mundo olfativo en la sociedad, hacen que resulte interesante disponer de un método de reconocimiento y caracterización de olores y gases. Una vía alternativa a la medición de alta resolución es la determinación sistemática de los perfiles olfativos de las sustancias olorosas, de tal manera que no es necesario conocer la concentración de la traza sino el perfil que esa traza otorga al olor. La utilidad de un sistema así en el mundo industrial se centraría principalmente en campos como la alimentación, perfumería y cosmética, y medicina. La idea no es novedosa, ya que la determinación de un perfil olfativo es justamente la tarea de un perfumista (si hablamos de cosmética y perfumería) de un catador (si hablamos de alimentación) o del propio médico (dentro del mundo de la medicina), pero si que es poco usual disponer de una herramienta sistemática capaz de emular esas tareas de forma automática.

1.3.1 Cosmética y perfumería

La historia de la cosmética y la perfumería se desarrolla paralelamente a la evolución de la cultura. Los primeros perfumes fueron de origen vegetal, y se obtenían de la combustión de gomas, resinas y líquenes y, de hecho, el término perfume proviene de estas antiguas prácticas: per fumum, “a través del humo”. Existen numerosos documentos dan fe de la utilización de los perfumes en la antigüedad; e incluso en la Biblia (Exodo 30 22-26) aparecen recetas para la fabricación de perfumes y óleos sagrados. Sin pretender desarrollar en este punto la historia de la cosmética, sí que es interesante reseñar que en la actualidad la industria de este sector mueve una cantidad de dinero del orden de centenares de miles de millones de pesetas. Es evidente que este volumen de negocio precisa de un control de los procesos de fabricación y un control de calidad muy bien establecido. Esencialmente, los encargados de realizar estas tareas son la ciencia química y los perfumistas. A grandes rasgos, la tarea de un perfumista consiste en calificar el producto mediante un test olfativo así como también desarrollar nuevos perfumes y fragancias. La profesión de perfumista requiere varios años de entrenamiento riguroso y especializado y es exigible una cierta habilidad en el sentido del olfato, pero no es necesaria una habilidad extremadamente alta. La habilidad de un perfumista radica más en como el cerebro trabaja con los datos que en el sentido del olfato por si mismo. Este concepto debe ser también aplicado a los sistemas de medida artificial para el reconocimiento de patrones; dicho de otro modo, disponer de buenos

sensores siempre es deseable, pero no es fundamental. Lo verdaderamente importante es la forma de procesar los datos que se obtienen de los sensores.

En cualquier caso, sí es cierto que no es deseable que un perfumista sufra anosmias parciales pronunciadas, pero la mayoría de ellos, al igual que la mayoría de personas, no están libres de algún tipo de anosmia. De hecho, algunos de los perfumistas más habilidosos son incapaces de oler ciertos aromas puros (usualmente almizcle o aromas amaderados), aunque si pueden reconocerlos en mezclas. Este es un fenómeno familiar en fisiopsicología, por el cual, el umbral de detección de una sustancia es diferente en estado puro que en mezclas. También es conocido que el sentido del olfato cambia con el tiempo, y no sólo a largo plazo (en escala de años) sino que también lo hace a lo largo del día, y puede depender de factores poco controlables como la comida o el tabaco. Por todo ello, el hábito de trabajo de los perfumistas es muy sistemático, y aun así, es difícil que siempre esté al mismo nivel de eficacia. De todos estos argumentos expuestos, se hace evidente que un dispositivo artificial (como la nariz electrónica) aportaría un valor añadido a la tarea del perfumista, puesto que no se vería afectada por aspectos fisiopsicológicos ni por otros factores poco controlables.

1.3.2 Alimentación

En las industrias de alimentación los olores y aromas juegan un papel básico. Por un lado, la medición aromática es un factor que se utiliza para el control de la calidad de los productos, ya sean carnes, pescados, vinos, cafés, etc. Por otro lado, dado que el sabor de un alimento viene determinado más por el olfato que por el gusto, se usan aromas para conferir artificialmente sabores concretos a los diferentes productos. Dentro del mundo industrial, las tareas que en perfumería y cosmética realiza un perfumista tienen su analogía en alimentación en la figura del catador. De hecho, tal como ya se ha dicho, la percepción de los sabores está íntimamente relacionada con el olfato y por tanto, esta analogía es clara y directa. Por esta misma razón, un experto catador no está exento de estar sometido a los mismos problemas que un perfumista y en este sentido, lo reseñado acerca de los sistemas de medida artificiales como complemento útil en las tareas de reconocimiento puede también aplicarse al mundo de la alimentación.

Existen numerosos trabajos científicos relativos al reconocimiento y clasificación de olores y aromas con aplicaciones en este sector. Trabajos sobre reconocimientos de diferentes cafés, whiskys, vinos, aceites o sobre la determinación del buen o más estado de alimentos son habituales en las revistas especializadas^{23 24 25 26 27}. Por otro lado, se están estudiando y se plantean como factibles sistemas basados en las técnicas de fusión de datos. Dentro del departamento de electrónica de la Universidad de Barcelona se han realizado avances en este sentido, compaginando datos provenientes de una nariz electrónica y una lengua electrónica para la determinación de la calidad de patatas chips y la identificación de las salsa que las acompañan⁷¹. En cualquier caso, lo que debe quedar claro es que las aplicaciones de los sistemas artificiales de medida de gases basados en sensores de bajo costo, en campos con tanto mercado como la perfumería y la alimentación son plenamente factibles; existe la necesidad y existe la tecnología y la tendencia actual de generar sistemas orientados a la aplicación hacen que el sistema final pueda resultar a un precio competitivo. Con todas estas variables, no cabe duda que, aunque ahora en sus primeros pasos, el reconocimiento de patrones artificial acabará siendo utilizado con asiduidad por la industria.

1.3.3 Medicina

El mundo de los olores y aromas también tiene presencia y un cierto papel a jugar en cuestiones relacionadas con la salud⁶³. Desde un punto de vista más comercial que de rigurosidad científica, aparece la aromaterapia como disciplina que se basa en la creencia que ciertos aromas pueden aliviar o incluso curar ciertas enfermedades (sobre todo de carácter psicológico). Se utilizan para ello ciertos aceites esenciales con fines tanto terapéuticos como estéticos. Así, por ejemplo, se cree que el uso de la albahaca o el vetiver inciden directamente sobre problemas de ansiedad o tensión, la mejorana o el ylang-ylang ayudan a superar procesos depresivos, y la salvia actúa sobre el cansancio mental y la irritabilidad. Sin embargo, y aunque existe un mercado relativamente importante relacionado con la aromaterapia, no existen pruebas concluyentes que certifiquen los supuestos resultados.

Desde otro plano, si que es cierto sin embargo, que la medicina ha utilizado y utiliza los olores para realizar diagnósticos de ciertas enfermedades⁶³. Las pruebas en aire expirado se pueden utilizar como alternativa a otras técnicas invasivas que siempre implican un cierto grado de peligrosidad y comportan molestias. Los avances conseguidos en las técnicas analíticas permiten identificar cada vez mayor número de compuestos en el aire expirado que permiten realizar diagnósticos precoces de muchas enfermedades. Si el aire expirado contiene información sintomatológica es porque el aire de los alvéolos pulmonares sólo está separada de la sangre por una delgada membrana. Los compuestos orgánicos volátiles tales como el etanol pueden atravesar dicha membrana incorporándose al aire que posteriormente se expirará. Desde tiempos de Hipócrates ya se conoce el hecho de que ciertos olores en el aliento ayudan a realizar diagnósticos. Olores dulces y afrutados indican la presencia de acetona que entronca con la existencia de problemas de diabetes, olor mohoso y a pescado indica problemas hepáticos, olor similar a la orina sugiere disfunciones renales y olores putrefactos avisan de la posibilidad de abscesos pulmonares. Continuando en ello y concretando, la exposición de los sujetos a ambientes contaminados puede acelerar la aparición de ciertas enfermedades como la arteriosclerosis que pueden prevenirse a través de las pruebas de aliento. Existe una alteración intestinal denominada síndrome de malabsorción que provoca diarrea crónica severa ya que el intestino no es capaz de absorber los alimentos de manera normal. Una consecuencia de este síndrome es que la mayoría de azúcares llegan intactos al colon y allí se metabolizan por las bacterias existentes. Esta metabolización genera hidrógeno que pasa a la sangre y se excreta a través de los pulmones. Para detectar el síndrome, se le suministra al paciente xilosa, un azúcar que, en condiciones normales, es completamente absorbido en el intestino delgado. La aparición de cantidades inhabituales de hidrógeno en el aire expirado confirma el diagnóstico. Diferentes refinamientos de estas técnicas permiten diagnosticar diversos problemas intestinales, problemas pancreáticos, úlceras pépticas, gastritis crónicas o problemas hepáticos e incluso esquizofrenia, cuyo proceso agudo provoca la aparición de pentano en el aire expirado, decreciendo la concentración a medida que el sujeto mejora su estado.

De nuevo se hace evidente que un sistema de detección artificial de gases y olores tiene aquí también un papel a jugar como elemento de medida alternativo y como elemento adicional de apoyo al diagnóstico de ciertas enfermedades.

Otra aplicación importante de la detección y reconocimiento de olores y aromas, aparte de las tres ya expuestas, está centrada en los procesos de control y calidad, por ejemplo

aplicados al medioambiente con mediciones de aire, aplicados a clasificación de calidad de aceites industriales, o de materiales plásticos o tejidos, etc.

En el marco actual del presente trabajo la atención se centra en el estudio de gases no estrictamente relacionados habitualmente con la olfacción. Así, aunque se plantea la posibilidad a medio plazo, actualmente no se han realizado pruebas efectivas con sustancias olorosas ya sean fragancias o aromas.

Cabe resumir finalmente, que realizar clasificaciones olfativas no es tarea sencilla y es un tema abierto a discusión. En general se suelen utilizar referencias olfativas, denominadas notas pero aspectos fisiológicos y psicológicos pueden afectar en la percepción del olor y en general al rendimiento del sistema olfatorio. Entonces, así como resulta altamente interesante conseguir un sistema de medición gaseosa sencillo, potente, portátil y económico, también ha de resultar útil desarrollar un sistema capaz de realizar las tareas de reconocimiento de olores con la eficacia del sistema olfatorio de un perfumista profesional, un catador experto o un médico, pero sin la subjetividad inherente a la persona.

Revisados hasta este punto la problemática separada entre gases y olores según se presentaba en la figura 1.1, y siguiendo precisamente dicho esquema, se desarrollan a continuación de forma diferenciada, los diferentes mecanismos (sensores como dispositivos de medida artificial y el olfato como sistema de medida natural) de captar la presencia de gases y olores.

1.4. Problemática de la medida: Sensores

La tarea de identificar diferentes especies químicas o bioquímicas en atmósferas se soluciona aplicando distintas técnicas analíticas. Los sensores químicos y bioquímicos son una pequeña parte de este mercado de instrumentos de análisis pero tienen un indudable valor a causa de su bajo costo, su pequeño tamaño y al hecho de que pueden ser integrados en pequeños sistemas electrónicos autónomos.

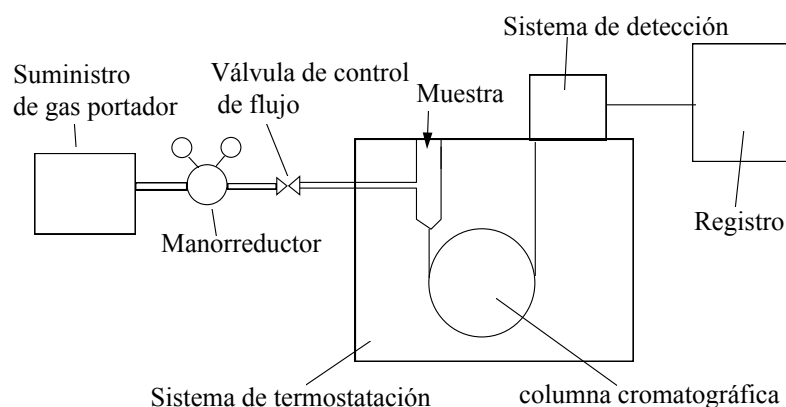


Figura 1. 11 Esquema de un cromatógrafo de gases

Durante los últimos decenios las técnicas de separación cromatográfica de alta resolución en combinación con la espectroscopía de masas o espectrómetros de infrarrojo en transformada de Fourier han permitido adentrarse profundamente en la composición química de las complejas mezclas naturales que componen los olores naturales. Estas modernas técnicas han incrementado notablemente el nivel de sensibilidad y selectividad de las medidas²⁸. Un ejemplo, la vainilla, contenida en un porcentaje de hasta 4.5% en las vainas de vainilla, ya fue bien caracterizada en el siglo

pasado. Sin embargo, gracias a los modernos métodos de análisis, han podido aislarse hasta 120 compuestos volátiles diferentes en vainas de vainilla. Esta es sin duda una enorme potencia de análisis, pero también supone unas necesidades mayores. Precisan este tipo de aparatos, aparte de la mayor o menor inversión económica, un laboratorio fijo donde ubicarlos y una cierta infraestructura que incluye sistemas de control del aparato mismo, diferentes equipos y dispositivos de apoyo y generalmente uno o varios PCs para realizar el control, la medición y/o el almacén de los datos.

Estos condicionantes hacen que ciertos tipos de mediciones no puedan ser satisfechas de manera eficiente por estos grandes equipos y se hace imprescindible buscar alternativas que, sin llegar a las precisiones de los grandes aparatos, tenga prestaciones satisfactorias, sean portátiles y de costo más asequible. Se debe tener en cuenta además, que para aplicaciones relacionadas con reconocimiento de olores no es estrictamente necesario, e incluso es una tarea inabordable, tratar de conocer cuantitativamente todos los componentes de un compuesto. Bastaría con disponer de un sistema capaz de diferenciar cualitativamente la información que recibe, del mismo modo que actúa el olfato de los mamíferos.

Tabla 1.3 Algunos sensores utilizados en la detección y análisis de gases

Tipo	Principio de operación	Disponibilidad	Ventajas	Desventajas
Óxidos metálicos	Conductivo	Comercial, muchos tipos	Baratos	Operan a alta temperatura
Polímeros	Conductivo, Capacitivo,...	Comercial bajo petición específica	Operan a temperatura ambiente	Grave interferencia con humedad
Microbalanzas de cuarzo QMB	Piezoeléctrico	Comercial, varios tipos	Tecnología bien conocida	Fabricación, interfaz electrónica
Surface Acoustic Wave (SAW)	Piezoeléctricos	Comercial, varios tipos	Pueden ser muy sensibles	Interfaz electrónica
Ópticos	Fluorescencia, Luminiscencia, Interferometría ...	Comercial e investigación	Inmunidad a ruido eléctrico	Necesidad de una fuente luminosa
CHEMFETS	Capacitivo acoplamiento, función de trabajo	Comercial bajo petición específica	Buena integración con la electrónica del sistema	El analito debe penetrar en la puerta
Calorimétricos (pellistores)	Calor de reacción	Comercial	Baratos	Operan a alta temperatura. Se envenenan con facilidad
Electroquímicos	Conductivo	Comercial, varios tipos	Buena linealidad	Se envenenan fácilmente

Como se comentó anteriormente, la concentración de especies de las muestras olorosas no están directamente relacionadas con el olor que se percibe, y en casos extremos, componentes que sólo se encuentran en trazas, pueden determinar por sí solos el perfil olfatorio de una muestra. Estas trazas pueden incluso no llegar a ser detectadas ni por

los aparatos analíticos más sensibles. Así, un conjunto de sensores no necesariamente muy específico en cuanto a su respuesta a diferentes gases, más un sistema de tratamiento de la información adecuado puede resultar tan útil en tareas de reconocimiento como un preciso cromatógrafo o cualquier técnica analítica de gran precisión. Si de lo que se trata es de alto poder de separación y alta resolución, qué duda cabe que los grandes aparatos analíticos aportan una precisión difícilmente igualable; pero a costa de perder ventajas ya mencionadas como portabilidad, sencillez de diseño, economía o facilidad de realización de medidas on-line. A menudo es conveniente sacrificar algo de precisión (por otra parte no siempre exigible) para obtener las ventajas.

El objetivo, pues, es desarrollar un sistema de detección gaseosa, a base de sensores de gas, relativamente barato, fácilmente transportable, de sencilla implementación y capaz de realizar tareas de reconocimiento cualitativo y de medición cuantitativa.

El abanico de posibilidades que ofrece el mercado de sensores de gas es amplísimo^{29 30} y a fin de poder tener una cierta clasificación sistemática de los mismos se recurre a ordenaciones en base a parámetros como principio de operación, tecnología utilizada, soporte físico, etc. Aun así, las clasificaciones siempre resultan difusas produciéndose múltiples intersecciones entre los conjuntos establecidos. Asumida esta dificultad se presenta la tabla 1.3 que incluye un listado de algunos de los sensores más comúnmente utilizados en análisis de gases junto con su principio de operación.

En los siguientes apartados se profundiza en la descripción de algunos de estos sensores.

1.4.1. Sensores electroquímicos

Es bien sabido que en toda reacción química tiene lugar una reorganización eléctrica en la cual se modifican las distribuciones de densidad electrónica entre los átomos de las especies que intervienen. Cuando este balance electrónico es negativo, se dice que se ha oxidado, mientras que en caso contrario, se dice que se ha reducido. Los sensores electrolíticos y las denominadas pilas de combustible aprovechan este fenómeno para ofrecer una señal eléctrica relacionada con el gas combustible. El principio de funcionamiento de las pilas de combustible y de los sensores electroquímicos es el mismo y, por tanto, aunque el campo de aplicación de ambos dispositivos es bien diferente (los sensores se utilizan obviamente para realizar mediciones y las pilas de combustible funcionan como fuentes de energía) en este apartado ambos términos se confunden. Esencialmente, un sensor electrolítico es un dispositivo capaz de convertir la energía química de los reactivos en energía eléctrica basándose en el principio REDOX.

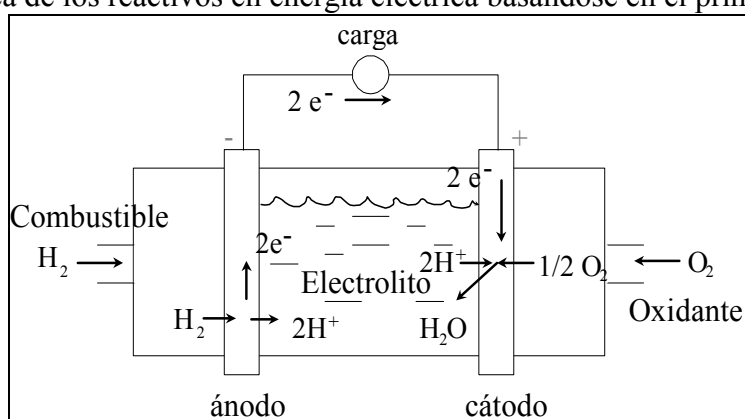
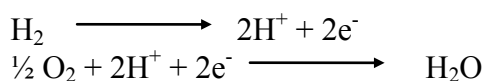


Figura 1.12 Esquema de un sensor electrolítico

En cierta manera, existe una gran similitud entre un sensor electrolítico o una pila de combustible y una pila o un acumulador. Sin embargo, mientras que una pila primaria o un acumulador consume el propio material que la constituye, en las pilas de combustible, el reductor y el oxidante vienen aportados desde el exterior y los electrodos no sufren, idealmente, ninguna transformación irreversible. Idealmente, estos dispositivos, cuyo funcionamiento se esquematiza en la figura 1.12, funcionan tanto tiempo como se aporten los reactivos necesarios.

Los sensores electrolíticos constan, en esencia, de un ánodo o electrodo de combustible y de un cátodo o electrodo oxidante, separados por un electrolito que realiza la función de conductor iónico. Los electrodos, se conectan exteriormente a través de una carga y mediante un circuito metálico por el cual fluyen los electrones provenientes de la reacción anódica. Mientras tanto, a través del electrolito fluyen los iones desde el ánodo al cátodo. En el cátodo los iones y los electrones se recombinan en una reacción de reducción de la sustancia oxidante. Los electrodos han de poseer ciertas características especiales, por ejemplo deben ser porosos para que la superficie de la reacción sea la mayor posible y para permitir el paso de las especies iónicas a su través, además han de ser buenos conductores y poseer un cierto poder catalítico para que la reacción tenga lugar en proporciones significativas. Un buen candidato a electrodo que cumple con estos mínimos exigidos es el platino negro, que se puede obtener como resultado de la precipitación de ciertas sales de platino. En la figura 1.12 se ha presentado un sensor con reacción hidrógeno-oxígeno. El hidrógeno llega al ánodo, donde gracias a la acción catalítica del electrodo se descompone fácilmente en iones H^+ y electrones. Los electrones fluyen por el circuito exterior hasta llegar al cátodo, mientras que los iones fluyen por el electrolito hasta el mismo cátodo. Allí confluyen junto con el oxígeno, provocando la formación de agua.

Las reacciones de oxidación-reducción son:



Aunque el esquema básico presentado es de dos electrodos, también existen en el mercado sensores electroquímicos que trabajan con tres electrodos. Este tercer electrodo actúa simplemente como electrodo de referencia y confiere al sensor una mayor selectividad frente a la configuración con dos electrodos que, por otra parte, obviamente es más barata.

La teoría sobre los sensores electroquímicos es bien conocida, pero su realización práctica es compleja y relativamente cara^{31 32}. El problema fundamental radica en la fabricación de los electrodos ya que suelen ser metales preciosos y la tecnología de fabricación no es sencilla. Además del platino, se pueden utilizar metales como níquel, iridio o aleaciones de carbón platino o estaño platino.

Como sensores de gas, tienen una aplicación inmediata, sólo es necesario que el gas a detectar sea fácilmente oxidable, por ejemplo cualquier tipo de alcohol, o hidrocarburos en general. Sin embargo, presentan graves problemas de envenenamiento que hacen que la vida media de estos sensores en comparación con otros tipos, sea baja y que su sensibilidad vaya disminuyendo a lo largo del tiempo.

Comercialmente, pueden encontrarse sensores de diferentes fabricantes como Dragger, City Technologies, Endress+Hauser o FIS (serie "ED" electroquímica).

En cuanto a sus aplicaciones más típicas, se centran en la detección gaseosa en la industria petrolífera, en la automovilística y en la minería para detectar gases como

monóxido de carbono y óxidos de azufre y de nitrógeno. Otra aplicación muy común es la detección de etanol en sangre a través del aire expirado. La mayoría de aparatos portátiles que utiliza la policía para este fin incorporan sensores electroquímicos de electrolito líquido. Prototipos de este tipo de dispositivos fueron desarrollados por el autor bajo un proyecto de colaboración universidad empresa. Los esfuerzos invertidos en este proyecto fructificaron en artículos¹⁷ y trabajos de investigación^{18 19} así como en un modelo de utilidad⁴⁴. Los prototipos de los sensores se desarrollaron a partir de polvo de PVC que, mediante un proceso térmico se sinterizan unas obleas que se sumergen en ácido sulfúrico. El ácido queda retenido en los poros de la oblea y harán la función de conductor electrolítico. La fabricación de los electrodos se realizó a través de una deposición química a partir de cloruro de platino. Del cloruro de platino se obtiene el platino poroso que se deposita sobre la oblea hasta conseguir un electrodo de buena conductividad y gran poder catalítico. El gas combustible, en este caso el etanol se oxida liberando electrones e iones H^+ . Los electrones fluyen por el circuito eléctrico mientras que los protones atraviesan el electrolito hasta llegar el otro electrodo donde todo se recombina en una reacción de reducción que cierra el circuito^{18 19}.

Pero no sólo existen sensores electrolíticos con electrolito líquido, también se han desarrollado este tipo de sensores pero con un electrolito sólido. La estructura de estos últimos es similar a la anterior; consisten básicamente en un disco electrolítico encerrado en un tubo cerámico, rodeado de un calentador que permite trabajar a alta temperatura, donde el electrolito es más eficiente. El disco, forma dos cámaras separadas, la cámara de referencia y la cámara de muestra. En cada cámara, y cubriendo al disco, se encuentran los electrodos.

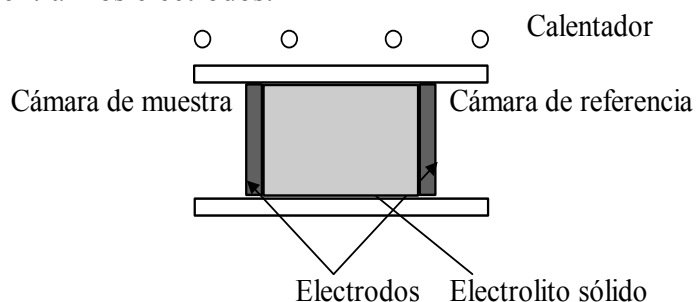


Figura 1.13 Esquema de un sensor de electrolito sólido

Como electrolito se suele utilizar óxidos de circonio, aunque también es habitual utilizar membranas de intercambio iónico como el Nafion o el Nasion.

Las prestaciones e inconvenientes de los sensores de electrolito líquido continúan apareciendo en los de electrolito sólido. La ventaja de estos últimos sobre los primeros radica precisamente en el electrolito sólido, que evita los problemas de corrosión que los electrolitos líquidos pueden presentar.

1.4.2. Sensores de polímeros

El fundamento sensible de un sensor de polímeros se basa, esencialmente y sin entrar en consideraciones sobre actividad de superficie o de absorción en el cuerpo, en la capacidad del polímero de interactuar con las moléculas de las especies químicas que se le presentan. A través de esta interacción se producen cambios en las propiedades del polímero que pueden ser aprovechadas para obtener una variable relacionada con la concentración de la especie química. Aunque estas variaciones pueden ser muy diversas,

las que habitualmente se usan para fabricar dispositivos sensibles son: variaciones de conductancia que permiten fabricar sensores de polímeros conductivos, variaciones de índice de refracción o de grosor que permiten fabricar sensores de polímeros ópticos, variaciones de permitividad eléctrica o de grosor que permiten fabricar sensores de polímeros capacitivos y variaciones de masa que permiten fabricar sensores de polímeros SAW (Surface acoustic wave) o QMB (quartz microbalance) de alguna de las propiedades del polímero (conductividad, permitividad, capacitancia, masa,...) ^{36 37 38 39}. Los polímeros habitualmente usados son polisiloxanos, polypirrolas, thiopenos, indoles o furanos y algunos de ellos son conocidos y usados en cromatografía como material de las columnas de retención cromatográficas. En la tabla 1.4 se listan algunos de estos polímeros.

Por tener relación con medidas de variable eléctrica, en este apartado se van a desarrollar una breve explicación sobre los sensores conductivos y los capacitivos basados en polímeros, mientras que para los sensores de masa y los ópticos se han reservado apartados específicos.

Centrándose pues en los sensores conductivos, el mecanismo por el cual la interacción con la especie química produce cambios de la conductividad del polímero está producido por los enlaces químicos que las moléculas de esta especie forman con el cuerpo del propio polímero, siendo el enlace habitualmente iónico, aunque también es posible que se produzcan enlaces covalentes. La consecuencia fundamental de la generación de estos enlaces es que se modifica la transferencia de electrones a través de la cadena polimérica y por tanto, varía su conductividad.

La forma básica de los sensores de polímeros conductivos consiste en dos electrodos separados por unos 10 o 20 μm entre los cuales se inserta el polímero mediante, habitualmente, procesos de electropolimerizado.

En cuanto a los sensores capacitivos basados en polímeros, hay que diferenciar entre el efecto de variación de la capacidad basada en el incremento del grosor del polímero al interactuar con la especie química, y el efecto de la variación de la permitividad eléctrica del polímero. Aunque, por supuesto, es factible aunar ambos efectos en un mismo sensor hay que decir que la contribución de la variación de grosor a la variación de capacidad del dispositivo, salvo en casos concretos en la que la permitividad del polímero es especialmente grande, suele ser siempre muy pequeña y prácticamente no puede resolverse con medidas de capacidad. En cualquier caso, en la figura 1.14 se presenta el esquema de un sensor capacitivo basado en polímeros.

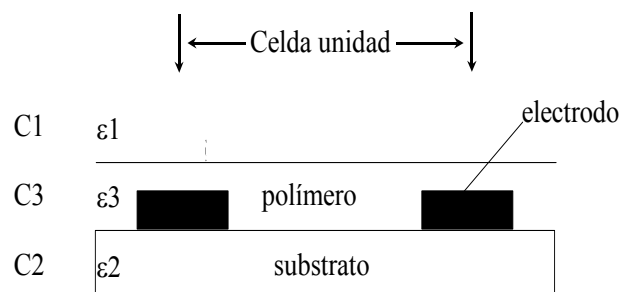


Figura 1.14 Esquema de un sensor capacitivo.

La capacidad total del dispositivo es la suma de las contribuciones de la fase gas (C1), el substrato (C2) y el polímero (C3). Las tres capacidades presentan diferentes permitividades eléctricas, pero sólo en C3, a través de la interacción de la especie

química con el polímero, se modifica esta permitividad, así como también el grosor tal como ya se ha dicho, mientras que el resto de capacidades permanecen prácticamente constantes. En realidad, un dispositivo basado en sensores capacitivos consiste en una serie de celdas unidad como la del esquema, interdigitadas.

Entre las ventajas de este tipo de sensores de polímeros, cabe destacar el hecho de que no necesitan, al igual que los electroquímicos, de calefactor puesto que trabajan a temperatura ambiente, lo que simplifica su diseño y su modo de operación. Además, los procesos de interacción del polímero con la especie química siempre son reversibles y el costo del dispositivo final es bajo. Sin embargo, en contra tienen el hecho que es habitual que presenten graves problemas de sensibilidad cruzada con la humedad y, en el caso de los capacitivos, los accesorios de medida, si se requiere mediciones de impedancia, son caros. Algunas empresas que comercializan este tipo de sensores son PHILLIPS, Neotronic Scientific, MOTET,...

Tabla 1. 4 Polímeros usados para la fabricación de sensor de gas	
DMPS	Dimetilpolisiloxano
MTFP	Metiltrifluopropilpolisiloxano
PCM	Policianopropilmetilpolisiloxano
CMP	(10-12%)cianopropilmetil(88-90%)dimetilpolisiloxano
MPD	(45-55%)Metilfenil(45-55%)difenilpolisiloxano
PS	Polistrirola
PE	Polietileno
PVC	Cloruro de polivinilo
PAN	Poliacrinitrilo
PVP	Polivinilpirrolidona
PVAC	Acetato de polivinilo
PEG	Polietilenglicole
PVPYR	Polivinilpirrol

1.4.3. Sensores SAW y microbalanzas de cuarzo (QMB)

En el desarrollo del apartado anterior no se incluyeron los sensores poliméricos basados en cambio de la masa del polímero a causa de la interacción con la especie química. Es en este apartado, continuación el anterior, donde se van a presentar los dispositivos que permiten aprovechar esta característica de los polímeros. En este sentido, los dos dispositivos esenciales que son capaces de realizar esta tarea son los dispositivos SAW (Surface Acoustic wave) y los QMB (Quartz Microbalance).

El esquema esencial de un dispositivo SAW se presenta en la figura 1.15.

Un dispositivo SAW está configurado por un material piezoeléctrico sobre el que se colocan dos electrodos denominados IDT (Interdigital Transducers). Cuando se aplica una señal r.f. a uno de ellos, el material piezoeléctrico comienza, alternativamente a comprimirse y expandirse, generándose ondas acústicas que se transmiten por la superficie del material a una cierta velocidad. Si la frecuencia es la adecuada y la distancia entre electrodos es correcta, se formarán interferencias constructivas de tal manera que la onda podrá llegar al segundo electrodo donde se producirá el proceso contrario, generándose una señal r.f a partir de la onda acústica que llega.

Un dispositivo acústico puede utilizarse para medir la cantidad de masa depositada en su superficie, convirtiéndose así en un sensor másico. La variación de frecuencia respecto

de la frecuencia de referencia causada por un incremento de masa viene dada por una relación de este tipo:

$$\Delta f/f_{\text{res}} = S \Delta m \quad (1.1)$$

siendo f_{res} la frecuencia de resonancia, Δf la variación de frecuencia, Δm la variación de masa y S un factor de sensibilidad.

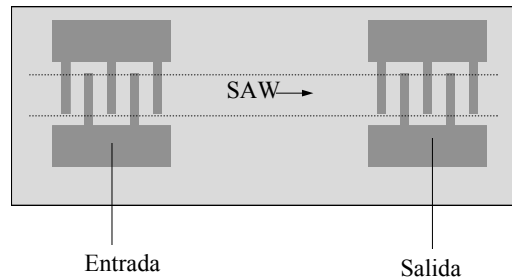


Figura 1.15 Esquema de un dispositivo SAW

Los sensores másicos clásicos por excelencia son las microbalanzas de cuarzo o Quartz Microbalance (QMB) y son de hecho, dispositivos BAW (Bulk Acoustic Wave) diseñadas específicamente para la detección de masa. Las estructuras BAW tienen un fundamento físico idéntico a los dispositivos SAW, pero la transmisión de ondas se hace a través del volumen del material en vez de por la superficie (ver figura 1.16). La utilización estructuras SAW para generar sensores másicos es posterior a los primeros diseños de QMBs. Históricamente, la idea de los detectores basados en el efecto piezoeléctrico y la absorción de moléculas fue introducida por King en 1964, usando sensores BAW⁵⁰, mientras que los desarrollos basados en tecnologías de sensores SAW aparecieron por primera vez en 1979⁵¹. Desde entonces, este tipo de sensores ha venido siendo investigados y utilizados cada vez con mayor frecuencia.

El típico sensor QMB se basa en un disco piezoeléctrico fino con los electrodos dispuestos en las caras opuestas del disco. La frecuencia de resonancia depende del material piezoeléctrico y de su grosor. Si al dispositivo se le añade algún polímero capaz de adsorber moléculas, el cambio de frecuencia de resonancia que se produce debido al incremento de masa viene dado, en primera aproximación y suponiendo que el incremento de masa es despreciable frente a la masa total del dispositivo, por la siguiente ecuación:

$$\Delta f = -\frac{1}{\rho k_f} f_0^2 \frac{\Delta m}{A} \quad (1.2)$$

donde ρ es la densidad del cristal, k_f es una constante, A es la superficie sensible, f_0 es la frecuencia fundamental de oscilación y Δf e Δm son las variaciones de frecuencia y masa.

El incremento en la masa está relacionado a su vez con la concentración de gas analito, y en consecuencia, también lo está la frecuencia según demuestra en primera aproximación la ecuación (1.3).

Los sensores másicos basados en estructuras SAW responden, en primera aproximación según una ley similar a las estructuras BAW de los QMB:

$$\Delta f = (k_1 + k_2) f_0^2 \frac{\Delta m}{A} \quad (1.3)$$

donde k_1 y k_2 son constantes del material.

El material sensible que cubre al piezoeléctrico debe poseer unas propiedades básicas y esenciales. Principalmente debe ser un material no volátil y debe permitir una fácil difusión de los vapores. Además, el material no debe sufrir alteraciones a causa de la exposición a las sustancias volátiles ni deben llevar al dispositivo a sufrir problemas de histéresis.

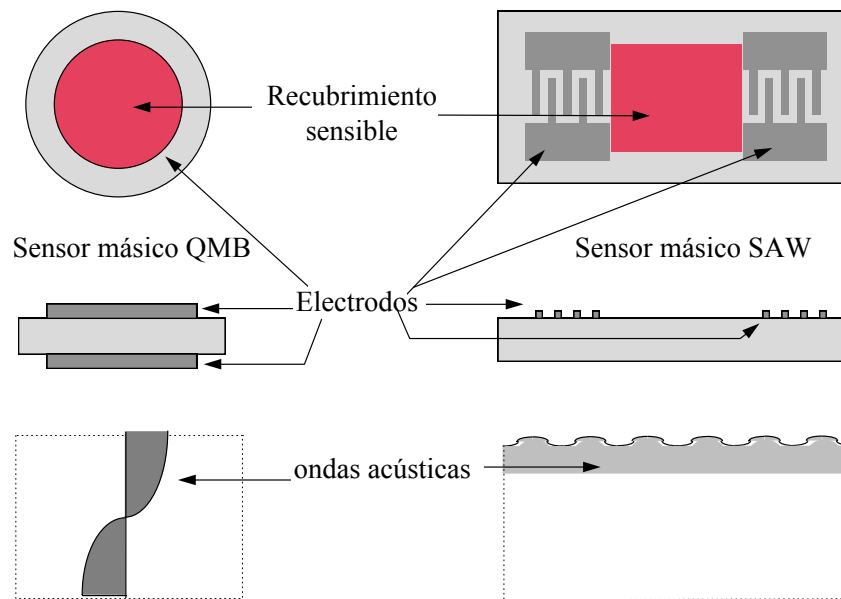


Figura 1.16 Esquema de estructuras BAW y SAW

A pesar de que las estructuras BAW y SAW funcionan bajo el mismo principio, existen varias diferencias importantes. En primer lugar, las estructuras SAW son capaces de operar a frecuencias de al menos dos órdenes de magnitud superior que las BAW. Además, la geometría de las estructuras SAW permiten trabajar con montajes de una sola cara, que proporciona mayor robustez. También el tamaño los diferencia, las estructuras SAW son intrínsecamente un microsensors y ocupan un volumen mucho menor que las BAW a la vez que presentan menor no-linealidad. Para finalizar, los dispositivos basados en estructuras SAW se pueden fabricar de forma múltiple, uno al lado de otro, en el mismo sustrato. Actualmente, no es fácil encontrar en el mercado este tipo de sensores. Su fabricación prácticamente está reservada a nivel de laboratorio y centros especializados, aunque algunas empresas disponen de líneas de fabricación (SRI International, Femtometrics Inc, MOTEC,...).

1.4.4. Sensores ópticos

Al igual que con otros sensores y otras categorías, la clasificación de los sensores dentro de la categoría de sensores ópticos siempre es difusa. Dentro de ella pueden considerarse dos clases principales: los sensores basados tecnología interferométrica y los basados en absorciones en el infrarrojo⁴⁰. El primer tipo de sensores, expresamente excluidos junto con los dispositivos másicos del apartado de sensores poliméricos, aplica técnicas interferométricas para determinar el cambio en el grosor o cambios en el

índice de refracción que experimentan los polímeros sensibles usados. Los límites mínimos de grosor detectables son del orden de amstrongs. Con esta precisión, se pueden llegar a límites de detección de 1 ppm para algunos gases. Hoy en día, la posibilidad de integración de todo un interferómetro hace de los sensores ópticos una opción interesante. La integración más simple consiste en realizar un Fabry-Perot a través de un canal de guía de onda monomodo y con dos extremos parcialmente espejados. Variaciones sobre la misma idea son el interferómetro Mach-Zehnder y el interferómetro de Michelson.

Las desventajas de este tipo de sensores son la necesidad de un tipo especial de materiales poliméricos sensibles, y la necesidad de accesorios caros (fuentes de luz y guías de luz).

El segundo tipo de sensores se basan en las absorciones de radiación infrarroja que se producen en gracias a los enlaces atómicos de las moléculas de gas. Simplemente, el gas se introduce en una cámara óptica y se somete a una radiación infrarroja bien determinada. Usualmente, la fuente de luz infrarroja suele ser de banda ancha, y la determinación de la banda de interés se realiza mediante filtros interferométricos. Las moléculas de gas absorberán parte de esta radiación y a la salida se detectará el resultado en forma de picos de absorción a diferentes frecuencias, dependiendo del tipo de gas y de diferente amplitud, dependiendo de la concentración de gas. Un sistema de este tipo es complejo, de hecho, el elemento sensor real suele ser de tipo piroeléctrico, pero es necesario incluir en el sistema de una fuente luminosa adecuada y de una cámara de absorción. Es por ello que propiamente, debe hablarse de un detector de infrarrojo en lugar de sensor. Estos detectores son, por otra parte, muy precisos pero son caros y complejos. Empresas que disponen de sensores y sistemas de medida ópticos son: Cal Sensors, Dexter Research Center Inc. , EG&G Heimann Optoelectronics,...

Cabe destacar que el el departamento de electrónica de la Universitat de Barcelona, se han diseñado y construido detectores IR para CO, CO₂, CH₄ bajo contrato para Gas Natural sdg⁷³.

1.4.5. Sensores de óxidos metálicos.

Fue en la década de los cincuenta cuando se descubrió la propiedad de los óxidos metálicos semiconductores tipo n de cambiar de conductividad de ante la presencia de gases inflamables^{45 46}. Se iniciaron entonces los estudios para comprender la naturaleza de esta propiedad pero la utilización del fenómeno para la detección de gases no fue concebida hasta 1962 cuando Seiyama el al⁴⁷ propuso la detección de gases inflamables a través de cambios en la resistencia de una capa fina de óxido de Cinc (ZnO) a la vez que Taguchi⁴⁸ propuso un sinterizado poroso de óxido de estaño (SnO₂). Desde entonces, los sensores de óxidos metálicos semiconductores, y especialmente el óxido de estaño, se han convertido en quizás el tipo de sensor más importante en la industria de los sensores de gas^{33 34 35}.

Aunque se han desarrollado una gran variedad de estructuras sensores basadas en óxido de estaño, los mecanismos sensibles son complejos³⁴. A modo de sumario breve se puede decir que el fundamento sensible se basa en los cambios de conductividad que presenta este material durante su interacción con las moléculas del gas analito. En presencia de aire puro, las moléculas de oxígeno presentes son adsorbidas en la superficie del SnO₂ en forma de especies iónicas, principalmente O⁻ y O₂⁻. Así, el óxido de estaño cede electrones de su banda de conducción al oxígeno, la conductividad de esta zona superficial disminuye respecto al resto de material, formándose una capa de

depleción. El resultado general es que el material pierde electrones, la densidad de carga electrónica disminuye, y, por tanto, la conductividad general del material también decrece. En resumen, la resistencia eléctrica de los óxidos metálicos semiconductores aumenta ante la presencia de oxígeno. Esta explicación general y simple es muy intuitiva y muy clara, pero los procesos concretos que se producen realmente son complejos y su estudio y comprensión es difícil. Por ejemplo, parece haber quedado demostrado que la conductividad global del óxido de estaño viene determinada fundamentalmente por las contribuciones de nanocristales de óxido de un cierto radio cristalino. Sin embargo, salvando estas particularidades, es comúnmente aceptado que la estructura granulosa del material juega un papel importante. Para una buena conducción, los electrones deben moverse libremente de un grano a otro, pero el hecho de que el oxígeno atrape electrones para formar especies iónicas, genera una barrera de potencial que dificulta este tránsito. Esta barrera de potencial se debe a la aparición de una zona de carga espacial en la frontera de cada grano a causa de la donación de electrones de la banda de conducción por parte del óxido de estaño. Al mismo tiempo, la capacidad de adsorción de oxígeno por parte del material decrece, llegando a un estado en el que queda compensada la adsorción y la donación de electrones y que proporciona el valor global de la resistencia.

Cuando el sensor es colocado en una atmósfera con el gas reductor a analizar, normalmente las moléculas del gas reaccionan con el oxígeno adsorbido y liberan los electrones que habían quedado atrapados. De esta manera, la conductividad aumenta, disminuyendo la resistencia eléctrica del material^{42 43}.

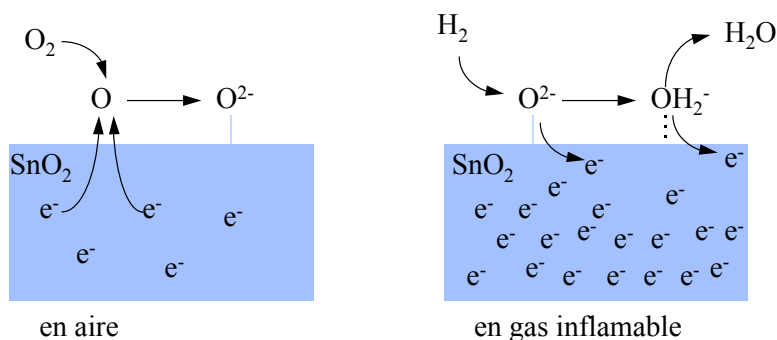


Figura 1.17 Funcionamiento de un sensor de óxido de estaño

En base a este fundamento sensible, es aceptado que la dependencia de la conductividad con la concentración sigue una ley aproximada potencial de la forma:

$$\Delta G = k C^r \quad (1.4)$$

donde C es la concentración del gas y k y r son constantes características para cada par gas-sensor.

Para alcanzar buenas prestaciones de selectividad, sensibilidad y estabilidad hay que tener en cuenta y optimizar factores como la geometría de los contactos, el grosor de las capas, el dopaje o la temperatura de operación. Estos parámetros llevan a diferentes comportamientos de los sensores frente al gas analito y mediante dichos parámetros se puede conseguir que los sensores estén optimizados para la detección principal de los gases que interesen primordialmente.

La estructura clásica de un sensor de óxido de estaño se representa en la figura 1.18

Consta, principalmente, del elemento sensible, consistente en óxido de estaño sinterizado el cual contiene además algún catalizador, generalmente paladio. Además dispone de una resistencia calentadora que permite alcanzar una temperatura de trabajo óptima.

El óxido de estaño sensible se obtiene habitualmente mediante un proceso húmedo a partir de cloruro de estaño (SnCl_4). Aunque este proceso está bien establecido, existen muchas variantes en la forma de aplicarlo a la fabricación de sensores de gas y de hecho, la tecnología de fabricación es aun una cuestión de debate en la que intervienen muy diversos parámetros. En cualquier caso, existen dos tendencias generales en la fabricación de los sensores de óxido de estaño: la tecnología de capa fina y la tecnología de capa gruesa (la figura 1.18 corresponde a un esquema de sensor de capa gruesa).

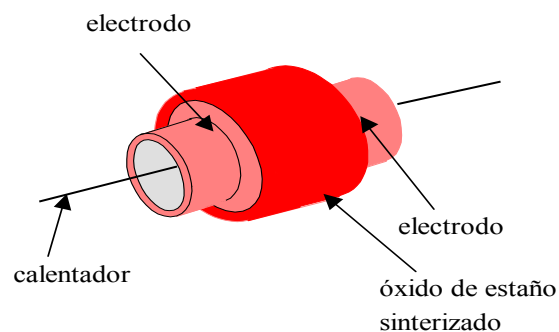


Figura 1.18 Esquema de un sensor de óxido de estaño

Ambas tecnologías presentan sus ventajas e inconvenientes; así, con tecnología de capa fina es posible fabricar sensores de menor consumo que con tecnología de capa gruesa y además, permiten beneficiarse de las posibilidades de la tecnología microelectrónica. A favor de la tecnología de capa gruesa están factores como mayor sensibilidad y selectividad.⁴⁹

Los sensores de óxido de estaño comerciales se fabrican según ciertas estructuras determinadas que se presentan en la figura 1.19.

Coexisten en el mercado varias empresas dedicadas a la fabricación de sensores de gases (FIGARO, FIS, NEMOTO, MICROSENS...) aunque de hecho las dos primeras, FIGARO Y FIS, dominan el mercado mundial.

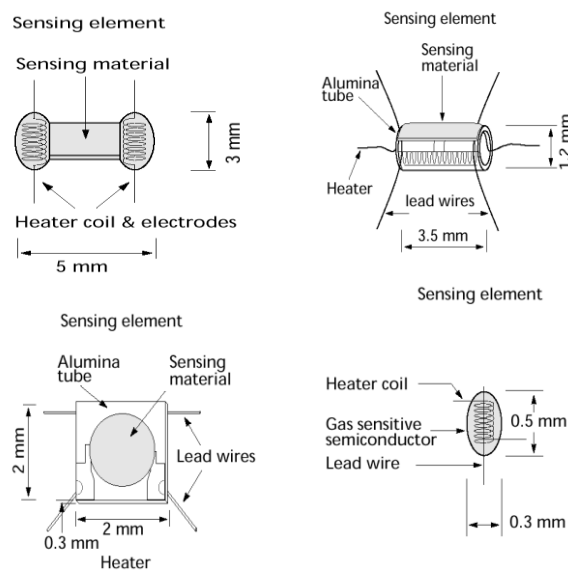


Figura 1.19 Estructuras básicas del elemento sensible de un sensor de óxido de estaño (sensores de FIS)

En las siguientes imágenes se presentan el aspecto externo de diferentes sensores comerciales, así como una imagen de microscopía del elemento sensible de un sensor de capa fina sobre sustrato de silicio.

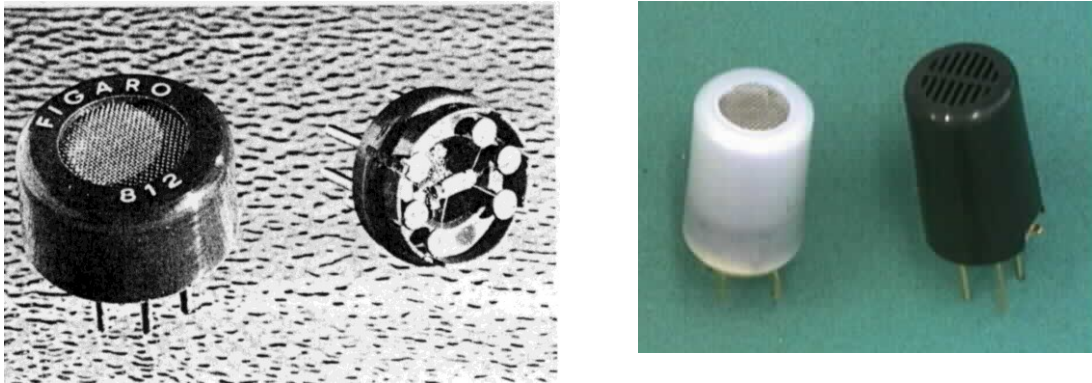


Figura 1. 20 Sensores comerciales de óxido de estaño

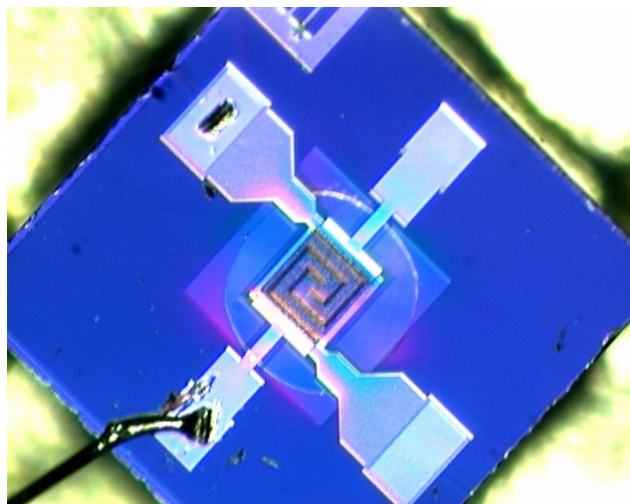


Figura 1. 21 Estructura interna de un sensor de óxido de estaño de Microsens.

A pesar del relativo éxito comercial de este tipo de sensores, no están exentos de problemas. Los principales inconvenientes se resumen en que son sensores poco específicos y aunque en ocasiones pueden llevar incorporados filtros de carbón activo para aumentar la selectividad, en general responden a una gran cantidad de gases diferentes.

Presentan también alta dependencia con la temperatura y humedad y además, suelen exhibir derivas a largo plazo y sufren de cierto grado de no linealidad. Por el contrario, y como ventajas hay que hacer constar que son sensores muy baratos, responden rápidamente y son muy sensibles.

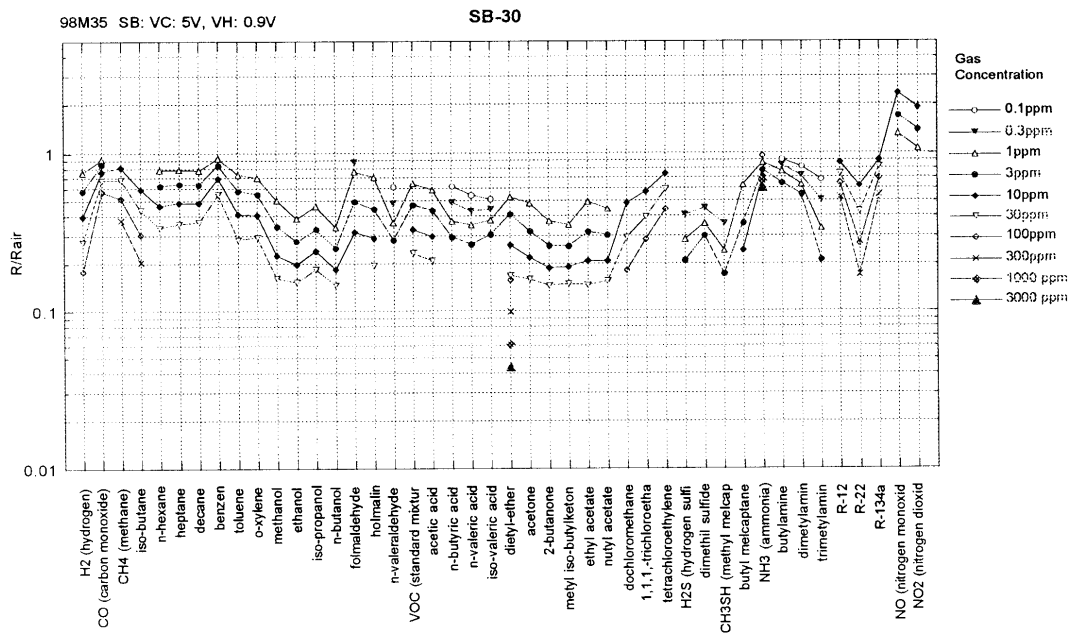


Figura 1. 22 Característica sensitiva del sensor SB30 de FIS a diferentes gases

1.4.6. Sensores calorimétricos o catalíticos

Los sensores catalíticos son dispositivos sensores que se caracterizan por el hecho de que el material sensor actúa también como calentador del dispositivo. En general consisten en una gota de alúmina activa que contiene ciertas cantidades de catalizadores como platino, paladio o rutenio. Cada gota esta montada sobre un hilo de platino, tal como muestra la figura 1.23.

Los pellistores son el ejemplo típico de sensor catalítico. Consisten en dos arrollamientos de hilo de platino dentro de sendas gotas de alúmina, y conectados eléctricamente a un circuito puente. Una de las gotas es catalíticamente activa y promueve las reacciones de oxidación mientras que la otra se trata especialmente para inhibir dicha reacción. El funcionamiento del sensor se basa en alcanzar, a través de una corriente circulando por el arrollamiento, una cierta temperatura para la cual se produzca oxidación en la gota catalizadora. La reacción

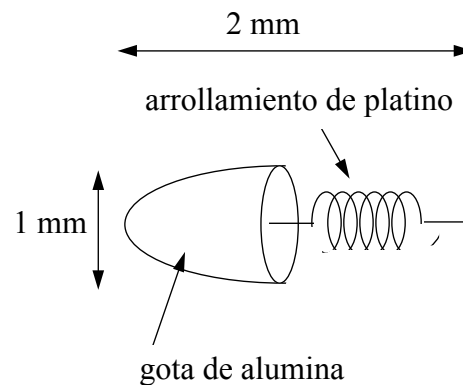


Figura 1. 23 Esquema de un sensor catalítico

provoca una variación de temperatura que supone una variación de la resistencia del hilo de platino. Esta diferencia provoca un desequilibrio en el puente que puede ser registrado. Los pellistores se usan sobre todo para gases combustibles, son baratos y presentan buena linealidad. La principal desventaja es que se envenenan con mucha facilidad. Pueden encontrarse en el mercado varias empresas que desarrollan líneas de producción de sensores catalíticos. Algunas de ellas son Pewatron, Crowcon, Rae, City Technology,...

1.5. Detección natural: La percepción olfativa.

Frente a este repaso general de diferentes sensores de gas, se va a abordar a continuación, en oposición a la detección artificial, una breve descripción de algunos aspectos interesantes de la detección natural y, aunque esta descripción es, en cierta manera, un contrapunto a la detección artificial, no es incompatible con ella, sino que la complementa en el camino hacia el concepto de percepción artificial inteligente.

En la diversidad existente en el mundo actual, como resultado de procesos evolutivos producidos a lo largo de toda la historia, pueden observarse especies animales muy dispares y adaptadas perfectamente a su entorno natural. A pesar de esta disparidad, las células sensoriales involucradas en la recepción de los olores han cambiado poco. No solamente la base química esencial de la percepción olfativa es similar en todos los animales, sino que se mantiene también en la mayoría de especies el nexo de unión entre la percepción olfativa y funciones fisiológicas como la alimentación o la reproducción, por ejemplo.

Lo que sí que ha evolucionado es la capacidad de los animales para reconocer un amplio espectro de estímulos químicos, bien por reconocimiento innato o por aprendizaje, y la habilidad para aprovechar estos conocimientos en su conducta. De hecho, el papel del sentido del olfato en el conjunto de la percepción no es ni más ni menos importante que el resto de sentidos, y sin embargo, interviene en funciones tan básicas como la búsqueda y selección de alimentos a través de la regulación de las funciones digestivas y nutritivas o como guía de los fenómenos sexuales, ayudando al encuentro entre los dos sexos. Interviene también, como mecanismo de defensa, descubriendo situaciones peligrosas o posibles enemigos a través de olores característicos que delatan dichas situaciones a distancia.

Desde el punto de vista de la fisiología humana, el olfato, junto con el gusto, conforman los denominados sensores químicos. Ambos sentidos se basan en la respuesta más o menos selectiva, de células sensoriales especializadas a la presencia de moléculas de determinados compuestos. Tanto el gusto como el olfato colaboran conjuntamente en la sensación de sabor de los alimentos, sin embargo estos sentidos son claramente diferentes tanto morfológicamente como fisiológicamente. Las células de las papilas gustativas, por ejemplo, al igual que otras células epiteliales tienen una vida corta (unos 10 días), renovándose constantemente por células nuevas, mientras que las células olfativas tienen una capacidad de regeneración muy limitada.

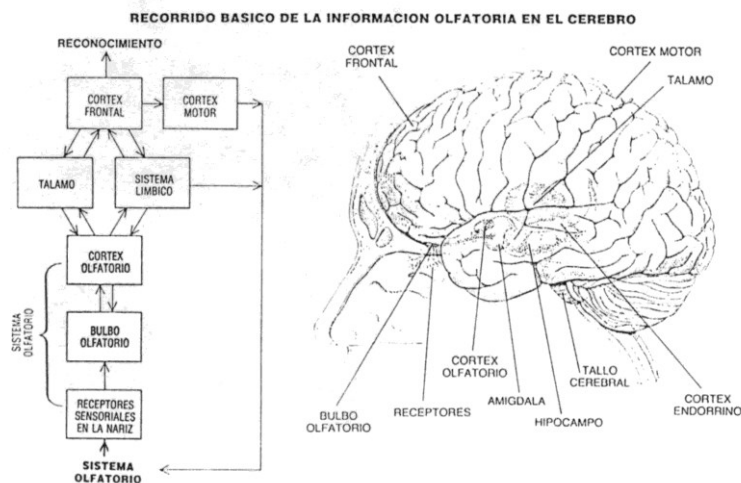


Figura 1. 24 Esquema del sistema olfatorio humano.

La tabla 1.5 especifica algunas diferencias fundamentales entre estos sentidos. Centrando ya la atención sobre el sistema olfatorio humano principal, puede decirse que éste consta de: los receptores, situados en las mucosas de las fosas nasales, el bulbo olfatorio y las proyecciones olfatorias que conectan con el sistema nervioso que activa la percepción. La activación de la percepción actúa sobre diversos sistemas somáticos y produce diversas respuestas tanto emocionales como racionales y relacionales. Los siguientes apartados se dedican a describir con algo más de detalle cada uno de estos elementos.

Tabla 1.5	Gusto	Olfato
Receptores	Células sensoriales de fácil regeneración	Células sensoriales de difícil regeneración.
Situación	En la lengua	En la cavidad nasofaríngea
Estímulo	Moléculas orgánicas y no orgánicas, sustancias la mayoría volátiles. Fuente de estímulos en la vecindad o directamente en contacto con el órgano sensorial.	Moléculas casi exclusivamente orgánicas, compuestos volátiles en forma de gases. Fuente de estímulos la mayoría de las veces a gran distancia.
Estímulos distinguibles	Bajo, cuatro cualidades fundamentales	Muy alto (miles), numerosas clases de cualidades
Sensibilidad absoluta	Baja: al menos 10^{21} moléculas/ml solución	Alta: 10^7 moléculas/ml aire
Caracterización biológica	Sentido cercano, control de alimentación, control de la ingestión (reflejos salivares)	Sentido lejano y cercano, control de medio ambiente (higiene), control de alimentación, comunicación, procreación.

1.5.1. Los receptores olfativos

La nariz contiene los receptores olfatorios, aunque de hecho, estos ocupan sólo una pequeña parte de la cavidad nasal. Hay que remarcar que la función principal de la nariz es la respiración, pero tanto para la función respiratoria como para la función olfatoria es necesario que el aire inspirado sufra un proceso de calentamiento, humidificación y filtración. Estos procesos se realizan en las fosas nasales gracias al alto riego sanguíneo existente en el interior de las cavidades nasales (la temperatura media del aire en las fosas nasales oscila entre los 36 y los 37 grados centígrados) y a la secreción seromucosa de una serie glándulas distribuidas por la superficie nasal interna.

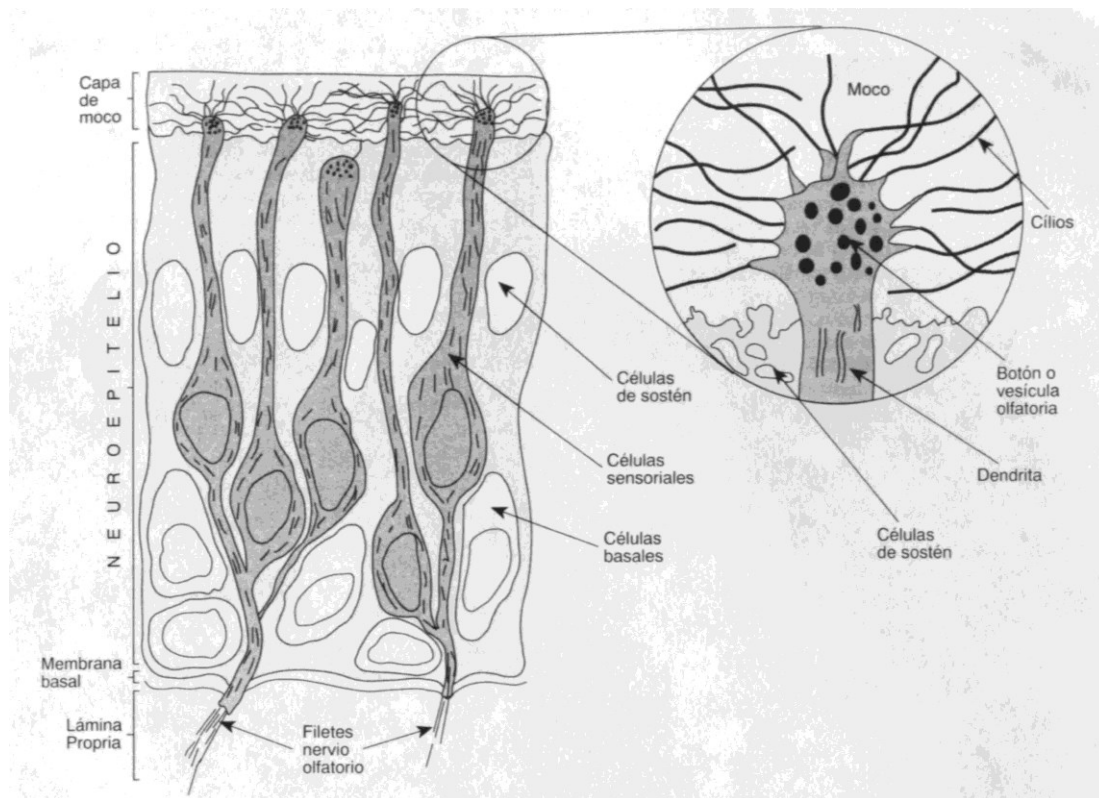


Figura 1. 25 Esquema de las células receptoras del epitelio olfatorio

El área olfatoria (epitelio olfatorio) donde se encuentran los receptores olfatorios se halla alojada en la parte superior de las fosas nasales y, en el hombre, ocupa unos 10 cm². Las células sensoriales receptoras son pequeñas células bipolares, que envían una conexión sensible (dendrita) hacia la superficie mucosa y una vía de conducción del estímulo (axón) que termina en el bulbo olfatorio. Entrando en más detalle, la dendrita acaba en una dilatación denominada botón olfatorio del que surgen cilios en un número que oscila entre cinco y veinte, y que se introducen en el moco que cubre el epitelio. Se piensa que los olores, una vez que se han disueltos en el moco y difundidos hacia el receptor, interactúan con los cilios, produciendo los primeros cambios que darán lugar a la transducción químico-eléctrica.

Los axones de las células sensoriales cruzan el hueso etmoides para entrar en el bulbo olfatorio, que es un primer centro de procesamiento de las impresiones. Es destacable en este sentido la cercanía entre las células receptoras, similares a neuronas, y el primer centro de procesamiento de datos^{52 53}.

1.5.1.1. Electrofisiología de las células olfatorias

La estimulación de los receptores olfatorios por los olores, desencadena en la célula una serie de eventos electrofisiológicos, básicamente similares a los que ocurren en cualquier receptor sensorial. Los electro-olfatogramas (EOG), basados en el registro de las señales emitidas por el epitelio olfatorio merced a situar electrodos en el mismo epitelio, son una herramienta básica para estudiar y entender los procesos que se producen en la olfacción⁵². El EOG representa la suma de la actividad eléctrica generada en la mucosa olfatoria debido a la aplicación de un olor en la superficie ciliar. Un EOG viene definido por las siguientes características: su amplitud, que está directamente

relacionado con la concentración del olor aplicado; el periodo de latencia, que refleja el tiempo de partición de la sustancia en el moco y su difusión hacia la membrana ciliar; el decremento de la señal en función de la localización de la medida en la célula; y la duración, que depende del tiempo durante el cual se aplica el estímulo. Una aplicación prolongada (15 seg.) de un estímulo odorable de lugar a un EOG con dos componentes: una respuesta transitoria inicial y un componente tónico, de menor voltaje, que dura hasta la desaparición del estímulo. Se observa en los EOG, que una aplicación repetida de olores da lugar a una disminución de la amplitud de la respuesta transitoria inicial. Es mas que probable que este fenómeno se deba a un proceso de adaptación que se estudiará con más detalle más adelante.

Estudios de los EOG y otros estudios electrofisiológicos de las células olfativas permiten definir una serie de propiedades interesantes de las mismas:

- a) La mayoría de las células incrementa su potencial de acción con el aumento de la concentración del olor.
- b) Cada célula responde a muchas sustancias odorables y cada olor es capaz de estimular muchas células.
- c) Sin embargo, pueden llegar a distinguirse regiones del epitelio olfatorio con mayor sensibilidad a unos u otros olores.
- d) Las células sensoriales olfatorias tienen propiedades de adaptación lenta y son sensibles a la fatiga.

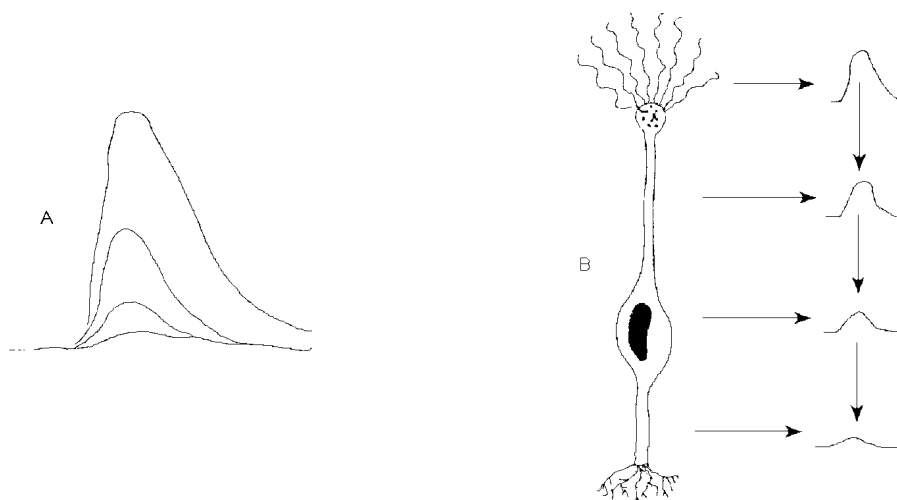


Figura 1. 26 Esquematación de la electrofisiología de las células olfativas

1.5.2. El bulbo olfatorio y las conexiones con los centros nerviosos

La integridad del bulbo olfatorio es fundamental para la capacidad olfativa. En el hombre, lesiones del bulbo producen trastornos olfatorios como la anosmia, de la que hablaré mas adelante. En palabras de Ramón y Cajal, el bulbo olfatorio es “el primer centro cerebral en el cual terminan los verdaderos nervios olfatorios. Desde el punto de vista anatómico y funcional, hay, por consiguiente, una analogía completa entre el bulbo olfatorio, la retina y los centros bulbares que sirven de estación terminal a los nervios acústicos”.

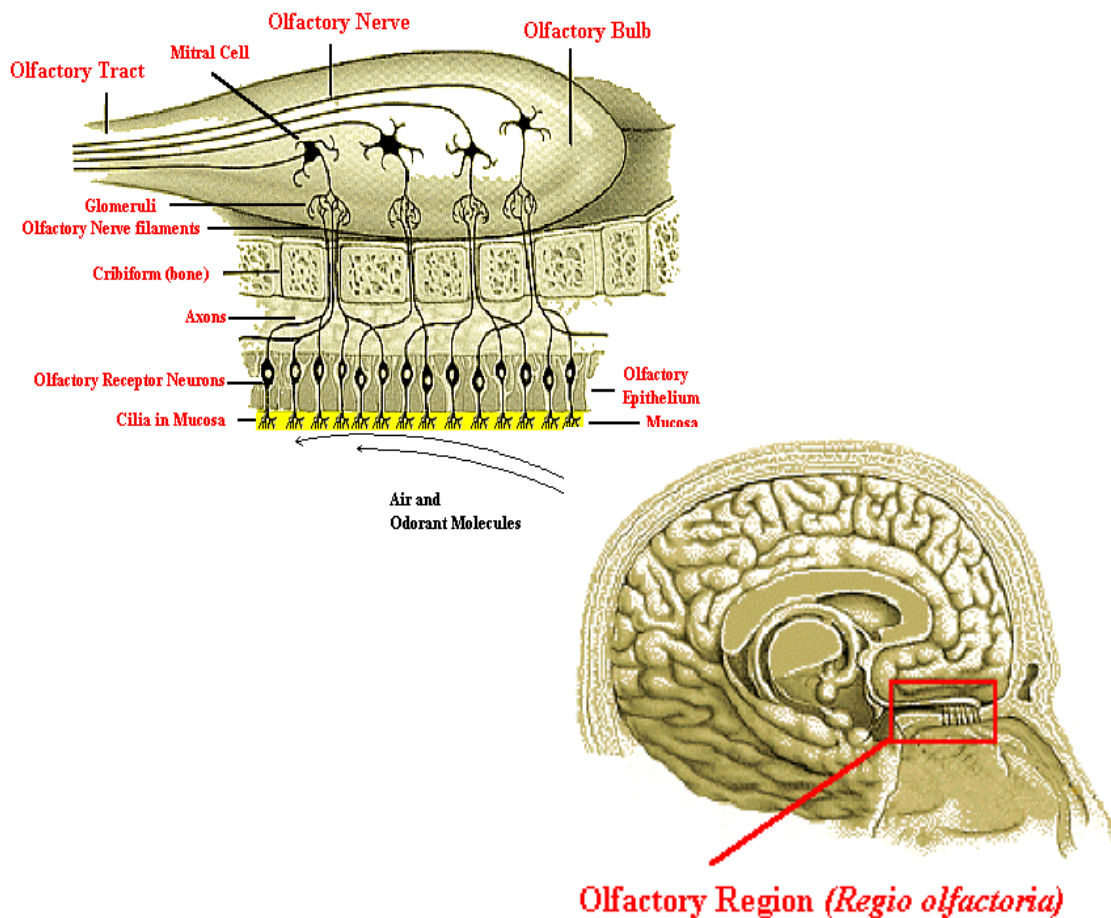


Figura 1. 27 Estructura celular del bulbo olfatorio

La estructura del bulbo está representada en la figura 1.27. Una cantidad cercana a 25.000 axones de las células sensoriales olfatorias hacen sinapsis excitatorias en el bulbo olfatorio, sobre las dendritas de unas 25 células mitrales y unas 70 células en penacho, formando el denominado glomérulo olfatorio. La convergencia de todo ello no es indiscriminada, en el sentido que distintos olores provocan grados de actividad en diferentes regiones del bulbo olfatorio. Curiosamente, este aspecto se retomará en los sistemas de reconocimiento artificiales; concretamente en la descripción de determinadas redes neuronales artificiales cuyo funcionamiento se verá más adelante.

La discriminación funcional del bulbo olfatorio indica que la primera codificación de la información odorable tiene lugar a través de las vías que distribuyen la actividad a diferentes áreas del bulbo olfatorio. Puesto que los campos receptivos del epitelio olfatorio no están distribuidos en un mapa espacial, se piensa que el bulbo olfatorio posee diferentes áreas funcionales para diferentes olores⁵².

La información recogida por las células mitrales y en penacho salen del bulbo olfatorio por los axones de estas mismas células, pero sufren una modulación (según se ve en la figura) por las llamadas células periglomerulares (que hacen sinapsis con células mitrales y en penacho de glomérulos adyacentes) y por las células granulosas (que también hacen sinapsis con las células mitrales y en penacho). Estas sinapsis son todas inhibitorias pero, por otra parte, las células del bulbo también están influenciadas por señales inhibitorias procedentes de otros sistemas fisiológicos (límbico e hipotálamo)

que se relacionan con el aprendizaje, y por señales procedentes de centros de procesamiento olfatorio posteriores, que ayudan a la localización de la fuente odorable.

Las señales ya moduladas se distribuyen en diversas zonas denominadas áreas olfatorias secundarias, donde se producen las percepciones conscientes de los olores, la discriminación de los mismos y donde se conforman las reacciones (afectivas o de conducta) al estímulo.

1.6. Trastornos y disfunciones del sistema olfativo

Para producir sensación olfativa, la sustancia olorosa debe actuar sobre los receptores con una concentración mínima y en un tiempo mínimo. Cada sustancia olorosa tiene su propio umbral y éste permanece fijo a través de los años. En la tabla 1.6 se pueden observar concentraciones umbrales de ciertas sustancias de entre las cuales destaca el caso del metilmercaptano, responsable del olor del ajo, que presenta una bajísima concentración umbral⁵⁴.

El tiempo mínimo necesario de actuación en el hombre es de unos 0.18 segundos que en comparación con otras sensaciones, es un tiempo un tanto grande. Este tiempo tiene que ver con lo que tarda en solubilizarse la sustancia olorosa en agua o lípidos de la mucosa a fin de alcanzar al receptor. La temperatura y humedad de la sustancia olorosa no importan demasiado, pues como ya se ha comentado, la propia nariz se encarga de regular estos aspectos.

Tabla 1. 6 Umbral de sensación para ciertas sustancias olorosas

Alcoholes	ppm	Sustancia	ppm
Metílico	1	éter etílico	0.0058
Etílico	0.025	Cloroformo	0.0033
Propílico	0.005-0.01	Yodoformo	0.00002
Butílico	0.001	Almizcle sintético	0.00000004
Isoamílico	0.0001	Metilmercaptano	0.0000000004

Existen multitud de factores que modifican la sensibilidad del olfato y el mismo paso del tiempo es, irremediamente, uno de ellos ya que el olfato, que se desarrolla a partir del nacimiento, se atrofia considerablemente en la vejez. La falta de percepción olfativa se denomina anosmia^{55 52} y puede ser producida por un mal funcionamiento de las células receptoras o por disfunciones de los centros de procesamiento de la percepción. Es conocida por la genética la existencia de anosmias parciales a determinados olores, en particular al ácido cianhídrico (olor a almendras), al mercaptán y al almizcle, e incluso se utilizan pruebas con estos olores para realizar investigaciones de paternidad. Si la anosmia consiste en la ausencia total de sensibilidad olfativa para ciertos olores, la hiposmia es la disminución de la sensibilidad olfativa, en contraposición a la hiperosmia, consistente en la exageración de dicha sensibilidad.

Otro factor ligado con trastornos de la olfacción es el sexo y las hormonas sexuales. Esto se demuestra en el hecho de que existe una cierta insensibilidad al olor a almendras en aproximadamente el 8 % de los hombres, mientras que para las mujeres, este porcentaje baja hasta el 4.5%. Es también muy conocido la hiperosmia de las mujeres al principio del embarazo y antes de cada menstruación y está constatado el efecto de desaparición del olor al almizcle después de la extirpación de los ovarios, recuperándose este olor por acción de los estrógenos.

Por otro lado, junto con estos trastornos, se pueden producir falsas percepciones, denominadas parosmias, y alucinaciones olfativas denominadas fantosmias. Tanto las parosmias como las fantosmias pueden ser consecuencia de rinitis alérgicas o producirse después de traumas craneales. Existe un tipo especial de fantosmia de carácter desagradable, denominada cacosmia, que son experimentadas sobre todo por personas esquizofrénicas, histéricas o epilépticas, aunque en algunos casos tienen un origen somático (sinusitis, bronquitis crónica, etc.).

1.7. Cuantificación de la sensación olfativa: Psicofisiología olfativa

Hasta el momento se ha hablado de la percepción olfativa de una manera cualitativa, y resultaría de gran utilidad encontrar una manera de cuantificar dichas sensaciones. El problema que se plantea para poder precisar cuantitativamente la percepción olfativa consiste en encontrar una función que determine la relación entre la intensidad olorosa percibida y la intensidad del estímulo, desde el olor más débil hasta el más fuerte que el sujeto pueda soportar.

El método más común para estudiar esta función es la estimación directa. Consiste en que los individuos bajo test, asocien números según el efecto que les produzca el estímulo oloroso. Los resultados son promediados para grupos, generalmente de entre diez y veinte personas. Las conclusiones de estos estudios⁵⁶ indican que la intensidad del olor percibida sigue una ley del tipo:

$$R = kS^b \quad (1.5)$$

donde R representa la intensidad olorosa percibida, S es la concentración del estímulo y k y b son constantes. Es destacable el hecho de que esta ley es semejante a la respuesta de los sensores de óxidos metálicos semiconductores.

La ecuación 1.5 ha demostrado un buen comportamiento en la descripción de la percepción olfativa en numerosas pruebas experimentales y, en consecuencia, se debe concluir que la dependencia potencial de la percepción con la concentración es una correcta descripción del sentido del olfato humano.

Sin embargo, esta ley no tiene en cuenta diferentes efectos asociados a la olfacción que distorsionan el comportamiento normal de este sentido. Tal vez el efecto distorsionante más importante es el proceso de adaptación.

El fenómeno de la adaptación al estímulo aparece en todos los sentidos, pero es muy acusado en el sentido del olfato. Cuando un individuo está sometido de manera continua a un estímulo oloroso, disminuye la sensación hasta llegar a la anosmia específica para ese estímulo, aunque persiste la sensibilidad para otros estímulos olorosos diferentes. Este efecto se experimenta con asiduidad, y la anécdota típica que lo ilustra consiste en entrar en una habitación llena de gente. En estos casos suele experimentarse una sensación olfativa fuerte, que desaparece poco tiempo después de permanecer sometido al estímulo continuamente.

Otro fenómeno interesante asociado a la adaptación es el hecho de que un individuo sometido a un proceso de adaptación a un estímulo mantiene sensibilidad a estímulos olfativos diferentes, pero el umbral de sensación para los mismos varía. Por ejemplo, después de la adaptación al olor de alcanfor, se eleva el umbral de sensación para el agua de colonia o la esencia de eucalipto.

Para estudiar como influye la intensidad del estímulo oloroso en la adaptación, se realiza el siguiente experimento: Se somete a los sujetos a tres inhalaciones de tres concentraciones del odorante antes de realizar una prueba normal del mismo odorante. Los resultados están en la figura 1.28 e indican que para las pruebas realizadas después de la adaptación con concentraciones altas, las concentraciones más bajas no son detectables. Por tanto, cuanto más alta es la concentración de adaptación, menos puntos son detectables en la prueba olfativa estándar de tal manera que las bajas concentraciones no se perciben.

Se ha comentado anteriormente la capacidad de las células sensoriales para responder a muchas sustancias odorables. Además de esta característica, se observa que ligeros cambios en la estructura de estas sustancia pueden llevar a variaciones notables en la cualidad sensorial e incluso puede llegarse a la total ineficacia de la célula. Probablemente, el tamaño de la molécula olorosa o la distribución de carga eléctrica afecta en el proceso sensitivo. Sin embargo, aun no está claro porqué moléculas muy diferentes químicamente, proporcionan las mismas sensaciones olorosas. Así, por ejemplo, producen olor a almizcle los tres compuestos representados en la figura 1.29.

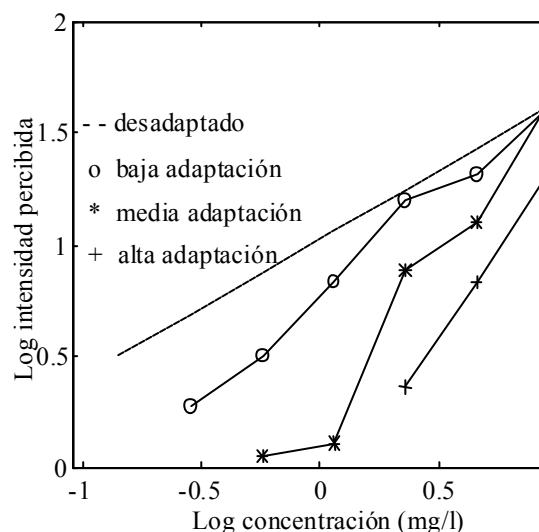


Figura 1.28 Efecto de la adaptación para pentanol con concentraciones de 0.5, 2.1 y 9.2 mg/l

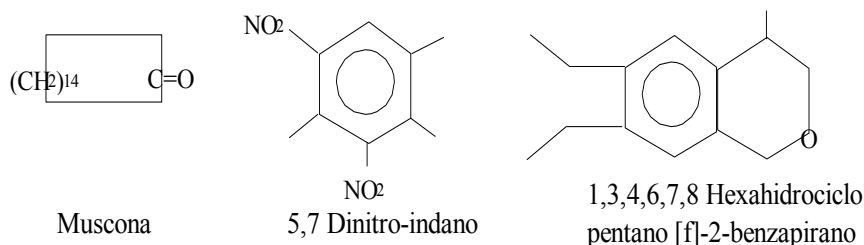


Figura 1.29 Configuración química de tres moléculas diferentes que dan la misma sensación olfativa.

Esta curiosidad olfativa puede producir adaptación cruzada es decir, adaptación para una sustancia producida por la inhalación de otra^{53 56}. Se han realizado pruebas con diferentes alcoholes que demuestran que, aunque el efecto se produce en menor grado que en la “autoadaptación” (adaptación de una sustancia para sí misma), realmente existe la adaptación cruzada para la percepción olfativa.

Todos estos fenómenos, son típicos de la detección olorosa natural, pero no son exclusivos de ella. Los sensores artificiales también sufren efectos parecidos, y en cierta forma, las derivas temporales, la variación en el nivel de cero o las interferencias cruzadas de sustancias no dejan de ser los mismos problemas de adaptación, fatiga, umbrales de detección, anósmia, hipósmia, etc., descritos hasta este punto. Es tarea del procesado de la señal, corregir estos errores de la forma adecuada. Una cuestión diferente, que se aborda brevemente en el siguiente apartado, son los aspectos

psicológicos de la percepción frente a los cuales son inmunes los procesos artificiales de detección.

1.8. Subjetividad en la percepción olfativa

De acuerdo con la teoría clásica⁵⁶, el nivel umbral de percepción juega un papel principal en la percepción olfativa. Según esta teoría, el estímulo que llega a los receptores inicia la respuesta neuronal que acaba con la percepción del olor por el cerebro. La magnitud de este efecto, varía con la intensidad del estímulo, la sensibilidad del receptor, la eficiencia de las conexiones neuronales y la actividad del cerebro. Si el estímulo tiene una cierta magnitud, este provoca una respuesta perceptiva relacionada con esta intensidad. La magnitud mínima del estímulo necesaria para producir respuesta perceptiva se denomina magnitud umbral. Los diversos factores listados anteriormente afectan al umbral instantáneo en cada experiencia olfativa. Para la teoría clásica la variable fundamental es la magnitud del estímulo y sostiene que con cuidadosos experimentos se pueden eliminar los efectos de variables no relacionadas con el estímulo.

La teoría clásica es básicamente correcta, pero es demasiado simple ya que no tiene en cuenta determinados efectos psicológicos. Uno de estos efectos, quizás el más claro, es el producido por la expectación por parte del sujeto sometido a un estímulo de que cierto olor está presente en la muestra, y la relación que esta expectación supone en el rendimiento de la detección. Existen varios experimentos que demuestran que la sugestión influye en la percepción olorosa provocando alucinaciones^{56 53}. En uno de ellos se demuestra como los sujetos sometidos a las pruebas, acuerdan que el agua destilada tiene olor desagradable. Estos efectos se deben al hecho de que el nivel umbral no es un punto claro y bien definido sino que se asemeja más bien a una región donde los efectos del estímulo varían vagamente entre lo perceptible y lo imperceptible. La teoría de detección propone que tanto estimulaciones débiles como fuertes producen un efecto perceptual continuo, pero este efecto debe ser valorado junto con el efecto del ruido y ruido de fondo que también tienen efectos perceptivos. El ruido siempre está presente, y su contribución puede venir por parte del sujeto, del ambiente, o como resultado que para diferentes experimentos estimulantes, de la misma magnitud, el número de moléculas que llegan a los centros receptores es diferente.

1.9. Percepción artificial: la nariz electrónica

A lo largo de este capítulo se ha desarrollado dos vías de detección gaseosa. Por un lado, se ha presentado el olfato como vía natural para la detección de olores y aromas, y se han desarrollado brevemente, aspectos relativos a la fisiología y a la percepción olfativa. Anteriormente se había presentado la medición de gases a través de sensores como método de detección artificial y se repasaron, también brevemente, algunos de los sensores típicos para esta aplicación.

De los argumentos expuestos, necesariamente hay que concluir que en tareas de reconocimiento cualitativas, el olfato saca ventaja a los sensores artificiales gracias a su configuración sensorial y perceptiva no encaminada hacia la especificidad sino hacia la extracción de pautas sobre la base de un aprendizaje a través de conexiones neuronales^{52 57}. Esta configuración hace que el olfato sea extremadamente eficiente en tareas de reconocimiento de olores y aromas, ya que no busca una detección cuantitativa sino un reconocimiento e identificación de los patrones característicos. Sin embargo, el olfato

tiene un rendimiento pobre a la hora de cuantificar los resultados, resulta afectado por aspectos ajenos al proceso de medida como el estado psicológico del individuo y además es incapaz de detectar la mayoría de gases (CO, O₂, H₂, etc.). La medición artificial a través de sensores es capaces de realizar tareas de medida cuantitativas de gases concretos, pero por contra no son capaces de alcanzar, por si solos, las prestaciones que el olfato proporciona en reconocimiento de olores.

Sería deseable aunar las ventajas de ambas vías de detección gaseosa en algún dispositivo de tal manera que las carencias quedasen compensadas. Este dispositivo objetivo es la nariz electrónica o nariz artificial y para alcanzarlo se combina la objetividad cuantitativa de los sensores con el eficaz tratamiento de la señal que efectúa el proceso perceptivo natural. Esta especial eficacia en el procesado de datos es fruto de la inmensa capacidad de interconexión que ofrece el sistema nervioso. Es por ello que una buena manera (aunque no la única) de imitar este tipo de procesado pasa por imitar también el comportamiento neuronal.

1.10. Breve descripción de la fisiología de una neurona

18. Txe7, Axe7; 19. C4!, y Kasparov abandona. Con estas jugadas acabó la partida en la cual Deep Blue derrotó al campeón mundial de ajedrez. Quizás a partir de este reto, ganado por la máquina, algo cambie en la concepción de la “inteligencia” de las máquinas o quizás no, pero lo que es indudable es que es en las computadoras que juegan al ajedrez donde mejor se exhibe lo que puede ser considerado como “conducta inteligente”. Con esto no se pretende reflexionar acerca de la supuesta inteligencia de un ordenador más o menos sofisticado; en cierta manera, no son más inteligentes que una máquina de calcular o que un ábaco. Es cierto, sin embargo, que poseen una increíble capacidad de cálculo, basada en algoritmos muy potentes ejecutándose sobre chips de alta velocidad. El ejemplo de los ordenadores que llegan a ser maestros de ajedrez es fruto de todo un conjunto de teorías, nacidas en entre principios y mediados del siglo XX, como la teoría de la información de Shannon, la identificación de sistemas y sistemas realimentados de Wiener y una teoría desarrollada por McCulloch y Pitts en 1943 que dio origen a lo que hoy se denomina teoría de redes neuronales. Ciertamente el nombre de esta teoría es paradójico ya que las redes neuronales artificiales imitan, en cierta manera, la manera de trabajar de las estructuras nerviosas de los mamíferos y por abstracción del ser humano. Es decir, que para conseguir unas estructuras capaces de aparentar inteligencia se imitan las estructuras ya creadas por la naturaleza.

Así, para comprender el funcionamiento de las redes neuronales artificiales (ANN: Artificial Neural Networks) resulta conveniente conocer aunque sea superficialmente las estructuras neuronales naturales.

Sin profundizar en descripciones fisiológicas ni anatómicas, los sistemas nerviosos animales, y cerebro humano en particular, utilizan como elemento básico de funcionamiento un conjunto de células denominadas neuronas. El descubrimiento del funcionamiento de las neuronas se debe a Ramón y Cajal a finales del siglo XIX. Se conocen al menos cinco clases diferentes de células neuronales, pero para comprender su funcionamiento no es necesario analizar cada una de las diferentes clases; es suficiente describir un esquema general más simple representado en la figura 1.30.

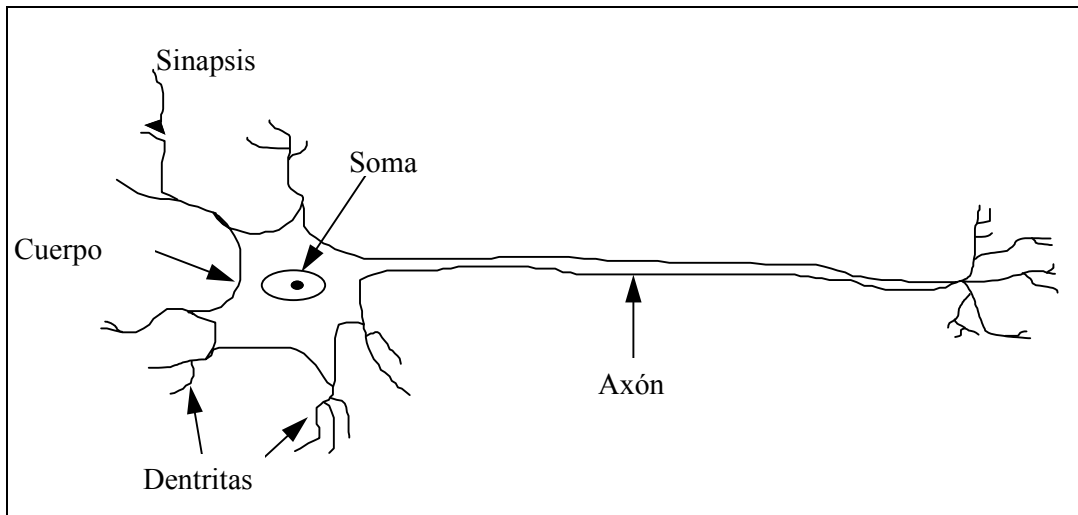


Figura 1. 30 Esquema simplificado de una neurona

Una neurona típica consiste en un cuerpo central o soma del que parten dos tipos de extensiones: las dendritas y el axón.

Las dendritas reciben señales portadoras de información mediante ciertas conexiones (denominadas sinapsis) con otras células. Estas señales se envían a través del soma, hacia el axón, que a su vez, finaliza en unas ramificaciones coronadas por unos pequeños bulbos que permiten realizar nuevas sinapsis con otras neuronas.

Las señales que llegan a las sinapsis son impulsos eléctricos y la transmisión se realiza mediante un complejo proceso químico basado en ciertas sustancias que reciben el nombre de neurotransmisores. Estos neurotransmisores, según se esquematiza en la figura 1.31, cruzan el gap sináptico y atraviesan la membrana que rodea a las dendritas, introduciéndose en una segunda neurona. Cada neurona tiene un potencial de activación, que el neurotransmisor es capaz de modificar, elevándolo o disminuyéndolo. La neurona se activa si su potencial eléctrico alcanza un cierto umbral, en ese caso, en la dendrita, el neurotransmisor genera una nueva señal eléctrica, un pulso lo suficientemente fuerte y de la duración adecuada como para atravesar el axón, llegar a las sinapsis e iniciar el proceso para traspasar la información a una tercera neurona. Una vez realizado todo el ciclo la neurona entra en un periodo de relajación durante el cual no puede volver a ser activada.

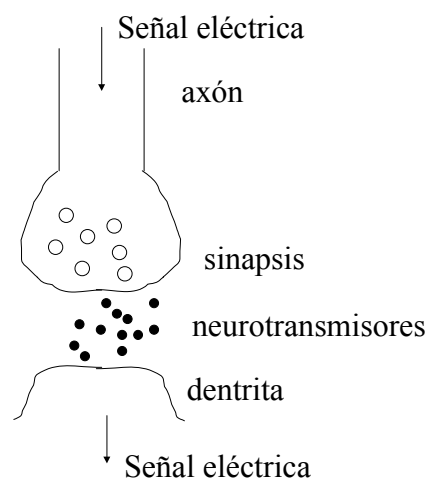


Figura 1. 31 Representación esquemática de una sinapsis

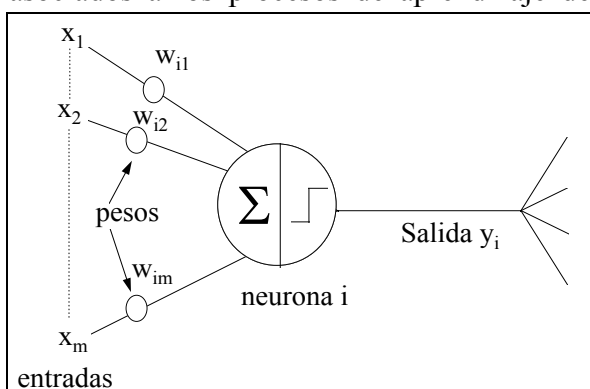
Se dice que las sinapsis son excitatorias si permiten el paso de impulsos que causen la activación de la siguiente neurona. Por contra, son sinapsis inhibitorias las que permiten el paso de impulsos que dificultan la activación de la siguiente neurona. Es de destacar el hecho de que las sinapsis son de un solo sentido, es decir que la transmisión de información funciona de sinapsis a dendritas, pero no al revés.

Curiosamente, este complejo y altamente sofisticado y proceso de transmisión de información no es excesivamente rápido si lo comparamos con la velocidad de transmisión de datos de los circuitos basados en silicio. Mientras que en una puerta lógica digital basada en tecnología de silicio, dicha velocidad es del orden de nanosegundos, en las neuronas es de milisegundos. Sin embargo, esta relativamente baja velocidad se ve ampliamente compensada por el increíblemente alto número de neuronas y su asombrosa capacidad de interconectarse entre ellas. Hay, aproximadamente 10^4 sinapsis por neurona en un cerebro humano, y hay que tener en cuenta que el sistema nervioso humano contiene unas 10^{11} neuronas. El resultado global es una herramienta de procesado altamente eficiente tanto funcional como energéticamente⁵⁸.

El cerebro humano es pues, desde el punto de vista de un sistema de procesado de datos, una herramienta altamente compleja, no lineal y capaz de realizar computación en paralelo. Es más rápido en tareas como reconocimiento, percepción y control que cualquier computador que pueda existir hoy en día. Uno de los secretos de tan alto rendimiento, aparte de la ya comentada increíble capacidad de interconexión neuronal, consiste en la aptitud del cerebro para acumular experiencia. El cerebro acumula experiencia cerrando y abriendo sinapsis concretas, y lo hace a lo largo de toda la vida, aunque de hecho es en las primeras etapas del crecimiento cuando más se desarrolla este ejercicio de conectividad neuronal.

1.11. La neurona artificial

Los primeros intentos de generar un proceso artificial que imite el comportamiento de las neuronas naturales se realizaron en la década de los 40. La figura 1.32 representa una neurona artificial, realizando las operaciones según el modelo biológico. Esta neurona fue propuesta por McCulloch y Pitts en 1943, y posteriormente, con los algoritmos asociados a los procesos de aprendizaje desarrollados por Rosenblatt entre los años



1958 y 1962, desembocó en lo que hoy en día se conoce como perceptrón⁵⁸.

En este modelo la neurona recibe las entradas x_j que contienen algún tipo de información. A su vez, estas entradas están compensadas mediante unos pesos w_{ij} . La neurona procesa estos datos en dos pasos. El primer consiste en realizar la suma de las entradas x_j ponderadas con los pesos w_{ij} . El resultado se aplica, en el segundo paso, a una función escalón que ofrece una salida $y_i=1$ o $y_i=0$, según si la suma de las entradas

Figura 1. 32 Esquema de la neurona McCulloch-Pitts

pesadas está por encima o por debajo de un valor umbral θ .

$$y_i(n) = F\left(\sum_{j=1}^m w_{ij}x_j(n) - \theta_i\right) \quad (1.6)$$

Donde n indica un índice en la secuencia temporal y $F(\cdot)$ es la función escalón:

$$F(f) = \begin{cases} 1 & \text{para } f \geq 0 \\ 0 & \text{para } f < 0 \end{cases}$$

En esta aproximación matemática al proceso biológico de transmisión neuronal, los pesos w_{ij} representan la fuerza de la conexión sináptica. Un peso positivo indicara una sinapsis excitatoria mientras que un peso negativo viene a representar una sinapsis inhibitoria. Si acaso el peso adoptase un valor nulo, indicaría la ausencia total de sinapsis para esa entrada entre las dos neuronas.

A pesar de la aparente simplicidad de este modelo matemático, una neurona así establecida y con los pesos y valor umbral adecuados puede simular funciones de alto nivel como por ejemplo las funciones lógicas NOT, OR o AND; y a partir de ellas, otras funciones lógicas más complejas. Sin embargo, la neurona de McCulloch-Pitts aunque contiene la esencia del procesado neuronal, representa sólo el primer modelo de neurona que se ideó. En la figura 1.33 aparece un modelo más general para una neurona y aun siendo efectivamente más general, no puede representar a todos los tipos de neuronas artificiales existentes, que son muchas y muy variadas.

En el esquema es posible diferenciar claramente tres zonas:

- Un conjunto de sinapsis, caracterizadas por los pesos w_{ij} .
- Un sumador o combinación lineal de las entradas y los pesos
- Una función de activación que limita la salida de la neurona.

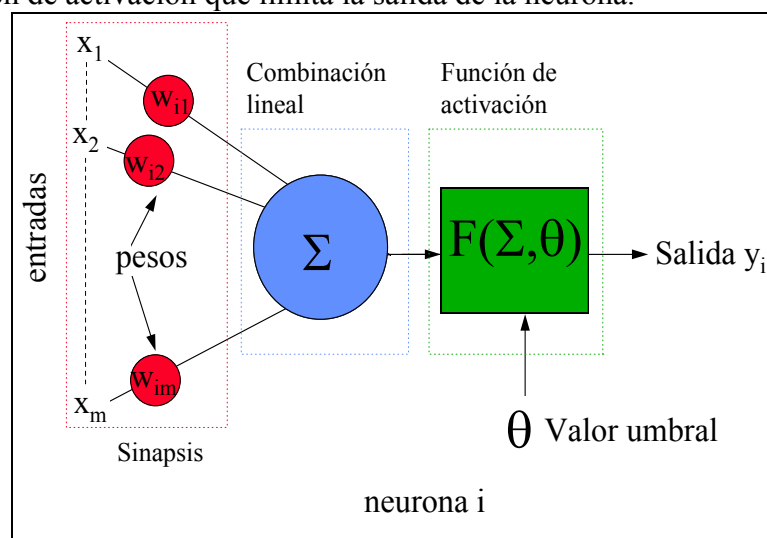


Figura 1.33 Generalización de la neurona de McCulloch-Pitts

El proceso que sigue la información en una neurona es el siguiente: a la neurona llegan una serie de señales x_j que quedan ponderadas a través de los pesos w_{ij} . El conjunto de las entradas ponderadas se combina, generalmente de forma lineal, para dar como resultado un valor v_i según la ecuación 1.7.

$$v_i = \sum_{j=1}^m w_{ij} X_j \quad (1.7)$$

La neurona inicia a continuación el segundo proceso donde se aplica una función de activación o función de transferencia que se encarga de transformar la señal formada por v_i menos el valor umbral θ_i , limitando la salida y_i de la neurona. Típicamente esta salida se limita, a través de esta función de activación o transferencia F , a un intervalo normalizado que puede ser $[0,1]$ o $[-1,1]$, dependiendo de la utilización de la red.

$$y_i = F(v_i - \theta_i) \quad (1.8)$$

No hay una norma rígida para el uso de funciones de activación para las neuronas, sin embargo, algunas funciones típicas son la función escalón, función rampa y la función sigmoide.

La función sigmoide no es una función concreta sino que engloba aquellas funciones que cumplen las características de ser no lineal, creciente, suave y asintótica. Ejemplos de función sigmoide son la función logística:

$$F(n) = 1/[1 + \exp(-n)] \quad (1.9)$$

o la función tangente hiperbólica:

$$F(n) = [1 - \exp(-n)]/[1 + \exp(-n)] \quad (1.10)$$

Siguiendo con la analogía entre las neuronas naturales y las artificiales, cabe remarcar, tal y como ya se ha comentado en el modelo de Mc Culloch-Pitts, que los pesos sinápticos w_{ij} de una neurona tratan de reproducir las conexiones sinápticas entre las neuronas naturales; por otro lado, el valor de threshold tiene que ver con el potencial de activación de las neuronas. Para acabar con la analogía, la función de transferencia está, evidentemente, relacionada con la forma en la cual las neuronas naturales transmiten la información. Las diferentes funciones intentan cubrir de la manera más simple posible el abanico de posibilidades de cálculo que las neuronas naturales abarcan, desde clasificación lineal con la función escalón hasta el procesado no lineal con la función sigmoide.

1.12. Redes neuronales artificiales: Arquitectura

Por supuesto que una sola neurona, ni en el reino animal ni en el mundo de los algoritmos neuronales, es suficiente para realizar procesos de alto nivel. Las neuronas, por si solas, constituyen los elementos de procesamiento de una estructura más general y compleja denominada red neuronal y que se conforma, precisamente, con un cierto número variable de neuronas interconectadas entre sí según una cierta arquitectura.

Comenzando desde una arquitectura sencilla y avanzando en complejidad, podemos encontrar las siguientes arquitecturas:

Redes de una sola capa.

Este tipo de arquitectura organiza las neuronas de la red neuronal en una sola capa. Las entradas se proyectan hacia una capa de neuronas, y no hay retorno en el otro sentido. A

este tipo de conexión en sentido entrada salida se denomina *feedforward*, y por defecto las redes así conectadas se denominan también redes feedforward. La figura 1.34 representa una red feedforward de una sola capa.

Aunque en la figura se representan cuatro entradas y cuatro salidas, la arquitectura no exige, ni esta restringida a tener un mismo número de entradas que de salidas.

Un sistema de memoria asociativa lineal es un ejemplo de red neuronal feedforward de una sola capa.

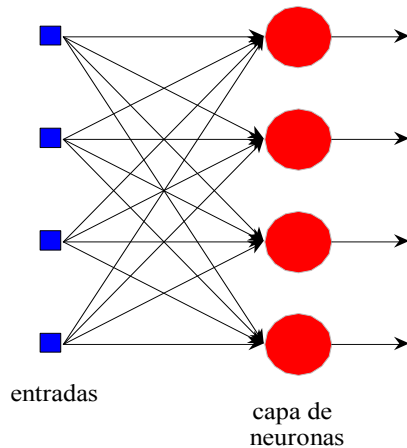


Figura 1. 34 Red de una sola capa

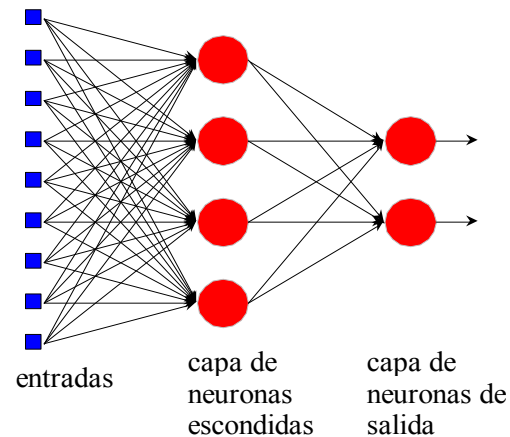


Figura 1. 35 Red feedforward multicapa

Arquitectura multicapa feedforward

En estas arquitecturas existe más de una capa de neuronas. Comenzando por el final de la estructura, la capa de neuronas que se corresponde con la salida de la red recibe como entrada, las salidas de una capa anterior de neuronas, que a su vez, pueden recibir como entradas, las salidas de sucesivas capas de neuronas, hasta remontarse a las señales de entrada. Las capas de neuronas que no son la capa de salida se denominan capas escondidas. La función de estas capas es intervenir entre la entrada y la salida, consiguiendo, en cierta forma una visión más global del problema y permitiendo, en principio, extraer información de orden mayor, a costa claro está de una mayor complejidad. La figura 1.35 representa una red neuronal multicapa totalmente conectada, es decir en la figura todas las neuronas están conectadas con todas. Esta no tiene porque ser y de hecho no siempre es la mejor estrategia. Es posible construir una arquitectura neuronal en la no todas las posibles conexiones estén realizadas y, a pesar de ello, ofrezca mejores prestaciones. A este tipo de arquitectura con no todas las conexiones habilitadas se la denomina arquitectura parcialmente conectada⁵⁸.

La red típica de esta arquitectura es la denominada red de perceptrones multicapa. Cabe destacar como aplicación importante de esta arquitectura el hecho de que es un aproximador universal (con el grado de aproximación que se desee) para cualquier función continua no lineal⁶¹. Para ello sólo es necesario disponer del número suficiente de neuronas en una única capa escondida en la estructura. El problema es que no se conoce, a priori, el número de neuronas que se van a necesitar para un cierto grado de aproximación, y es, por tanto, necesario ir realizando, paso a paso, un proceso constructivo.^{59 60}

Redes recurrentes

En el caso de las arquitecturas anteriores, la información fluye siempre en un determinado sentido, sin realimentaciones. De esa manera, las entradas de una capa de

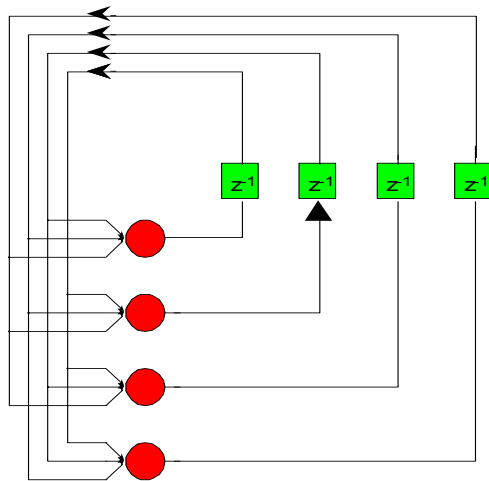


Figura 1. 36 Red recurrente

neuronas sólo pueden proceder de las salidas de una capa anterior. Cuando en una red se permite reenviar las salidas hacia atrás a la propia entrada o a entradas de capas anteriores, se dice entonces que las redes son *feedback*. Cuando la realimentación se realiza en bucle cerrado se obtienen las llamadas redes recurrentes. Un ejemplo de redes recurrentes se muestra en la figura 1.36, consistente en una capa de neuronas que realimentan su salida hacia la entrada, aunque en este caso no es autorealimentada, es decir la salida de la neurona no se realimenta hacia la propia entrada. La presencia de realimentaciones incide directamente en las capacidades de aprendizaje y en el rendimiento de la red ya que, de alguna forma, la red tiene “memoria”. De hecho, una red de este estilo puede llegar a emular a cualquier autómata determinístico con un número de estado finito. Pertenecen a este tipo de arquitecturas las redes de Elman y Jordan (que no son mas que redes de perceptrones multicapa realimentadas).

1.13. Redes neuronales artificiales: Entrenamiento

La principal característica de las redes no radica en su estructura, sino en su capacidad para aprender. El objetivo de una red neuronal, tanto natural como artificial, es aprender; es decir obtener información de la realidad circundante y almacenarla, generando modelos útiles para interpretar dicha realidad y/o realizar actuaciones concretas. Puede ser que el término aprender no parezca adecuado para un proceso matemático, sin embargo, en un sentido amplio, aprender es un proceso automático por el cual una serie de parámetros del proceso, cambian en virtud de los datos que recibe. No importa, en este sentido si la entidad receptora de la información es un ser vivo o un algoritmo matemático. Sólo resulta necesario que disponga de la capacidad suficiente para realizar el proceso, aunque esto no implica que el ente sea “inteligente”, en el sentido de tener conciencia de sí mismo y del aprendizaje que está realizando.

El proceso de adquisición de conocimiento para una red neuronal se conoce con el nombre de entrenamiento y se basa tanto en una buena elección de la arquitectura de la red como en su capacidad para procesar la información. En función del entrenamiento las redes neuronales son capaces de imitar, más o menos sofisticadamente, los procesos de aprendizaje de las estructuras neuronales naturales, desarrolladas a través de una evolución de siglos es decir, aprenden de la experiencia y mejoran las actitudes y comportamientos a partir de ella. Hablando de forma general, un entrenamiento permite la modificación de la estructura de la red (fundamentalmente, variación de los pesos w_{ij}) como consecuencia de la historia de estímulos proporcionados a la red y sus respuestas. Específicamente, en el proceso de entrenamiento se producen tres fases:

1. Estímulo de la red: A la red se le presentan los datos entrada-salida.

2. Modificación de los parámetros: Las diferentes neuronas de red modifican los valores de sus pesos y valor umbral en virtud del estímulo

3. Modificación de la respuesta: La red neuronal responde de forma diferente debido a los cambios producidos en las neuronas gracias al estímulo.

Al acabar el proceso de entrenamiento, la red presenta una estructura de pesos y valores umbrales que conforma un modelo de respuesta a los estímulos. En cierto modo, los propios datos van conformando el modelo, ya que los pesos se van modificando de forma iterativa durante el entrenamiento. Es decir, en el momento temporal n , se aplica un ajuste $\Delta w_{ij}(n)$ al peso $w_{ij}(n)$ para obtener $w_{ij}(n+1)$.

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n) \quad (1.11)$$

y el ajuste $\Delta w_{ij}(n)$ dependerá, de alguna manera, del estímulo que reciba la red.

Así, el entrenamiento consiste meramente en un algoritmo, más o menos sofisticado, de modificación de pesos y valores umbrales en función de los estímulos de la red hasta obtener un resultado conveniente. Sin embargo, del proceso de entrenamiento y aprendizaje de una red hay que dos conceptos elementales pero a menudo confusos: el tipo de aprendizaje y la regla o algoritmo de aprendizaje.

El tipo de aprendizaje es la forma general o la filosofía del entrenamiento, mientras que el algoritmo de aprendizaje está constituido por un conjunto de reglas, bien establecidas, que permiten la resolución de la ecuación (justo la anterior)

De entre los distintos tipos de aprendizaje destacan dos: el aprendizaje supervisado y el aprendizaje no supervisado²¹.

En el proceso de aprendizaje supervisado, para cada instante del proceso, cuando a la red se le aplica la entrada, también se le ofrece la salida correspondiente que se desea obtener. De esta manera, a la red se le dice precisamente el tipo de salida que debe emitir. Más claramente, a la red se le proporcionan secuencias de datos $[x(1), y(1)]$, $[x(2), y(2)]$, ..., $[x(k), y(k)]$, ..., donde $x(k)$ es la entrada en el momento k e $y(k)$ es la salida en el momento k que se desea obtener. En virtud de la regla de aprendizaje y de la estructura de la red, esta proporciona una respuesta $\hat{y}(1)$, $\hat{y}(2)$, ..., $\hat{y}(k)$, ... a las entradas dadas. La diferencia entre la respuesta $\hat{y}(k)$ y la respuesta deseada $y(k)$ es el error cometido, que produce una señal de error en la red para que esta modifique todos sus parámetros de tal manera que la salida se aproxime a la salida deseada.

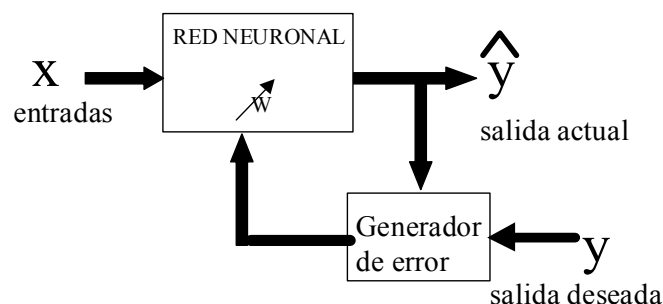


Figura 1. 37 Esquema de algoritmo supervisado

En el aprendizaje no supervisado no existe ninguna guía con la que, a través de la realimentación, indicar a la red que “camino” seguir. No existen realimentaciones indicando si la salida es correcta o no. La red debe descubrir por sí misma los patrones, regularidades, hechos, correlaciones o categorías de las señales de entrada y codificarlas

en la salida. Mientras la red descubre todo esto, ella misma autorrealiza los cambios en sus parámetros que sean necesarios. A este proceso se le llama auto-organización (“self-organizing”). El ejemplo típico de aprendizaje no supervisado es el de clasificación de objetos en ausencia de cualquier información acerca de las clases. Los propios grupos (“clusters”) se forman gracias al descubrimiento de similitudes o desavenencias entre los objetos.

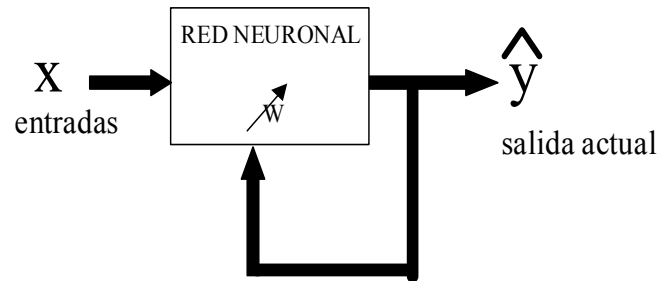


Figura 1. 38 Aprendizaje no supervisado

En cuanto las reglas de aprendizaje, hay de muy diversas clases: Algoritmos Hebbianos, aprendizaje competitivo, aprendizaje de Kohonen, regla de Widrow-Hoff, Backpropagation, etc. Unas son típicas de los tipos de entrenamiento supervisado (las dos últimas) y otras son características de los no supervisados (las tres primeras), pero todas proporcionan una fórmula fundamental de cambio de pesos. Por el interés particular en el presente trabajo, se presentan brevemente los aprendizajes de Kohonen y la regla backpropagation.

El aprendizaje de Kohonen es un tipo de aprendizaje denominado competitivo que consiste en reforzar el vector de pesos que más se “parezcan” a las entradas. El grado de “parecido” se calcula en este caso en función de la distancia euclidea existente entre el vector de pesos y el vector de entradas. Cuanto más pequeña sea esa distancia más “parecidos” son los pesos a las entradas. La ecuación de modificación de los pesos viene marcada por la siguiente expresión:

$$w_j(n+1) = w_j(n) + \eta(n)[x(n) - w_j(n)] \quad (1. 12)$$

El factor η es un parámetro que determina la velocidad del aprendizaje. De alguna forma, la modificación tiene que ver con la distancia entre $x(n)$ y $w(n)$

La regla backpropagation es una generalización de una regla más sencilla denominada regla delta, LSM (Least Mean Square) o Windrow-Hoff. Esta última se basa en minimizar el error entre el vector de salida de la red y el vector de salida deseado. La minimización se realiza buscando el camino de máximo descenso del gradiente. La regla backpropagation trata de adaptar este método para una arquitectura multicapa con redes que dispongan de funciones de activación no lineales. En el algoritmo backpropagation se han de distinguir dos procesos bien diferenciados, uno hacia delante (forward propagation) donde el flujo de los datos de entrenamiento x se propaga hacia delante por las capas de la red dirigiéndose hacia la salida, donde se produce una señal \hat{y} ; y otro hacia atrás (back propagation) donde el error, definido como la diferencia entre esta señal \hat{y} y la salida real y , se propaga hacia atrás, desde la capa de salida hacia las capas anteriores, para actualizar los pesos respectivos.

$$w_{ij}(n+1) = w_{ij}(n) + \eta f(e_i(n)) x_j \quad (1. 13)$$

donde η determina la velocidad de aprendizaje, x_j representa la entrada de la neurona j -ésima y $f(e_i)$ representa una función dependiente del error.

Con esta breve descripción genérica de lo que es una red neuronal y su estructura es suficiente para comprender los resultados que se presentarán más adelante. Para ahondar más en el tema se puede acudir a la bibliografía específica⁵⁸.

Ahora bien, en este punto del capítulo, ya se han repasado las dos vías de detección gaseosa: la artificial, a través de sensores, y la natural, a través del olfato. Además, se ha descrito una entidad matemática artificial, capaz de imitar los procesos naturales de las neuronas (y en cierta forma, por tanto, los procesos que llevan a la percepción), obteniéndose así una herramienta de trabajo capaz de aunar ambas vías. Es hora por tanto de iniciar el proceso de descripción del dispositivo que ha de ser capaz de llegar a realizar mediciones gaseosas inteligentes tanto de forma cuantitativa como cualitativa: la nariz electrónica.

1.14. La nariz electrónica

Con los apartados anteriores relativos a las redes neuronales artificiales, se ha introducido una herramienta matemática capaz de tratar datos de forma muy parecida a como lo hacen las neuronas biológicas. Continuando con este interés, expuesto varias veces a lo largo del capítulo y enfatizado en el apartado 1.9, de imitar artificialmente procesos naturales, se pretende en este apartado aunar los conceptos que se han ido exponiendo a través de la nariz electrónica. Efectivamente, con la nariz electrónica, se cierra de hecho el organigrama del capítulo representado en la figura 1.1, y a través de ella se establece la conexión entre la medición cuantitativa basada en sensores y la medición cualitativa basada en la percepción que tantas veces se ha reclamado a lo largo del capítulo. La noción de aunar la metodología analítica y la percepción sensorial no es nueva, y de hecho en algunas industrias que tienen relación con aromas, perfumes y pueden encontrarse cromatógrafos de gases cuyas medidas se complementan con mediciones sensoriales, convirtiendo la técnica de análisis en un método analítico-sensorial, son los llamados GC-Sniffing o técnica de olfatometría-cromatográfica de gases (CG-O)⁶².

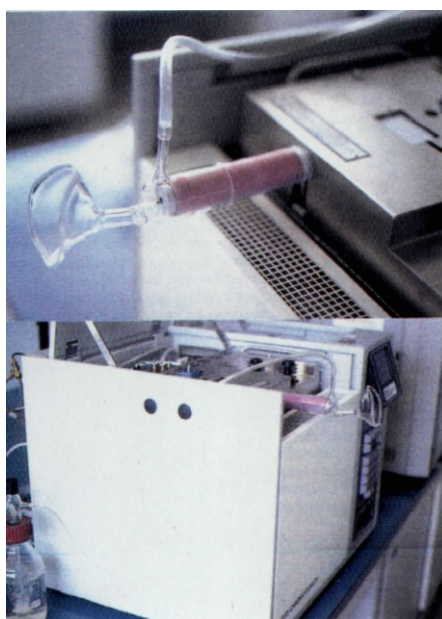


Figura 1.39 Cromatógrafo de gases con un sniffer para evaluación olfativa incorporado

En la técnica CG-O la mezcla de sustancias se separa en la columna cromatográfica, y seguidamente la corriente portadora de gas se divide al final de la columna y, simultáneamente, se registra mediante algún tipo de aparato detector (generalmente un espectrómetro de masas o un detector de ionización por llama) y se valora sensorialmente, oliéndola en aire humedecido. Esta técnica permite obtener una primera impresión rápida de la aportación olorosa de los componentes individuales al aroma total. Sin embargo, no es esta la idea que se pretende de la nariz electrónica donde de hecho no se permite la intervención del sentido olfativo sino que este ha de estar imitado por el conjunto de los sensores y el tratamiento de señal posterior (ya sea con redes neuronales artificiales o no).

Los primeros trabajos de desarrollo de un aparato dedicado a detectar olores datan del año 1961 y fueron llevados a cabo por Moncrief,⁶⁶ pero en realidad este aparato es una nariz mecánica. Las primeras propuestas de nariz electrónica fueron realizadas por Wilkens y Hatman en 1964⁶⁷, por Buck, Allen y Dalton⁶⁸ y por Draviex y Trotter en 1965⁶⁹. Sin embargo, posiblemente, el concepto de nariz electrónica como una matriz inteligente de sensores para la clasificación de olores nació en los últimos 20 años, a raíz de los estudios y publicaciones de Persaud y Dodd⁵⁷ en la Universidad de Warwick en Gran Bretaña. El término “nariz electrónica” nace a finales de 80 en el 8th Int. Congress of European Chemoreception Research Organization. En 1989 una sesión en la NATO Advanced Workshop on Chemosensory Information Processing estuvo dedicada a olfacción artificial. Finalmente, el primer congreso dedicado a narices electrónicas se llevó a cabo en 1990.

Retrocediendo de nuevo a los primeros modelos de nariz para la discriminación de olores descritos por Dodd y Persaud⁵⁷, se demostraba en dichos modelos que un sistema con tres sensores es suficiente para imitar de una manera grosera la capacidad de discriminación del sistema olfatorio de los mamíferos. Es en principio sorprendente que con tan pocos sensores sea posible llevar a cabo una tarea como esta. Sin embargo, si se analiza la capacidad que tiene el sistema de tres sensores para responder a diferentes patrones se demuestra que es perfectamente posible. Si la respuesta de los sensores es digitalizada por un convertidor A/D de 12 bits, entonces, tres sensores pueden generar 2^{41} patrones diferentes, lo que supone miles de millones de patrones ($\sim 6.8 \cdot 10^{11}$). Esta capacidad es más que suficiente para las aplicaciones de las narices electrónicas. Sin embargo, en la práctica esta capacidad es inalcanzable, como se expone posteriormente.

Una de las primeras descripciones de lo que es una nariz electrónica la dio Gardner empleando la siguiente definición:

“Una nariz electrónica es un instrumento que consta de una matriz de sensores químicos con especificidad parcial y un sistema apropiado de reconocimiento de patrones, capaz de reconocer olores simples o complejos”^{64 25}.

Sobre la base de esta definición, una nariz electrónica no sólo necesita un conjunto de sensores bien escogidos, con la especificidad y sensibilidad adecuadas, sino que también lleva asociadas las herramientas que permitan una adecuada discriminación de los olores. Hay que precisar la distinción entre análisis cuantitativo y el reconocimiento cualitativo. El reconocimiento cualitativo es el apropiado para imitar las características del olfato, pero no se debe olvidar que también resulta necesario para determinadas aplicaciones industriales disponer de sistemas capaces de realizar análisis cuantitativos. Aun y cuando el análisis cuantitativo no es una propiedad del olfato, se debe englobar este tipo de sistemas dentro de la definición de nariz electrónica, ya que en esencia consisten también en una matriz de sensores con especificidad parcial y un algoritmo de

tratamiento de señal, que aunque no es de reconocimiento, si que ofrece información (en este caso, cuantitativa) acerca de las muestras gaseosas analizadas. Así, una definición válida en este contexto es:

“Una nariz electrónica es un instrumento que consta de una matriz de sensores químicos con especificidad parcial y un sistema apropiado de tratamiento de la señal para la determinación cuantitativa y/o para el reconocimiento de patrones, capaz de ofrecer información cuantitativa de muestras gaseosas y/o reconocer olores simples o complejos”

La figura 1.40 pone de manifiesto, la analogía entre los sistemas naturales y los artificiales, dejando patente que las diferencias son sólo instrumentales pero no son, en absoluto, diferencias de fondo. Así, donde en los procesos naturales hay células receptoras olfativas, dispuestas, tal como se ha explicado, no para funcionar de forma específica, la nariz artificial incorpora matrices de sensores con especificidad parcial. El procesamiento de señal matemático es análogo en cuanto a funcionalidad al procesamiento de señal que realiza el bulbo olfatorio. El trabajo realizado por el bulbo, es transportado al cerebro donde se realiza la supervisión de todo el proceso, obteniéndose la percepción de la sensación, mientras que en la nariz electrónica existe un proceso capacitado para realizar el reconocimiento a partir de los datos ofrecidos por el sistema de tratamiento de la señal.

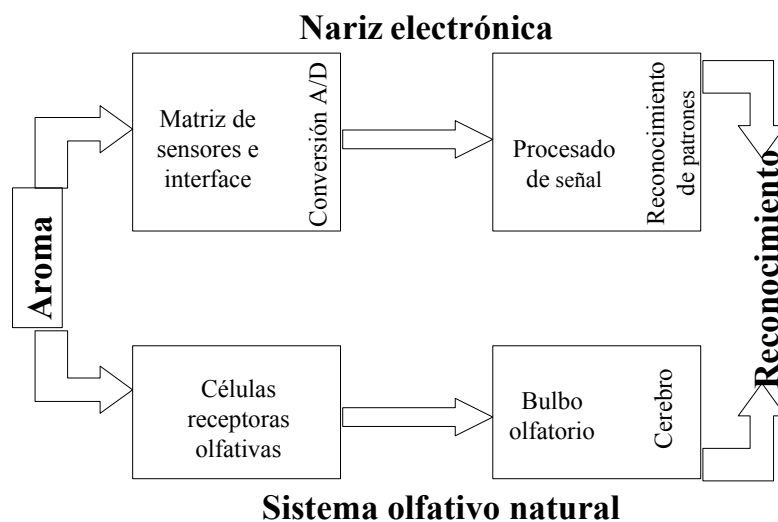


Figura 1.40 Comparación del funcionamiento de una nariz electrónica y el sentido del olfato

La matriz de sensores, que convierte la información química de entrada en una señal eléctrica, es sometida a la presencia de la sustancia gaseosa j . Cada sensor i de una nariz electrónica produce una señal eléctrica dependiente del tiempo $x_i(t)$, en respuesta a una sustancia gaseosa j . La respuesta de una matriz de n sensores es un vector \mathbf{X} , que se puede describir como:

$$\mathbf{X} = \{x_{1j}, x_{2j}, \dots, x_{nj}\} \quad (1.14)$$

En los capítulos sucesivos se desarrollarán más ampliamente los aspectos relativos a la adquisición de esta respuesta y el tratamiento matemático a aplicar a los datos para las aplicaciones de análisis cuantitativo y cualitativo.

En referencia a la matriz en sí, existen multitud de sensores desarrollados para la detección gaseosa. Algunos de los diferentes tipos de sensores ya han sido descritos anteriormente. Sin embargo, los requerimientos de los sensores de una nariz electrónica exigen que estos tengan especificidad parcial, es decir que respondan en general a un rango de gases en vez de responder específicamente a un gas. Esto es, por supuesto, lo contrario de lo que se le exigiría a un sensor de gas ideal. Sin embargo, una nariz electrónica, al igual que el olfato humano, saca ventaja, a través del procesado de las señales, de esta falta de especificidad. A esta ventaja hay que añadir otras de las cuales el olfato carece; por ejemplo, inmunidad frente a los aspectos psicológicos o fisiológicos de todo proceso olfativo natural, o la potencia analítica que se pueda extraer a través del procesado de señal.

En cuanto al tratamiento de la información, dependiendo del tipo de reconocimiento, se utilizan diferentes técnicas. Las principales están listadas en la tabla 1.7.

Tabla 1. 7 Algunas técnicas de tratamiento de señal en sistemas sensores de gas

Análisis cuantitativo	Reconocimiento
Regresión multivariante CLS, PCR, PLS, RR...	Clasificador bayesiano Funciones discriminantes
Aproximadores universales ANN, FIS, Splines...	K-Nearest neighbours Edited K-NN (VQ, LVQ, SOM)
Reducción de dimensionalidad PCA LDA.	ANN HCM, FCM...

1.15. Limitaciones de las narices electrónicas

El número teórico de patrones que una nariz electrónica puede discriminar es s^n , y que viene dado por el número de sensores de la matriz (n) y la cantidad de niveles permitidos en la salida del sensor (s). En la práctica, el valor teórico, (anteriormente se calculó un valor de $6.8 \cdot 10^{11}$ para tres sensores) es inalcanzable, y solo se puede discriminar con fiabilidad una pequeña fracción de patrones. Esto es debido principalmente a problemas achacables exclusivamente a los sensores. La falta de repetitividad, el problema de las derivas, el envenenamiento del material sensible o el ruido son dificultades ineludibles de disminuyen el rendimiento del sistema ideal. Además, otra fuente de restricción para problemas concretos es la falta de sensibilidad: si un sensor no responde a uno de los componentes gaseosos de una mezcla bajo test, la efectividad del sistema se reduce de n a $n-1$. Todos estos problemas son con los que el procesado de señal ha de enfrentarse y hacen necesario el uso sensores adicionales que compensen los errores del sistema⁶⁵.

En todo el conjunto del sistema, algunos parámetros, que juegan un papel importante como origen de problemas, pueden ser controlados. Por ejemplo los sensores de gas, en general, presentan fuertes dependencias con la temperatura y la humedad. En la nariz humana, estos parámetros están controlados por el riego sanguíneo y por la mucosidad pero en el caso de los sensores artificiales, es necesario incluir un control específico de los mencionados parámetros. Ciertamente es factible hacerlo y de hecho algunas narices electrónicas del mercado utilizan cámaras de sensores termostatadas para evitar problemas con la temperatura.. Sin embargo, la poca comprensión de los procesos que ocurren en los sensores hace que buena parte de los problemas no puedan ser, por ahora, solventados. A falta de sensores mejores, los algoritmos de procesado de señal

adquieren una importancia comparable a la de los propios sensores, ya que no sólo han de realizar la tarea de reconocimiento sino que han de elaborar estrategias que sorteen las limitaciones de los mismos. A pesar de los esfuerzos invertidos en mejorar las prestaciones de los sensores, hoy por hoy, los algoritmos de procesado precisan de una potencia de cálculo que prácticamente sólo puede ser abordada por un microprocesador.

1.16. Narices electrónica existentes

Dispositivos electrónicos que detecten un determinado gas existen comercialmente desde hace tiempo, pero no deben considerarse como narices electrónicas sino como meras alarmas más o menos inteligentes

Existen en la actualidad algunos sistemas comerciales basados en matrices de sensores dedicados al reconocimiento gaseoso y que pueden considerarse dentro de la definición de narices electrónicas. Por supuesto, estos sistemas no son capaces de igualar el rendimiento de la nariz natural, y de hecho no pretenden ser una nariz electrónica universal, sino un sistema para aplicaciones específicas. Esto quiere decir que se desarrollan las matrices de sensores apropiadas para la aplicación específica, y que dependiendo justamente de la aplicación, las matrices son unas u otras. Algunas aplicaciones típicas de narices electrónicas y los sensores utilizados se pueden ver en la siguiente tabla⁷⁰ :

Tabla 1. 8 Algunas aplicaciones de las narices electrónicas	
Aplicación	Tipo de sensor
Calificación de aromas de café	Óxidos metálicos
Calificación de licores	Óxidos metálicos
Calificación de cervezas	Óxidos metálicos
Calidad de grano de cereal	Electroquímicos
Clasificación de la frescura de pescado	Óxidos metálicos
Clasificación de la frescura de carne	MOSFETS
Calidad de aire	Polímeros
Perfumería	Piezoeléctricos

La figura 1.41 presenta una nariz electrónica denominada Intelligent Nose de la empresa Alpha MOS.



Figura 1. 41 Nariz electrónica comercial Alpha-MOS

La tabla 1.9 lista algunos dispositivos de monitorización de olores y las narices electrónicas existentes hoy en el mercado. Información relativa a la mayoría de las narices de este listado y otras está accesible a través de internet ^{w1 w2 w3 w4 w5 w6 w7 w8 w9}
 ,w12 ,w13 ,w14 w18...

Tabla 1.9 Sistemas comerciales similares a las narices electrónicas

Empresa	Origen	Tipo de sensor	Nº de sensores	Reconocimiento de patrones
Airsense Analysis GmbH	Alemania	Óxidos metálicos	10	ANN,PCA, SPR
Alpha MOS Multi-Organoleptic Systems	Francia	Polímeros conductivos, óxidos metálicos, QMB, SAW	6-24	ANN, PCA, DFA
Aromascan PLC	Gran Bretaña	Polímeros conductivos	32	ANN
Bloodhound Sesnsor LTd	Gran Bretaña	Polímeros conductivos	14	ANN, CA, PCA
Cyrano Sciences Inc	USA	Polímeros conductivos	32	PCA
EEV Ltd Chemical Sensor Systems	Gran Bretaña	Polímeros conductivos, óxidos metálicos, QMB, SAW	8-28	ANN, DFA, PCA
Electronic Sensor Technology Inc	USA	Gas chromatography, SAW	1	Quimiometría estandar
Hewlett-Packard Co	USA	Espectrometría de masas		Quimiometría estandar
HKR-Sensorsysteme GmbH	Alemania	QMB	6	ANN, CA, DFA, PCA
Lennartz Eelectronic GmbH	Alemania	QMB, óxidos metálicos	16-40	ANN, PCA, PCR
Nordic Sesnor Technologies AB	Suecia	Óxidos metálicos, QMB, MOSFET, Infrarrojos	22	ANN, PCA
RST Rostock Raumfahrt und Umweltshatz GmbH	Alemania	Óxioid metálicos, QMB, SAW	6-10	ANN, PCA
Sawtek Inc	USA	SAW	2	-

ANN: Artificial Neural Networks,
 DFA: Discriminant function analysis,
 SPR: Statistical pattern Recognition

PCA: principal Component analysis,
 CA: Cluster analysis
 PCR: Principal components regression

Bibliografía

- ¹R. Descartes, *El discurso del método* (1637).
- ²M^a Luisa Huidobro, “Tecnologías medioambientales y planes de ahorro y eficiencia,” *Química e Industria* **43 (9)**, 641-644 (1996).
- ³Redacción, “La cumbre del clima perfila un tímido acuerdo para reducir los gases que calientan la Tierra,” *La Vanguardia* (Barcelona, 1997), pp. 27-28.
- ⁴A. Aguirre Díaz, “Contaminación atmosférica en centrales térmicas convencionales,” *Química e Industria* **44 (2)**, 100-104 (1997).
- ⁵G. Massagué Roch, “Análisis de la contaminación atmosférica en áreas urbanas,” *Química e Industria* **44(2)**, 106-111 (1997).
- ⁶J.M^a Ramos Caro, “Problemática de la calidad del aire en ambientes cerrados, Síndrome del edificio enfermo,” *Química e Industria* , 147-155 (1995).
- ⁷Karen Walker, “Select a continuous emissions monitoring system,” *Chemical Engineering Progress* , 28-34 (1996).
- ⁸ Lana Sarrate s.a. “Los analizadores de combustión ayudan a los operadores de calderas a mantener un rendimiento óptimo”. *Química e Industria* **35(5)**, 524-525 (1989).
- ⁹E. Ezama, “Aspectos económicos y técnicos para el uso de motores a gas natural en Autobuses urbanos,” *Química e Industria* , 80-82 (1998).
- ¹⁰J. M. Sanz, *La contaminación atmosférica* (1991).
- ¹¹Instituto nacional de seguridad e higiene en el trabajo, Report “Ambientes cerrados: calidad del aire”, 1989.
- ¹²Edward N. Ruddy and Leigh Ann Carroll, “Select the best VOC Control strategy,” *Chemical Engineering Progress* , 28-35 (1993).
- ¹³Instituto nacional de seguridad e higiene en el trabajo, Report: “Síndrome del edificio enfermo: factores de riesgo”, 1989.
- ¹⁴M.G. Dorman P.K. Clifford, Report: “*Test protocols for residential carbon monoxide alarms*”, 1996.
- ¹⁵ A. Perera, T.Sundic, S. Marco: “A bio-inspired non-linear digital filter to integrate carbon monoxide concentration aiming to fulfil international standards” *ISOENS* **99** 211-214
- ¹⁶J.E. Peterson, R.D. Stewart. “Predicting the carboxyhemoglobin levels resulting from carbon monoxide exposures” *J. Appl. Physiol.* **39** 633-638 (1975)
- ¹⁷ J. Samitier, J.M. López Villegas, S. Marco, L. Cámara, A. Pardo, O.Ruiz and J.R, Morante “A new method to analyse signal transients in chemical sensors” *Sensors and Actuators B*, 18-19 (1994) 308-312
- ¹⁸A. Pardo, “Estudio y desarrollo de un sensor de alcohol basado en pilas de combustible”.
- ¹⁹ L. Cámara, “Análisis temporal de sensores de mezcla de gases: Reconocimiento de componentes”
- ²⁰L. Elias and A. H. Lawrence, “Portable trace narcotics detector for field use,” *Canadian Journal of Spectroscopy* **32 (2)**, 14A-15A (1987).
- ²¹A. H. Lawrence, “Characterization of benzodiazepine drugs by ion mobility spectrometry,” *Anal. Chem.* **61**, 343-349 (1989).
- ²²P. Neudorfl A.H. Lawrence, “Detection of ethylene glycol dinitrate vapors by ion mobility spectrometry usind chloride reagent ions,” *Analytical Chemistry* **60 (2)**, 104-109 (1988).
- ²³T B. Bourrounet, Talou, A. Gaset, “Application of a multi-gas-sensor device in the meat industry for boar-taint detection,” *Sensor ans Actuators B* **26-27**, 250-254 (1995).

- ²⁴J.W.Gardner and H.T. Chan V. Shunter, "The application of discrimination techniques to alcohols and tobaccos using tin oxide sensors," *Sensor and Actuators* **18**, 361-371 (1989).
- ²⁵Harold V. Shurmer and Julian W. Gardner, "Odour discrimination with an electronic nose," *Sensors and Actuators B* **8**, 1-11 (1992).
- ²⁶A. Fukuda y T. Moriizumi T. Nakamoto, "Perfume and flavor identification by odor sensing system using quartz resonator sensor array and neural network pattern recognition," *IEEE* , 355-358 (1991).
- ²⁷S. Tsubakino H. Nanto, M. Ikeda, F. Endo, "Identification of aromas from wine using quartz resonator gas sensors in conjunction with neural network analysis," *Sensors and Actuators B* **24-25**, 794-796 (1995).
- ²⁸E.J. Brunke P. Mair, "Métodos modernos en la investigación analítica de aromas," *Dragoco report* (1988), pp. 99-108.
- ²⁹W. H. Ko T. Grandke, *Sensors a Comprehensive Survey: Fundamental and general aspects* (VCH, New York, 1992).
- ³⁰H. Lüdi and H. M. Widmer N. Graber, "The use of chemical sensors in industry," *Sensors and Actuators B* **1**, 239-243 (1990).
- ³¹F.R. Foulkes A.J. Appleby, *Fuel cells handbook* .
- ³²H. Nakajima A. Katayama, K. Fujikawa, H. Kita, "Metal electrodes bonded on solid polymer electrolyte membranes (SPE)," *Electrochemical Acta* **28 N° 6** (1983).
- ³³F. Javier Gutierrez, "Semiconductor sensors for monitoring atmospheres and soils," presented at the Second European School on Sensors (ESS'95), Barcelona, 1995 (unpublished).
- ³⁴S.M. Sze, *Semiconductor Sensors* (John Wiley & Sons, INC., New York, 1994).
- ³⁵K.D.Schierbaum W. Göpel, "SnO₂ sensors: current status and future prospects," presented at the Eurosensors VIII, Toulouse, 1994 (unpublished).
- ³⁶U. Weimar A. Hierlemann, G. Kraus, M. Schweizer-Berberich, W. Göpel, "Polymer-based sensor arrays and multicomponent analysis for the detection of hazardous organic vapours in the environment," presentado en Eurosensors VIII, Toulouse, 1994 (unpublished).
- ³⁷W. Göpel, "New Materials and transducers for chemical sensors," *Sensors and Actuators B* **18-19**, 1-21 (1994).
- ³⁸W. Göpel, "Chemical sensors for cars in the future," presentado en the Second European School on Sensors (ESS'95), Barcelona, 1995 (unpublished).
- ³⁹A. Gerlach K.D. Schierbaum, M. Haug and W. Göpel, "Selective Detection of organic molecules whit polymers and supramolecular compounds: Application of capacitance, quartz microbalance and calorimetric transducers," presentado en Eurosensors V, 1991 (unpublished).
- ⁴⁰R. Dändliker E. Wagner, K. Spencer, *Optical sensors* (VCH, New York, 1992).
- ⁴¹T.A. Jones W. Göpel, M.Kleitz, *Chemical and biochemical sensors* (VCH, New York, 1992).
- ⁴² H. Ohnishi, T. _Matsumoto, M. Ippommatsu, "Highly sensitive methane sensor" International Gas Union, 19th World Gas Conference (1994)
- ⁴³ N. Yamazoe, "New approaches for improving semiconductor gas sensors" *Sensors and Actuators B*, **5**, 7-19 (1991)
- ⁴⁴Modelo de utilidad. Número de solicitud 9501043. "Dispositivo indicador de alcoholemia por aire expirado" (1996)
- ⁴⁵ C. Wagner, *J.Chem Phys*, **18**, 69 (1950)
- ⁴⁶ K. Hauffe and H.J. Engell, *Z. Electrochem.*, **56** 366 (1952)

- ⁴⁷ T. Seiyama, A. Kato, K. Fujiishi and M. Nagatani, "A new detector for gaseous compounds using semiconductor thin films" *Anal. Chem.* **34** (1962) 1502-1503
- ⁴⁸ N. Taguchi US Patent 3 695 84 October (1972)
- ⁴⁹ G. Sberveglieri "Classical and Novel techniques for preparation of SnO₂ thin-film sensors" *Sensors and Actuators B*, **6**, 239-247 (1992)
- ⁵⁰ W.H. King, "Piezoelectric sorption detector," *Anal Chem* **36**, 1735-1739 (1964).
- ⁵¹ H. Wohljen and R. Dessy, "Surface acoustic wave probe for chemical analysis," *Anal. Chem.* **51**, 1458-1475 (1979).
- ⁵² J.E. Sanchez O.A. Mora, "Fisiología del olfato," in *Neurofisiología* .
- ⁵³ J. Boeckh H. Altner, "Gusto y olfato," .
- ⁵⁴ V. G. Foglia, "Los sentidos químicos: el olfato y el gusto," .
- ⁵⁵ John E. Amoore, "Olfactory genetics and anosmia," *Handbook of Sensory Physiology*, edited by Springer (New York, 1971), Vol. 4, pp. 245-256.
- ⁵⁶ Trygg Engen, "Olfactory psychophysics," *Handbook of Sensory Physiology*, edited by Springer (New York, 1971), pp. 217-244.
- ⁵⁷ K. Persaud and G. Dodd, "Analysis of discrimination mechanism in the mammalian olfactory system using a model nose," *Nature* **299**, 352-355 (1982).
- ⁵⁸ S. Haykin, *Neural Network: a comprehensive Foundation* (1994).
- ⁵⁹ Kolmogorov, A.N. "On representation of continuous functions of several variables by superposition of continuous functions" *Dokl. Akad. Nauk. SSSR*, **114**, 953-956 (1957)
- ⁶⁰ Kurková, V. "Kolmogorov's Theorem and Multilayer Neural Networks", *Neural Networks* **5**, N.3, 501-506 (1992).
- ⁶¹ Siegelman, H. y Sontag, E.D., "Neural networks are universal computing devices" Technical Report SYCON-91-08, Rutgers Center for Systems and Control, (1991)
- ⁶² Frank Rittler y Gerhard Schamaus Ernst-Joachim Brunke, "Nuevos resultados sobre los componentes traza sensorialmente relevantes," in *Dragoco report* (1996), pp. 5-21.
- ⁶³ Michael Phillips, "Pruebas de aliento en medicina," *Investigación y Ciencia* (1992).
- ⁶⁴ J. M. Gardner H. V. Shurmer, P Corcoran, *The electronic nose* Mc Graw-Hill (1990).
- ⁶⁵ H. V. Shurmer, "Basic Limitations for an electronic Nose," *Sensor and Actuators B* **1**, 48-53 (1990).
- ⁶⁶ R.W. Moncrieff, "An instrument for measuring and clasifying odours" *J. Appl. Physiol* **16** (1961) 742
- ⁶⁷ W.F. Wilkens and A. D. Hatman "An electronic analog for the olfactory processes", *Ann NY Acad Sci.*, **116** (1964) 608
- ⁶⁸ T.M. Buck, F.G. allen and M. Dalton, "Detection of chemical species by surface effects on metal and semiconductors", in T. Bregman and A. Dravnieks (eds.), *Surface effects in detection*, Spartan Books Inc., USA 1965
- ⁶⁹ A. Dravnieks and P.J. Trotter, "Polar vapour detection based on thermal modulation of contact potenciales", *J. Sci. Instrum.* **42** (1965) 624
- ⁷⁰ J. W. Gradner, P. N. Bartlett "A brief history of electronic noses" *Sensors and Actuators B*, **18-19** (1994) 211-220
- ⁷¹ T. Sundic, S. Marco, J. Samitier, P. Wide, "Electronic tongue and electronic nose data fusion in classification with neural networks and fuzzy logic based models" *Instrumentation and Measurement Technology Conference. IMCT2000* (2000)
- ⁷² N.F. Suprenant, "Emissions Assessment of Conventional Stationary combustion Systems", Vol 5: Industrial Combustion Sources, EPA Contact No. 68-02-2197, GCA Corp., Bedford, AM, 1980.
- ⁷³ C. Calaza, "Detector óptico de gases basado en absorción IR"

Direcciones web

- W1 <http://www.Alpha-Mos.com>
- W2 <http://chem.external.hp.com/cag/products/hp4440.html>
- W3 <http://www.aromascan.com>
- W4 <http://www.airsense.com/uk/main.html>
- W5 <http://www.intraders.com/jicc/cosmos/cos-top.html>
- W6 <http://www.raesystems.com/home1.htm>
- W7 <http://eco-web.com/register/02123.html>
- W8 <http://www.neotronics.com>
- W9 <http://nose.uia.ac.be>
- W11 <http://www.ine.es>
- W12 <http://chem.extrnal.hp.com/cag/products/hp4440.html>
- W13 <http://www.bloodhound.co.uk/bloodhound>
- W14 <http://www.cyranosciences.com>
- W18 <http://www.estcal.com/>

Capítulo 2

Sistema experimental

- *“No necesitas sesos. Estás aprendiendo algo cada día. Un bebé tiene sesos, pero no sabe mucho. La experiencia es la única fuente de conocimientos y cuanto más permanezcas en la tierra, tanta más experiencia tendrás sin duda.”*

L. Frank Baum: “El mago de Oz”

- *La ciencia es un intento de lograr que la diversidad caótica de nuestras experiencias sensoriales corresponda a un sistema de pensamiento lógicamente uniforme. En este sistema, cada experiencia debe estar en correlación con la estructura teórica de tal modo que la relación resultante sea única y convincente.*

Abert Einstein: “Sobre la teoría de la relatividad”

Prólogo del capítulo

En el capítulo anterior se introdujo el concepto de nariz electrónica así como algunas de sus aplicaciones más útiles. La realización de un dispositivo complejo como una nariz electrónica, o como, en general, un sistema de test de sensores de gases es una tarea ardua. Todo ello requiere de un elaborado montaje experimental que debe incluir un sistema de control exhaustivo y un sistema de medida adecuado a los experimentos que se pretendan realizar.

El presente capítulo se estructura en dos secciones cualitativamente diferentes. La primera trata de enunciar a priori las características y la problemática básica de un sistema experimental como el que se pretende montar. La segunda sección trata concretamente de cada una de las partes que conforma el sistema experimental. En esta segunda parte se profundizará en la explicación a nivel técnico y de detalle para cada una de las partes que se consideran fundamentales para el funcionamiento del equipo. Así, se dará una descripción completa de los gases utilizados, de los sistemas de manejo de mezclas, del diseño de las cámaras de medida, de los sensores que se utilizan y de los sistemas electrónicos de acondicionamiento de señal y control del equipo. Hay que puntualizar, sin embargo, que el sistema está sujeto a continua evolución, tratando siempre de incorporar mejoras que aumenten sus prestaciones. Por este motivo, quizás la descripción del sistema experimental que se expone en este capítulo no se corresponde con el estado actual del mismo. Por otra parte, y aunque es una precisión que se sobreentiende, ha de quedar patente que el desarrollo de un sistema así no es atribuible al mérito de una sola persona, sino que es fruto del trabajo en equipo llevado a cabo por el grupo de tratamiento de señal del departamento de electrónica.

El montaje experimental que se pretende construir es por tanto y fundamentalmente un sistema de test de sensores de gases pero que, además, ha de permitir, a través de una serie de algoritmos de procesado de señal, generar modelos de comportamiento del sistema y realizar identificación directa de la señal de los sensores, identificación inversa de la concentración gaseosa o reconocimiento de los patrones gaseosos introducidos en el sistema. Esta problemática, los algoritmos antes mencionados (ya sean de carácter general o diseñados específicamente para el problema en cuestión), y la verificación de su rendimiento se expondrán en los siguientes capítulos. El desarrollo de este sistema iniciado con este trabajo, recibió un impulso definitivo al adquirirse el grupo, los compromisos de test derivados del proyecto, financiado por Gas Natural y MCC, "Detector selectivo CO/CH₄" para el estudio de la detección de gases de combustión en cocinas domésticas.

También se presentará la descripción de un sistema similar alternativo para el test de sensores QMB del cual se extraerán los resultados experimentales concernientes a componentes orgánicos volátiles. Este sistema no ha sido desarrollado en el Departamento de Electrónica de la Facultad de Física de la Universitat de Barcelona sino que pertenece al grupo de investigación en sensores dirigido por el profesor Göpel, de la Universidad de Tübingen y con el cual se ha colaborado estrechamente tanto en el diseño de experimentos como en la presentación de los resultados en diferentes artículos y congresos^{17 18}. El hecho de que los datos presentados en los capítulos siguientes pertenezcan a este segundo sistema experimental, no invalida el desarrollo del presente ya que, en definitiva, la construcción de un sistema experimental orientado hacia el trabajo con alarmas de gas domésticas, fue uno de los objetivos a conseguir en el planteamiento de este trabajo. Por otra parte, debe aclararse que, las similitudes en la esencia de ambos sistemas, como se verá, son mayores que sus diferencias.

2.1. Introducción

Por supuesto, una parte fundamental en el desarrollo de una aplicación en análisis de gases es el propio sistema de test y medida de los dispositivos sensores. Las necesidades en este sentido se han visto incrementadas a causa del aumento del interés en estos últimos años en desarrollar sistemas multisensores más potentes y versátiles. Además, los progresos en instrumentación en los últimos años han revolucionado la concepción de los sistemas de medida, poniendo a disposición del usuario tecnologías avanzadas a un precio competitivo. Aun así, el coste de un sistema de medida dentro de un proyecto global (como pueda ser el que se presenta en esta trabajo), constituye generalmente un alto porcentaje del coste total del proyecto. Por este motivo, es fundamental plantear el diseño del sistema con rigurosidad y precisión, conociendo en el mayor grado posible cuál es la física que está involucrada en el problema, dónde residen las dificultades y cuáles son las alternativas.

El sistema que se pretende desarrollar, ha de ser capaz de generar concentraciones de gas con buena precisión, con rapidez y con versatilidad. Por precisión se ha de entender que las concentraciones mínimas están en el orden de las decenas de ppm, por rapidez se quiere indicar la capacidad del sistema para permitir cambios de concentración en pocos minutos, para minimizar la duración de los experimentos y por versatilidad se pretende expresar la capacidad del sistema de generar no sólo atmósferas de composición monocomponente sino también mezclas de gas multicomponente, y además, se desea que el sistema sea capaz de generar muestras gaseosas a partir de sustancias que no estén en estado gaseoso.

Ante el conjunto global de esta problemática, las alternativas se reducen esencialmente a dos: sistema de obtención de muestra estática o sistema de obtención de muestra dinámica²³.

Un esquema de un sistema de test que utiliza el método estático de generación de muestras se presenta en la figura 2.1. Consiste en introducir una cantidad específica y bien determinada de soluto, líquido o gaseoso, en una cámara rígida de volumen conocido. Los métodos de introducción pueden ser varios, aunque lo habitual es utilizar o bien una jeringa, caso de soluto líquido, o bien un controlador de flujo másico (MFC), caso de soluto gaseoso. Si la muestra que se ha de introducir es líquida y se utiliza una jeringa, se evapora mediante algún tipo de calefactor, y se mezcla con el disolvente que hay dentro de la cámara. Para que la mezcla sea eficiente suele introducirse algún elemento que genere turbulencia, como por ejemplo un ventilador. En estas condiciones, la concentración de gas i viene dada como relación de volúmenes según una ecuación como la que sigue:

$$C_i = v_i/V \quad (2.1)$$

siendo v_i es volumen de soluto i y V el volumen total de la cámara.

Otra opción que se baraja habitualmente para el caso de soluto líquido es utilizar un flujo de gas portador, habitualmente aire seco, controlado por un controlador de flujo másico (MFC), que circula por un recipiente con la sustancia líquida de manera que la propia corriente portadora arrastre moléculas del soluto. La cantidad de soluto presente en el flujo portador depende en gran medida de la temperatura de operación. Más adelante se profundizará en la explicación de este método. También es habitual como alternativa, disponer de una pequeña cámara de evaporación, previa a la propia cámara de medida. En esta precámara se evapora la sustancia líquida que luego se arrastra mediante un flujo portador al interior de la cámara de medida.

En el caso de utilizar introducción de muestra gaseosa a través de un MFC, la concentración de gas i vendría dada por la ecuación 2.2.

$$C_i = \frac{1}{V} \int f(t) dt \quad (2.2)$$

siendo f el flujo del controlador, t el tiempo que dura la introducción de muestra y V el volumen de la cámara. En este caso, conocer perfectamente el volumen de la cámara y controlar el tiempo resulta fundamental para obtener buena precisión en la generación de la muestra.

Varios departamentos de diversas universidades utilizan este tipo de sistemas estáticos. Por ejemplo, el esquema de la figura 2.1 representa un sistema de test implementado en el departamento de electrónica de la Universitat Rovira i Virgili¹⁹ (España), y un sistema similar ha sido desarrollado en el departamento de ingeniería de la Universidad de Tottori²⁰ (Japón). En ambos casos la obtención de las mezclas se basa en la inyección de la sustancia licuada en la cámara de test. Un sistema que utiliza gas portador para el arrastre de moléculas de sustancias volátiles es el desarrollado en el departamento de ingeniería de la Universidad de Warwick²¹ (Gran Bretaña). Para obtener más información y un listado más extenso de diferentes sistemas de test, consultar la bibliografía^{6 16 23}.

Este tipo de sistemas de test es económico y muy simple y es especialmente útil cuando no se precisa de una gran versatilidad en las medidas⁵. Sin embargo estos sistemas adolecen de ciertas carencias y sus prestaciones son limitadas. Un primer problema, aunque no el fundamental, proviene del estimar el volumen real de la cámara; si esta contiene objetos en su interior (ventiladores, los propios sensores, calefactores, etc.) y además, estos objetos pueden ser unos u otros, dependiendo del experimento, la precisión en la estimación de dicho volumen puede decrecer de forma importante, repercutiendo directamente en la estimación de la concentración de las muestras. Por otra parte, estos sistemas son incapaces de sobreponerse a problemas de absorción de muestra en la cámara o de fugas. Además, realizar cambios de concentración en este tipo de sistemas es un proceso lento y tedioso, y el tiempo necesario para que se produzca la mezcla de los gases depende del tamaño y forma del recipiente, de las características difusivas de los gases y de los fenómenos turbulentos que se produzcan.

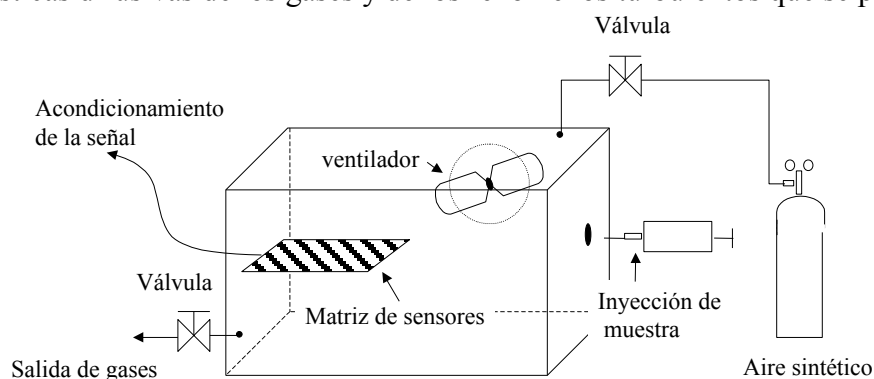


Figura 2.1 Obtención de muestras de gas : método estático

Un esquema de un sistema de generación de muestras dinámico se puede ver en la figura 2.2. En este tipo de sistemas se utilizan MFC para obtener un flujo constante de gas a la concentración deseada, que atraviesa la cámara de medida. En este caso, la concentración del gas i viene dada por la relación de los flujos portadores:

$$C_i = f_i / F \quad (2.3)$$

siendo f_i el flujo del gas i y F el flujo total.

Para generar muestras gaseosa a partir de sustancias en estado líquido es habitual utilizar el método de arrastre de moléculas mediante un flujo portador. Como ya se ha dicho anteriormente, más adelante se profundizará en la explicación de este proceso.

Algunos de los problemas que aparecían en los sistemas estáticos se reproducen en los dinámicos. Basados los dinámicos como están, en flujos de los gases, debe tenerse especial cuidado en no caer en regímenes de flujo laminar o estratificado. Para evitar esta eventualidad se suelen incorporar al sistema cámaras de premezcla que aseguren una perfecta mezcla de los flujos⁴, aunque la presencia de codos y bifurcaciones en el camino del flujo hasta la cámara suele asegurar una buena mezcla. La principal ventaja de éste método frente al sistema estático, es que se minimizan los problemas de absorción o desorción en la cámara. Además es mucho más versátil y permite de forma ágil modificar las concentraciones de los gases o generar mezclas multicomponente. Su inconveniente principal radica en su mayor complejidad y costo económico, no sólo en referencia al sistema en sí mismo sino también al alto consumo de gas que implica.

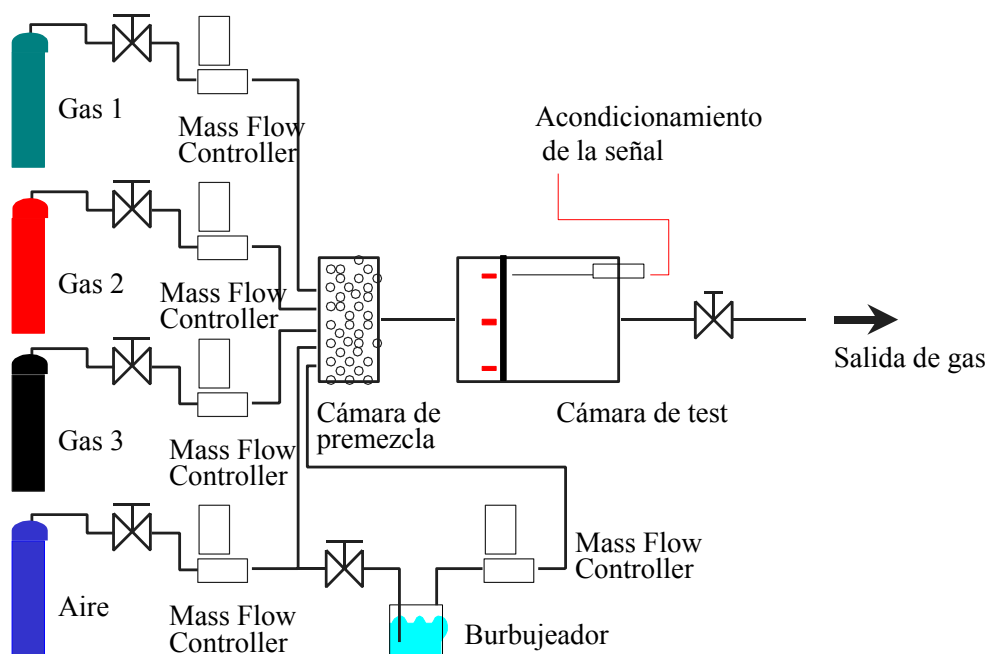


Figura 2. 2 Sistema de generación de muestras dinámico

Sistemas como este han sido desarrollados en el Fraunhofer-Institut für Festkörpertechnologie⁶ (Alemania) o en el Institute of Physical and Theoretical Chemistry, de la Universidad de Tübingen²² (Alemania). A partir de este último sistema se han obtenido los datos objeto de este trabajo.

2.2. Concepción del sistema

Antes de entrar a describir los pormenores del sistema resultará interesante presentar a priori algunas características básicas del sistema que se pretende desarrollar. Resulta conveniente abordar el problema del montaje experimental desde tres apartados diferentes: los gases y sus mezclas, los sensores, y el acondicionamiento de la señal y medición.

2.2.1 Sobre gases y mezclas

Como se ha mencionado en el prólogo, este trabajo se desarrolla como parte de un proyecto dedicado al estudio de los gases de combustión en cocinas domésticas con detectores de gas basados en sensores de óxidos metálicos. En este sentido, el sistema, aunque debe estructurarse de manera versátil, se va a desarrollar inicialmente para la detección de gases relacionados con procesos de combustión. Por lo tanto, en este entorno queda bien delimitado el tipo de gases a estudiar y, en una primera acometida al problema, se van a utilizar monóxido de carbono (CO), dióxido de carbono (CO₂) y metano (CH₄). Aunque el CO₂ apenas es detectado por los sensores de óxidos metálicos, su presencia se justifica a través del estudio del problema mediante otras tecnologías, y de hecho, existe una línea de trabajo abierta para el test de sensores ópticos.

Los rangos de concentración que se pretenden utilizar para el test de sensores son:

CH₄ : 500 - 20.000 ppm

CO : 30 - 1.000 ppm

CO₂ : 500 - 20.000 ppm

De alguna manera, estos valores están relacionados con los límites que dictan las normativas^{1 2 3 4} así como con las prestaciones que el sistema podrá alcanzar y que se detallarán más adelante. La generación de estas concentraciones y el tipo de estudio que se desea realizar a partir de ellas imponen condiciones al sistema. Así, la información que se pretende extraer a partir de la respuesta del sistema a las concentraciones no debe limitarse al estado estacionario. Debe plantearse un objetivo más ambicioso y estudiar la evolución de las respuestas del sistema en el tiempo, es decir, su dinámica. Dicha dinámica depende directamente de dos factores principales: de las respuestas de los sensores y de la respuesta del recinto que los aloje. En este sentido, si se desea que el sistema sea rápido, el recinto debe ser pequeño, pero por otro lado, las dimensiones físicas de los sensores imponen un cierto tamaño mínimo que puede, muy fácilmente, tener una característica dinámica no despreciable. En cualquier caso, la influencia de la cámara sensorial es ineludiblemente, y el planteamiento que hay que adoptar no es tanto el conocimiento de la dinámica sensorial sino más bien la dinámica del sistema global.

Pero para que esta disquisición sea factible en la práctica, es absolutamente necesario que a partir de la señal de concentración, que es la señal de entrada del sistema, sea posible realizar estimaciones dinámicas. Ya en la introducción se planteó el deseo de disponer de una capacidad de generación de concentraciones ágil y versátil. Estos adjetivos adquieren ahora mayor significado; no sólo hacen referencia a rapidez y facilidad en la capacidad de generación de mezclas sino también a la aspiración que el sistema sea capaz de generar concentraciones según diferentes formas de señal.

Ciertamente, en un estudio dinámico, interesa generar señales de concentración bajo ciertas formas típicas como escalones, sinusoides, rampas, ruido blanco,... Este tipo de prestación no es requerida en un estudio de características estáticas. De lo expuesto en el apartado anterior, el método adecuado para conseguir estas prestaciones respecto a la generación de mezclas gaseosas es, no cabía esperar otra cosa, el método dinámico.

2.2.2 Sobre Sensores

Además de los gases, las características sensoriales imponen nuevas condiciones al diseño del sistema. Este se plantea, por existir un mercado muy amplio y por su bajo

coste, para sensores de óxidos metálicos en una configuración matricial con buena, aunque no exclusiva, sensibilidad a los gases que se desean estudiar. El uso este tipo de sensores introduce dos nuevas necesidades: el acondicionamiento de la señal para adecuar la señal al sistema de medida posterior, y la activación y el control de los calentadores de dichos sensores. Tanto para una como para otra necesidad, es un requerimiento fundamental el que los sensores sean tratados de modo independiente. Así, el acondicionamiento, debe ser ajustable independientemente para cada sensor puesto que las diferentes características de los sensores de la matriz así lo requieren. En cuanto a las resistencias calefactoras de los sensores, debe preverse la posibilidad de gobernarlas de forma individualizada no sólo por el hecho de que cada sensor puede tener requisitos térmicos diferentes sino por la posibilidad que así se ofrece de realizar estudios particulares de dependencia con la temperatura del calefactor. Además, el sistema sensorial debe estar sometido a un control completo de la temperatura y de la humedad puesto que estos factores afectan directamente al rendimiento de los sensores.

2.2.3. Sobre acondicionamiento de la señal

Las señales que ofrezca la matriz de sensores deben ser recogidas por un sistema de medida que debe ser automático, fiable y preciso, permitiendo el tratamiento de toda la información a posteriori. Además, el conjunto debe ser absolutamente autónomo, es decir no debe precisar de ninguna actuación por parte del operador una vez definido el proceso de test y su protocolo de ejecución. Esto supone que todo los procesos (generación de mezclas, medición, archivado de la información, etc.) deben ir enmarcados en un preciso diseño del experimento en el cual se detallen todas y cada una de las acciones que el sistema debe efectuar.

En base a lo expuesto, y pretendiendo una concepción del sistema no dispersa, resulta conveniente centralizar el sistema de medida y el control del experimento en un ordenador complementado con el número necesario de tarjetas de entrada/salida analógico-digitales. De esta manera se evita la complejidad causada por múltiples subsistemas dedicados específicamente a una sola tarea, a costa de la necesidad de desarrollar interfaces específicas y adecuadas a las características del ordenador y de las tarjetas.

Se desea, por tanto, diseñar un sistema de medición de gases que disponga de:

- a) Una generación de mezclas gaseosas versátil. Esto elección exige, como ya se ha dicho, el uso de flujos gaseosos controlados,
- b) Un diseño individualizado de interfaces para cada sensor tanto desde el punto de vista de acondicionamiento de señal como de control del calentador y medición de las magnitudes de temperatura y humedad a las que se realiza el experimento.
- c) Hardware y un software de medida y control centralizado en un ordenador con la capacidad de medir y controlar al sistema de forma autónoma.

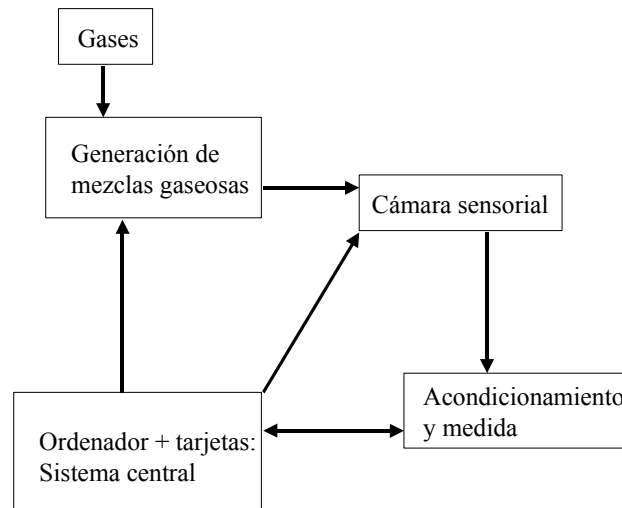


Figura 2.3 Diagrama de bloques del sistema

2.3 Sistema experimental

Repasados algunos sistemas ya existentes, establecida la necesidad de utilizar un método dinámico en la generación de mezclas gaseosas, (que acarrea por tanto, la necesidad de utilizar controladores de flujo másico), teniendo en cuenta además que el sistema sensor se basa en sensores de óxidos metálicos, y que resulta imprescindible centralizar todo el control y la medición en un ordenador, el diseño esquemático del sistema a montar puede representarse de la siguiente manera:

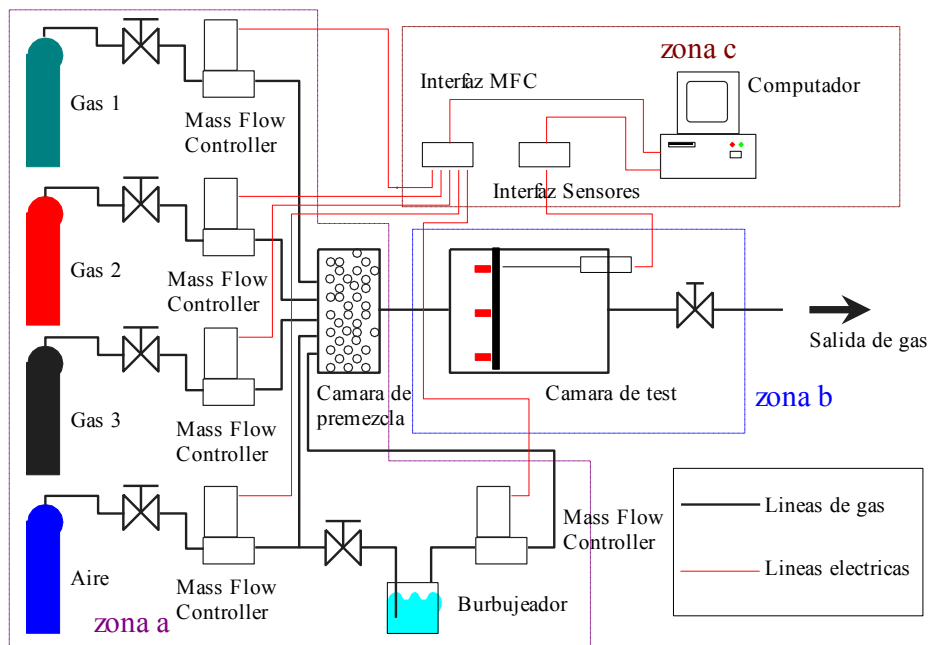


Figura 2.4 Esquema de un sistema de medición gaseosa

Tal y como cabe esperar, en el esquema se pueden diferenciar tres zonas fundamentales:

- zona de gases y controladores de flujo
- zona de cámara de test
- zona de control y medida de datos

En la zona *a* se disponen las bombonas de los diferentes gases y los controladores de flujo másico (MFC) para realizar el control de flujo y las diluciones de gas que se precisen. Aunque en el esquema sólo aparecen tres gases y el aire portador, el sistema permite la ampliación hacia un número mayor de gases.

Se incluye en esta zona, la cámara de premezcla, para asegurar la perfecta mezcla de los flujos. En la zona *b* se encuentra el sistema sensor propiamente dicho. Consta exclusivamente de una cámara donde se aloja la matriz de sensores pertinente, encargada de detectar la presencia de gases. La zona *c* contiene, principalmente un ordenador al que se le incorporan tarjetas de adquisición A/D y las interfaces de control. El conjunto controla el proceso de medida, y recoge los datos para su posterior análisis.

2.4. Gases

Tal como se menciona en el prólogo, el desarrollo de este sistema viene marcado por el proyecto “Detector inteligente CO/CH₄”. Centrándose por tanto en los gases de combustión procedentes de cocinas domésticas y si se revisan algunas normativas europeas^{1,2,3,4} relacionadas con el tema; éstas hacen hincapié en la necesidad de realizar determinados tests con ciertas sustancias que son incluso ajenas a los procesos involucrados en una combustión. Algunas de las sustancias mencionadas habitualmente en las normativas^{1,2,3,4} son el monóxido de nitrógeno, el dióxido de azufre, el etanol y el hexametildisiloxano. Por ejemplo, el estándar europeo que regula los aparatos eléctricos para la detección de gases combustibles en entornos domésticos menciona la realización de las siguientes pruebas:

- NO 5 ppm + 2 ppm durante 1 hora
- SO₂ 2 ppm + 1 ppm durante una hora
- Etanol 2000 ppm ± 10% durante 30 minutos
- Hexametildisiloxano 10 ppm durante 40 minutos

El monóxido de nitrógeno y el dióxido de azufre pueden ser, dependiendo del origen del combustible, productos procedentes de la combustión del gas natural. Obviamente, deben ser tenidos en cuenta por la normativa. Por otro lado, la razón de incluir otras sustancias como el etanol o el hexametildisiloxano, se basa en un doble motivo: son sustancias que pueden crear señales interferentes en los sensores del sistema y, además, son relativamente comunes en cualquier cocina doméstica. El etanol es un ejemplo claro de sustancia muy volátil, común en todos los hogares y que puede producir interferencias claras en la gran mayoría de sensores de gas. El hexametildisiloxano es una silicona muy volátil, también capaz de interferir con los sensores de gas y que se utiliza en numerosos productos de limpieza. Otro ejemplo de normativa que contempla pruebas con sustancias ajenas al propio proceso de combustión es la normativa europea sobre la regulación de detectores eléctricos para la detección de CO en el entorno doméstico. En este caso también aparece una prueba con etanol y hexametildisiloxano consistente en una secuencia de:

- aire limpio durante 15 minutos
- (600 + 50) ppm de etanol durante 10 minutos
- aire limpio durante 15 minutos
- (10+3) ppm de hexametildisiloxano durante 40 minutos

Además de estas consideraciones acerca de posibles sustancias interferentes, las normativas suelen incluir pruebas relacionadas con el efecto de la temperatura ambiente y la humedad sobre el sistema ya que ésta afecta también a la respuesta de los sensores.

El sistema inicial desarrollado consta de cuatro bombonas de gas que contiene metano,

CO₂, CO y aire sintético, pero es lo suficientemente versátil como para añadir en el futuro las líneas de gases necesarias para garantizar la realización de las pruebas que exigen las normativas tanto en lo referente a las sustancias interferentes como a pruebas de humedad, así como para añadir líneas de gases para otras aplicaciones no relacionadas con la combustión gaseosa.

Para garantizar los requisitos mínimos de detección y reconocimiento gaseoso e incluso para garantizar los requisitos de la normativa para los gases empleados (CO₂, CO y CH₄) no resulta necesario que estos sean altamente puros (ver tabla 2.1) ya que las impurezas que puedan contener no representan un porcentaje significativo una vez realizadas las diluciones en aire. Las concentraciones que se pretenden alcanzar se han explicitado anteriormente y para conseguirlas, una pureza media es más que suficiente. Sobre 1 ml de gas CO en un litro de disolución, que representa una concentración de 1000 ppm, y teniendo en cuenta una pureza del 99.3%, la cantidad de impurezas presentes es de 7 µl (7 ppm). Se demostrará en seguida que esta cantidad está por debajo del error que introduce las impurezas del aire disolvente.

En la siguiente tabla se especifica la pureza utilizada en estos gases y algunas de sus características relativas al riesgo de manipulación:

Características	CO	CO ₂	CH ₄
Pureza	C-23 (99.3%)	C-45 (99.995%)	C-35 (99.95%)
Peso molecular	28 gr/mol	44 gr/mol	16 gr/mol
LEL	12.5%	no definido	5%
TLV-TWA	25 ppm	5000 ppm	no definido
Estabilidad	Puede formar mezclas explosivas con el aire.	Estable en condiciones normales	Puede formar mezclas explosivas con el aire.
Toxicología	Produce daños a los glóbulos rojos (envenenamiento hemolítico)	Insuficiencia circulatoria. Dolor de cabeza, náuseas, vómitos o pérdida de conocimiento	No se conocen

LEL: Lower Explosive Level. Es el límite por debajo del cual el gas no se pueden formar mezclas explosivas. Se suele representar en % en volumen.

TVL-TWA: Threshold limit value-Time Weighted Average. Representa la concentración media en el tiempo a la cual puede estar sometido un trabajador durante una jornada laboral de 8 horas y 40 horas semanales sin notar efectos adversos.

El aire sintético es una mezcla elaborada a partir de oxígeno y nitrógeno puros, obtenidos por destilación criogénica del aire de la atmósfera. Este aire se utilizará como diluyente de los gases y por tanto, el nivel de pureza si que es crucial ya que un pequeño contenido mínimo de hidrocarburos puede modificar el valor de la disolución y afectar por tanto a la medida. El aire sintético utilizado tiene una pureza igual o superior al 99.999 %, y en un análisis típico, la composición del gas es: 21.0% oxígeno, 79.0% nitrógeno, 3 vpm agua, 1 vpm hidrocarburos C_nH_m y 0.5 vpm CO/CO₂. (vpm: volumen por millón). Este grado de pureza es equivalente a decir que tenemos un error

experimental de 10 ppm de impurezas. Resulta ahora evidente que, los 7 ppm de error que se cometen con el gas más impuro de los usados, el CO, están aun dentro de los límites que marcan las impurezas del aire.

La siguiente tabla muestra una comparación entre el aire atmosférico y el aire sintético:

Tabla 2. 2 Composición respecto al aire atmosférico (a 0 m de altitud, exento de humedad)		
Gases	Aire atmosférico(% Vol)	Aire sintético (% Vol)
N ₂	78.084	79.0
O ₂	20.946	21.0
Ar	0.934	*
CO ₂	0.033	*
Otros gases	0.003	*

Todas los gases encerrados en las bombonas están a alta presión, por lo que debe introducirse en la línea de gas algún tipo de sistema de reducción para adecuar la presión a las condiciones de trabajo de la siguiente etapa que son los controladores de flujo másico.

2.5 Controladores de flujo másico

Las diluciones de los gases con el aire sintético se realizan mediante controladores de flujo másico. El modelo utilizado es el 5850E de la casa Brooks Instrument⁷. Los controladores de flujo másico constan de tres unidades básicas: un sensor de flujo, una válvula de control de flujo y un sistema electrónico integrado de control del proceso. El sensor de flujo se basa en la medición de la diferencia de temperatura que aparece en los extremos de un capilar calefactado ante la presencia de un flujo de gas. Una fuente de potencia estable, proporciona una potencia constante a un calentador situado en el centro del capilar. Dos sensores de temperatura se hallan situados en los extremos del mismo capilar midiendo la temperatura a la entrada (T1) y a la salida (T2). Cuando no existe un flujo circulando por el capilar, ambas temperaturas son iguales pero si un cierto flujo lo atraviesa aparece una diferencia de temperatura entre la entrada y la salida que es proporcional al flujo de masa según la ecuación:



Figura 2. 5 Mass Flow Controler

$$\Delta T = A \cdot P \cdot C_p \cdot m \quad (2.4)$$

Donde:

ΔT es la diferencia de temperaturas (°K), C_p es el calor específico del gas a presión

constante ($\text{KJ/Kg}^\circ\text{K}$), P es la potencia del heater (KJ/s), m es el flujo másico (Kg/s) y A es una constante de proporcionalidad.

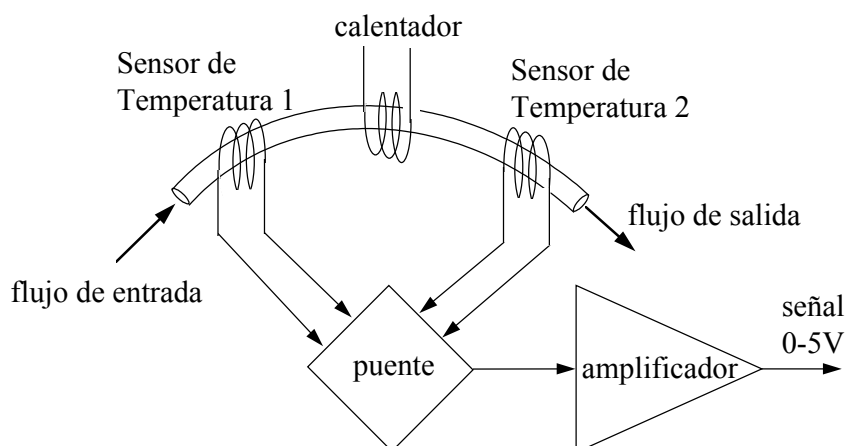


Figura 2. 6 Esquema del principio de funcionamiento de un MFC

Un circuito eléctrico, compuesto principalmente por un puente y un amplificador, genera señales lineales entre 0 y 5 Volts directamente proporcionales al flujo másico de gas. A través de esta señal y del control integrado es posible medir el flujo de gas y ajustar una corriente que modula la válvula para aumentar o disminuir el flujo.

Si, como ya ha quedado establecido, las diferentes concentraciones se forman a partir de las mezclas de los flujos, es claro que los flujos para soluto y disolvente han de ser diferentes. Por ello, el rango máximo de flujo para los MFC varía siendo de 3 ml/min para los gases soluto, mientras que para el disolvente, el rango máximo es de 1 l/min. Los flujos totales que se han de formar a partir de los MFC variaran en un rango que oscila alrededor de los 200 ml/min hasta un máximo de 500ml/min.

Las características principales del MFC que se utilizan en este montaje se listan en la siguiente tabla:

Tabla 2. 3 Características de los MFC Brooks 5850E	
Especificaciones	Valor
Rango operativo	3 ml/min-1 l/min
Precisión	± 1 % fondo de escala
Repetividad	± 0.25 % fondo de escala
Tiempo de respuesta	menos de 3 seg
Alimentación	+15 Vdc $\pm 5\%$, 25 mA -15 Vdc $\pm 5\%$, 180 mA 3.5 Watts de consumo
Limite temperatura	5 a 65 °C (modo operativo) -25 a 100 °C (modo no operativo)
Señal de salida	0-5 Volts en 2000 ohms o mayor. Rizado máximo 3 mV p-p.
Comand input	0-5 Vdc. Resistencia de entrada 200 K Ω
Presión de trabajo	100 bar max.
Conexión eléctrica	conector tipo D, 15 pins.
Conexión mecánica	compatible 1/4 " VCR

2.6. Líquidos y gases

El metano, CO_2 y CO son compuestos gaseosos a temperatura ambiente, pero no siempre las sustancias de interés lo son. Este es el caso del etanol que, aunque muy volátil, es líquido hasta los 78°C aproximadamente; o de la humedad por ejemplo, que debe ser introducida en las líneas gaseosas a partir de la evaporación de agua líquida. Para conseguir realizar experimentos con estas y otras sustancias líquidas resulta necesario incorporar algún dispositivo capaz de introducir en las líneas de gas del sistema experimental un flujo gaseoso, portador de una disolución con una concentración controlada de dichas sustancias. Si la nariz de cualquier mamífero es capaz de captar los olores provenientes de sustancias líquidas a partir de las moléculas extraídas del líquido gracias al arrastre por un flujo de aire, del mismo modo se puede idear un sistema capaz de transportar hasta el sistema sensor un flujo de aire mezclado con el soluto líquido que interese. De todas formas, si bien el control de flujo gaseoso es relativamente sencillo de realizar a través de un MFC, el control de flujos de vapores generados a partir de líquidos es algo más complejo.

Aunque existen diferentes métodos⁸, la esencia del proceso consiste en disponer de un recipiente conteniendo el líquido que se quiere evaporar. La presión en el interior del recipiente viene entonces determinada por la temperatura, que suele controlarse a través de algún tipo de calentador. Es bien conocido que pequeñas variaciones de temperatura producen grandes cambios en la presión de vapor y que, por tanto, el equilibrio líquido-gas es muy sensible a cambios de temperatura. Este es el motivo fundamental por el cual la temperatura debe permanecer constante y estar siempre bien controlada. Teniendo en cuenta este punto, puede optarse por tres vías para la generación del flujo. La primera opción consiste simplemente en elevar la temperatura y dejar que el propio proceso de evaporación produzca a la salida del recipiente un flujo de vapor (figura 2.7 izquierda). La segunda opción es forzar la salida de vapor mediante un gas portador que arrastre consigo las moléculas de vapor existentes en la interficie gaseosa (figura 2.7 centro). Si además se fuerza el paso controlado del flujo gaseoso portador a través del líquido (figura 2.7 derecha), entonces se habla de un dispositivo burbujeador. El flujo portador de un burbujeador, al circular a través del líquido, arrastra consigo moléculas del mismo, generando una cierta concentración de vapor en el flujo de salida, que también depende de la temperatura de trabajo.

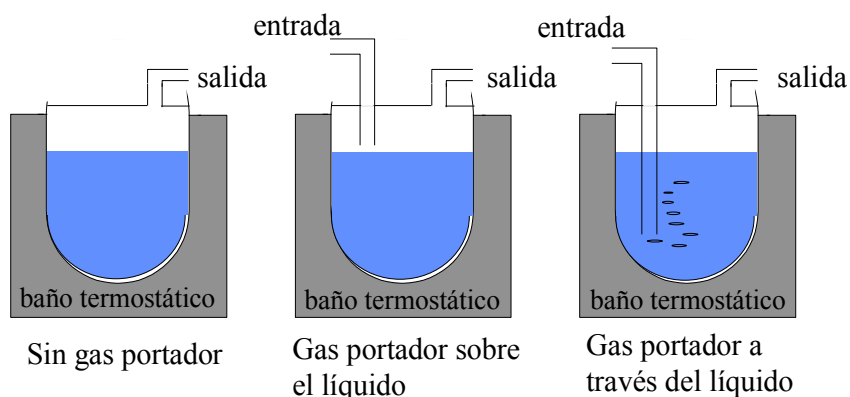


Figura 2. 7 Generación de concentraciones de gas a partir de líquidos

En el primer diseño, no hay un gas portador, y el control del flujo se realiza a través,

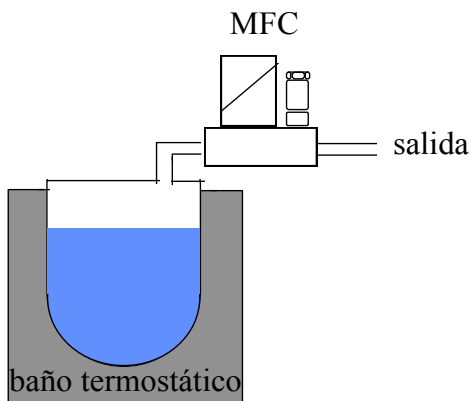


Figura 2. 8 Control del flujo de salida

exclusivamente de la temperatura. El diseño se puede optimizar colocando un MFC a la salida controlando el flujo de vapor en la salida (ver la figura 2.8). De todas formas, este MFC debe ser capaz de controlar el flujo aun con una baja presión en la entrada. Este tipo de MFC ha de tener una serie de características no comunes que hacen que esta opción sea, si mejor, pero también mas compleja y a la vez más cara.

Los otros diseños se basan en la presencia de un gas portador. La temperatura sigue siendo

un factor clave para mantener una cierta presión de vapor constante. El diseño de los

burbujeadores con gas portador puede también ser optimizado de las siguientes maneras:

- Incluyendo un control del flujo de gas portador.
- Controlando el flujo del gas portador y la presión del burbujeador.
- Midiendo la diferencia de conductividad térmica entre el gas portador a la entrada y en la salida. La diferencia de conductividad está relacionada con el contenido de vapor en el flujo. La cantidad de vapor en el flujo se puede controlar a través del flujo portador y de la temperatura.

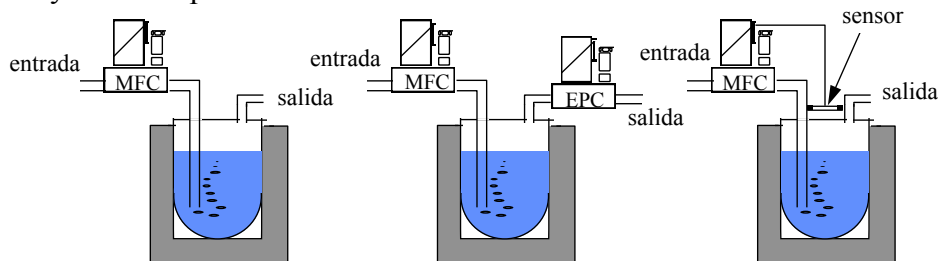


Figura 2. 9 Mejoras de los burbujeadores con gas portador

Todas estas opciones son, obviamente, mejoras del diseño básico, pero también aportan complejidad al sistema y lo encarecen en exceso. La opción escogida por sencillez y buen rendimiento sin encarecer excesivamente todo el sistema, es la del gas portador a través del líquido. La configuración es esta:

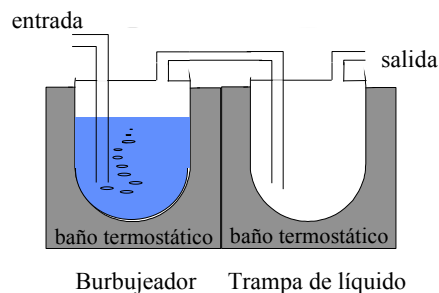


Figura 2. 10 Esquema de un burbujeador con trampa de líquido

El primer recipiente es el dispositivo burbujeador propiamente dicho. El flujo de entrada es aire seco y se controla con un MFC 5850E ya descrito. El segundo recipiente es una cámara de condensación para tratar de garantizar la concentración deseada en el flujo hacia la cámara sensorial⁶. Para ello, la cámara de burbujeo está a una temperatura un poco por encima de la de trabajo, y la cámara de condensación, un poco por debajo. Si este dispositivo se emplea para introducir humedad en el experimento, entonces la primera botella contiene agua, y el índice de humedad se obtiene en función de la relación de flujo de aire saturado y aire seco.

2.7. Matriz de sensores

La cámara sensorial propiamente dicha contiene una matriz de hasta seis sensores diferentes destinados especialmente a la detección de diferentes gases. La elección de los sensores de la matriz no viene tan sólo determinada por su eficiencia para determinada sustancia. Tal y como ya se ha mencionado en el capítulo anterior, el olfato humano contiene una serie de receptores olorosos sensibles a diferentes olores, pero que no son especialmente selectivos. De hecho, la mayoría presentan selectividades solapadas para varias sustancias olorosas. Esta forma de trabajar parece haber resultado evolutivamente más útil y rentable que no desarrollar receptores específicos y selectivos. Así la selectividad queda a cargo de un proceso analítico posterior.

Siguiendo esta tendencia natural, los sensores que se monten en la matriz deben también guardar una cierta relación entre selectividad y solapamiento de sus repuestas.

La matriz de sensores que se muestra en la figura 2.11 está constituida por seis sensores con los cuales se realizaron los primeros tests del sistema de medida.

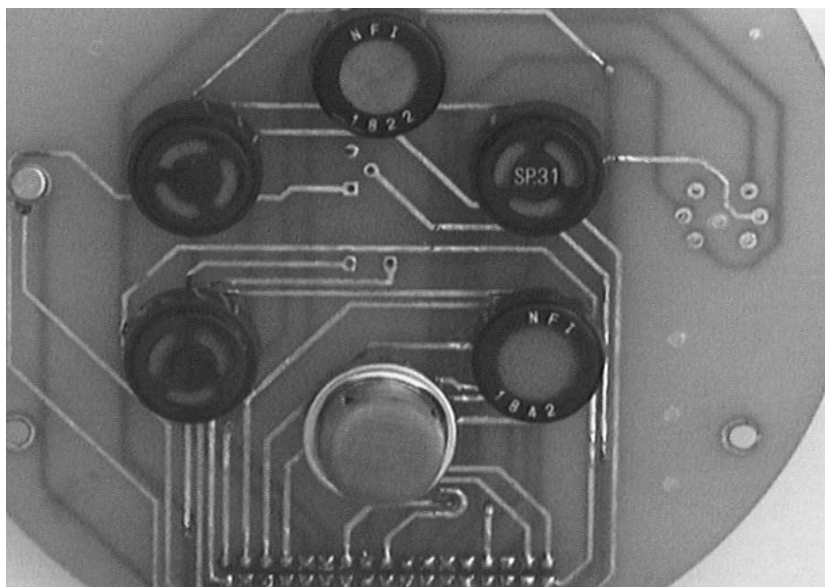


Figura 2. 11 Matriz de sensores

Tres de ellos son sensores del tipo ST31 de generaciones diferentes fabricados por la casa FIS. Los sensores ST31 son de carácter general, son sensibles a diferentes gases aunque se suelen usar par detectar etanol⁹. La matriz contiene también un sensor SP12 para la detección de metano, SB31 de carácter general, SB12 para gas natural^{10 11}.

Además de los sensores de gas propiamente dichos, al sistema se le incorporan sensores

de temperatura y de humedad para conocer y controlar, en todo momento ambas magnitudes en el interior de la cámara. Para ello se introducen en la matriz dos sensores AD590 para la medición de temperatura¹² y dos sensores PHILIPS H1 para la medición de la humedad¹³.

2.8. Electrónica de acondicionamiento y control

En esta parte del sistema se llevan a cabo todos los procesos de automatización, control y medida que, ineludiblemente, lleva asociado un dispositivo experimental semejante. A estas tareas está dedicado un ordenador Pentium, bus ISA a 120 MHz con dos tarjetas A/D: DT01-EZ y DT2815 de Data Translation, además del software adecuado para manejarlas. Las características esenciales de ambas tarjetas se exponen a continuación¹⁴:

DT01-EZ	
Entradas analógicas:	16 entradas single-ended o 8 entradas diferenciales 12 bits de resolución 27.5 KHz frecuencia Ganancia programable: 1, 2, 4, 8 Rango: 0-1.25,2.5,5,10V/±1.25,2.5,5,10 V dependiendo ganancia
Salidas analógicas:	2 salidas analógicas 12 bits de resolución 29,5 KHz frecuencia Rango: 0-5,10V/±2.25,5,10V
Digital I/O	16 salidas organizadas en dos grupos de 8
DT2815	
Salidas analógicas:	8 salidas analógicas 12 bits resolución 3.3 KHz frecuencia Rango: 0-5V/±5V

Resulta necesario, por tanto, desarrollar circuitería de interfaz para adaptar las señales del sistema a las tarjetas o para enviar señales desde las mismas al sistema y software para leer la información a través de las tarjetas de adquisición y desarrollar un entorno de usuario amigable.

2.8.1. Interfaz para Mass Flow Controlers

Esta interfaz permite generar, a través del ordenador, las señales de control y acción necesarias para los MFC. Con ella, es posible trabajar en tres modos diferentes: Modo de flujo controlado, modo de flujo cero o modo de máximo flujo.

Las entradas desde el ordenador son: una entrada analógica para comandar el MFC y 8 entradas digitales para hacer referencia al MFC y actuar sobre el comando de open/close. A la vez, la interfaz proporciona al ordenador una salida analógica indicando el flujo actual del MFC. Para todo ello se utiliza la tarjeta DT-01EZ cuyas características básicas se han expuesto antes. Para un mayor detalle sobre la estructura

de esta interfaz consultar bibliografía¹⁶.

2.8.2. Interfaz para los sensores

Por otro lado, tal y como se precisó anteriormente, los sensores demandan de dos atenciones diferentes por parte del sistema de control. Primero, por el hecho de ser sensores de óxido de estaño, poseen un calentador que debe ser alimentado a una cierta tensión (típicamente de un valor máximo de 5 Volts) para que el elemento sensor adquiera una cierta temperatura de trabajo. Segundo, el sensor ofrece como resultado de su exposición a un cierto gas una señal eléctrica relacionada con la concentración de dicho gas.

En referencia a las resistencias calefactoras de los sensores, su interfaz debe diseñarse, de acuerdo con las exigencias iniciales, de forma individualizada. Así, si se desea, pueden obtenerse temperaturas de trabajo diferentes para cada sensor en función de la tensión aplicada. La observancia de centralización del control planteada inicialmente requiere el diseño de una fuente variable de potencia integrada en el ordenador. La tarjeta DT-2815 se utiliza exclusivamente para conseguir este control individualizado bajo el prisma de un sistema centralizado. Esta manera de proceder redundante en una menor complejidad y mayor facilidad de programación e integración del control en el marco general del experimento. Sin embargo, la potencia necesaria para alimentar a los calentadores no puede ser suministrada por un ordenador. Cabe remarcar que los consumos típicos de los sensores de FIS o Fígaro más comunes oscilan entre los 300 y los 600 mW, y aunque pueden encontrarse sensores con consumos algo más reducidos, la interfaz de potencia siempre es la mejor opción. El siguiente diseño para dicha interfaz de potencia es capaz de realizar esta tarea:

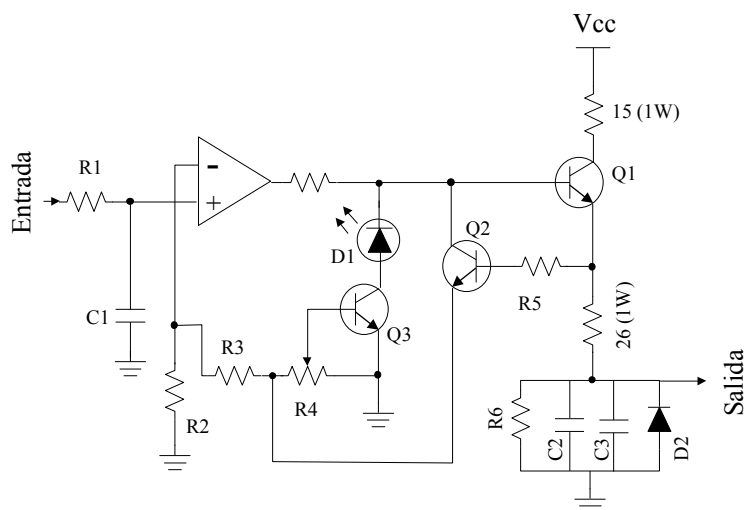


Figura 2.12 Circuito de potencia para las resistencias calefactoras

V_{in} es la señal de tensión programable suministrada por el ordenador. El operacional trabaja como seguidor de emisor ofreciendo una alta resistencia de entrada mientras que el transistor es el encargado de suministrar la potencia necesaria al calentador R_h . Todo el circuito trabaja en modo unipolar a una tensión $V_{CC} = 12V$ suministrada por una fuente exterior de 12 Volts y 2 Amp.

El operacional utilizado es el OP400 de baja tensión de offset, mientras que el transistor es el TIP31, suficiente para las necesidades que de él se requieren.

En referencia a la respuesta del sensor a concentraciones gaseosas el circuito de acondicionamiento de la medida de los sensores es el siguiente¹⁵:

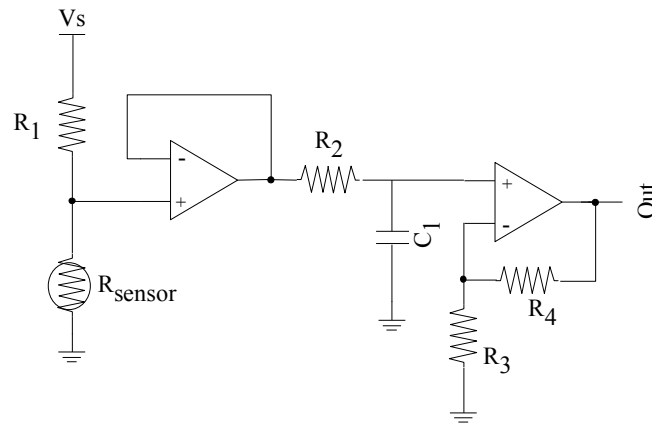


Figura 2.13 Circuito de acondicionamiento para el sensor de gas

Esencialmente está formado por un divisor de tensión a través de la resistencia del sensor, un seguidor de emisor, una etapa de filtrado y una etapa de ganancia. La tensión V_s proviene de una fuente de tensión estabilizada de 5 Volts, mientras que los operacionales están alimentados en modo bipolar entre ± 15 Volt. R_{Sensor} representa la resistencia del sensor. R_2 y C_1 conforman un sencillo filtro pasabajos con frecuencia de corte ajustada en el orden de la decena de Herz. R_3 y R_4 son las resistencias asociadas a la etapa de ganancia final en configuración inversor. Entonces, la tensión de salida de este circuito viene dada por la expresión:

$$\text{Out} = -\frac{R_4}{R_3} V_s \frac{R_{\text{sensor}}}{R_1 + R_{\text{sensor}}} \frac{Z_{C1}}{Z_{C1} + R_2} \quad (2.5)$$

Los operacionales utilizados son también del tipo OP400 de baja tensión de offset. La salida de cada uno de estos canales de medida, correspondientes a cada uno de los sensores, entran directamente a la tarjeta DT-01EZ instalada en el ordenador. En lo referente al control de temperatura y humedad, se han utilizado las interfaces que se describen a continuación.

Para el sensor de humedad capacitivo, se ha diseñado el circuito de la figura 2.14 consistente en dos timers 555 que proporcionan una salida PWM en función de las resistencias y capacidades. Un posterior filtro pasabajos (no dibujado en la figura) ofrece, a partir de la señal PWM, un nivel DC directamente proporcional a la capacidad del sensor, y por lo tanto relacionado a su vez con la humedad relativa presente.

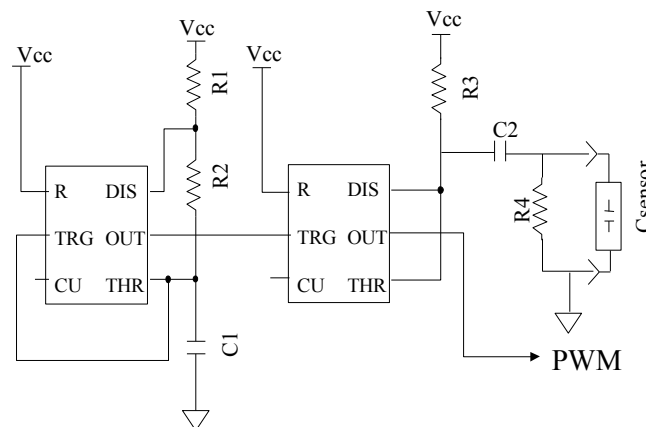


Figura 2.14 Circuito de acondicionamiento para el sensor de humedad

Para medir la temperatura se ha utilizado el siguiente circuito:

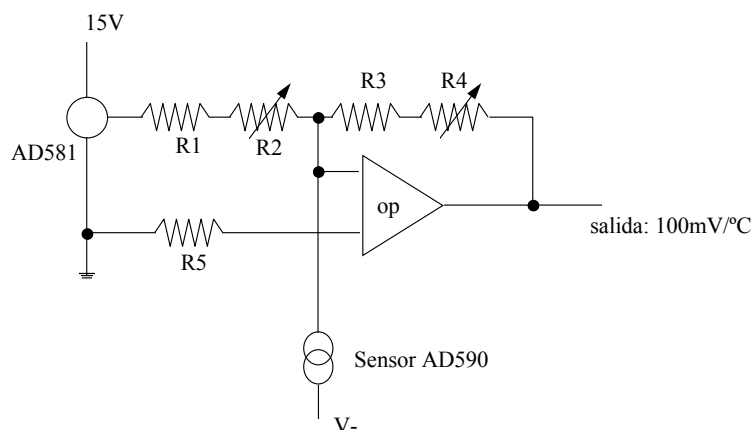


Figura 2.15 Circuito de acondicionamiento para el sensor de temperatura

El dispositivo AD581 es una fuente estable de 10 voltios. Con él y los valores adecuados de las resistencias, se consigue convertir la escala Kelvin a Celsius, ofreciendo la salida del circuito una señal en tensión equivalente a 100 mV por cada grado centígrado.

Todo este más o menos complejo sistema de medida y control está supervisado por el ordenador a través de las tarjetas y de un programa desarrollado en lenguaje HP-VEE (Hewlett-Packard Visual Engineering Environment, versión 3.12). Este es un software de programación gráfica pensado y optimizado para el control de instrumentación. Los programas en este entorno se conforman interconectando los diferentes iconos que representan a los diferentes equipos o funciones del sistema y generando así un flujo de datos desde el inicio (condiciones del experimento) hacia el final (resultados y medida). En definitiva, es un software de programación visual orientada a objetos que proporciona las herramientas necesarias para incorporar los diferentes equipos (como las tarjetas de Data Translations) como objetos del algoritmo, haciendo su manejo mucho más simple. El software diseñado genera las señales necesarias para el control de las resistencias calefactoras y de los MFC a la vez que lee de las tarjetas los datos procedentes tanto de los equipos a controlar como de la respuesta de los sensores a un experimento dado. El programa se encarga de generar varios ficheros de datos que contienen información acerca de las condiciones de medida, variables de control y, por supuesto, los resultados experimentales.

2.9. Estimación de las prestaciones

Además del diseño y más o menos complejo proceso de ensamblaje de todo el sistema, conviene realizar una estimación de las prestaciones del mismo con el fin de conocer cual es la precisión que puede obtenerse a partir de él. Esta estimación de las prestaciones se centra en tres puntos:

- a) Establecimiento de la concentración
- b) Control de los heaters
- c) Medida de las resistencias

a) Establecimiento de la concentración.

Para el sistema diseñado, la concentración de los gases queda determinada por la

ecuación 2.3 como relación entre el flujo de gas y el flujo total. Bajo esta premisa, el error absoluto en el establecimiento de la concentración viene dado por la siguiente ecuación:

$$\Delta C_i \approx \frac{\Delta f_i}{F_{\text{tot}}} + \frac{f_i}{F_{\text{tot}}^2} \Delta F_{\text{tot}} \quad (2.6)$$

Siendo F_{tot} la suma del flujo de gas f_i más el flujo de aire (f_{aire}). Dado que los errores en el flujo Δf_i y ΔF_{tot} ($\Delta F_{\text{tot}} = \Delta f_i + \Delta f_{\text{aire}} \approx \Delta f_{\text{aire}}$) son del 1% respecto del fondo de escala, es fácil derivar de esta ecuación que a mayor flujo, menor error relativo en el establecimiento de la concentración. La figura 2.16 muestra el error relativo en función de los flujos de los MFC.

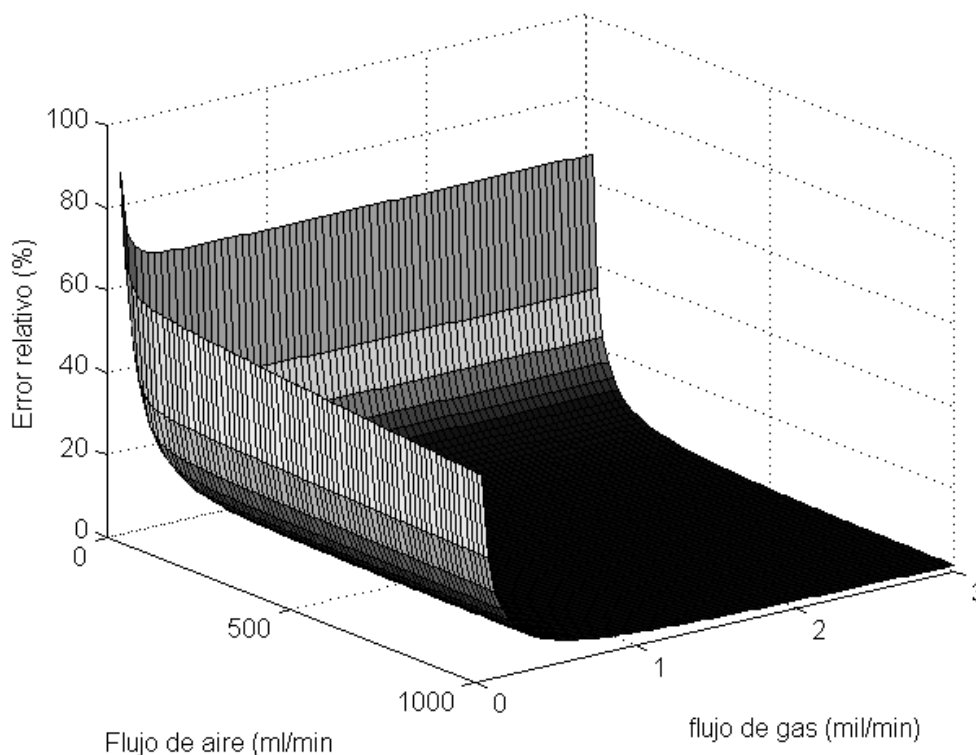


Figura 2.16 Error relativo en función de los flujos

A la vista de estos resultados, sería por tanto recomendable trabajar en el máximo rango de flujo de los MFC. Sin embargo, bajo esas condiciones no es posible alcanzar bajas concentraciones como las que se exigen para el CO por ejemplo. En estas circunstancias, hay que sacrificar precisión para obtener bajas concentraciones.

Las pruebas realizadas para metano con concentraciones desde 500 ppm hasta 10.000 ppm los errores en el establecimiento de dichas concentraciones son, como máximo del 8%, siendo el error medio de un 2%. Estos resultados se pueden extrapolar, con pequeñas variaciones, al resto de gases y son suficientemente precisos como para cumplir con las pruebas que se exigen en las normativas. Sin embargo, hay que remarcar que un error del 2% en el establecimiento de la concentración de 10.000 ppm supone un error en concentración de 200 ppm.

b) Control de los calefactores.

Se ha comprobado que la circuitería desarrollada para el control de los calefactores de

los sensores, permite obtener los niveles de tensión requeridos en cada canal y de forma independiente. Los errores relativos máximos que se obtienen son del orden del 4%, pero solo se dan para tensiones pequeñas, (por debajo de 0.5V); para valores habituales de tensión a aplicar a los calefactores, el error relativo disminuye hasta un 0.5%. El origen de este error puede asociarse a tres causas: la conversión analógico/digital (12 bits), el propio circuito y la longitud de las conexiones entre los calefactores y el control, que introduce un divisor de tensión variable.

c) Medida de las resistencias

En referencia al acondicionamiento de la señal de los sensores se ha comprobado que la medida de la resistencia de los sensores es fiable y está bien calibrada. A fin de establecer un orden de magnitud del error se realizaron pruebas con resistencias bien determinadas. Los resultados indican una buena correspondencia, con un error relativo de un 1% aproximadamente. El origen de este error puede atribuirse de nuevo a la inexactitud en la conversión analógico/digital, al circuito y a la longitud de las conexiones entre los sensores y el acondicionamiento.

Comprobado el funcionamiento del equipo, hay que remarcar que la limitación principal para la precisión radica en la capacidad de los MFC de generar mezclas gaseosas con suficiente precisión siendo el resto de errores de, como mínimo, un orden de magnitud inferiores.

La máxima precisión a la que se puede aspirar con un sistema basado en MFC de 1% de precisión es de 200 ppm para los gases metano y dióxido de carbono. Para el monóxido, ya que las concentraciones son menores, la precisión máxima a la que se puede aspirar es de 25 ppm.

Información más detallada sobre las prestaciones del sistema pueden encontrarse en la bibliografía¹⁶.

2.10. Sistema experimental de la universidad de Tübingen

Tal y como se mencionó al final del prólogo del presente capítulo, este trabajo es en parte fruto de la colaboración con el grupo del profesor Göpel de la Universidad de Tübingen. Toda una serie de datos, pertenecientes a análisis y estudios de compuestos VOC como tolueno u octano y que se utilizarán en los siguientes capítulos, provienen de un sistema experimental similar al descrito, desarrollado en Tübingen. Por tanto, para completar este capítulo de descripción experimental, conviene hacer una breve descripción de este sistema alternativo. En esencia este sistema experimental no difiere en su fundamento del que se ha presentado hasta ahora en este capítulo, constando también de las bombonas de gases a utilizar, controladores de flujo másico para controlar los flujos de los mismos y un ordenador dedicado a la tarea de controlar el sistema y medir las salidas de una matriz de sensores tal como se mostró en el esquema de la figura 2.4. Es precisamente en la matriz de sensores donde radica la diferencia fundamental ya que este sistema opta por el uso de sensores QMB.

Un conjunto de seis sensores QMB de frecuencia fundamental 10 MHz y recubierto de polímeros adsorbentes conforman la mencionada matriz. Los polímeros son: polidimetilsiloxano (PDMS), poli(cianopropil)metilsiloxano (PCPMS), polifenilmetilsiloxano (PPMS), poli(acido-isopropilcarboxílico)metilsiloxano (PiPCMS), poli-(aminopropil)metilsiloxano (PAPMS, 10% grupos amino) y poli[2-carboxi(D-valin-t-butilamida)propil]metilsiloxano. La matriz de sensores está introducida en una cámara de acero inoxidable estanca de 200 ml. La temperatura de la

cámara se mantiene a 303 °K gracias a una cinta calefactora dispuesta alrededor de la misma cámara y se controla dicha temperatura a través de una serie de sensores alojados en su interior. Las mezclas gaseosa se realizan con controladores MFC Tylan FC 280-S cuya precisión es también de 1% del fondo de escala y el flujo total habitual de trabajo se fija a 200 ml/min. Para vaporizar esencias, se dispone de un controlador de vapores (Tylan, Source V) en la línea de gases y, al igual que en el sistema anterior, toda línea es también de acero inoxidable. Se dispone además de sensores de humedad para el control de esta propiedad en la cámara, y para el control y la recolección de datos se utilizan tarjetas A/D de entrada-salida y un ordenador. Las prestaciones de este sistema son similares al anterior, y la precisión en las concentraciones que se generan dependen fundamentalmente de la precisión de los MFC. Para los gases que se utilizarán, la precisión se sitúa alrededor de las 100-200 ppm de gas.

Bibliografia

- ¹CENELEC, "Electrical apparatus for the detection of combustible gases in domestic premises," (1996).
- ²CENELEC, "Draft European standard guide for selection, installation, use and maintenance of domestic combustible gas detectors," (1996).
- ³CENELEC, "Electrical apparatus for the detection of carbon monoxide in domestic premises," (1996).
- ⁴norma italiana, "Requisiti generali per i rivelatori di gas d'uso domestico o similare," (1993).
- ⁵R.S. Barratt, "The preparation of Standard gas mixtures: A review," *The Analyst* **106**, 817-849 (1981).
- ⁶H.E. Endress, H. D. Jander, W. Göttler, "A test system for gas sensors", *Sensors and Actuators B*, **23**, (163-172) 1995
- ⁷Fisher Rosemount, "Mass Flow Controller Model 5850: Installation and Operating Manual," .
- ⁸H.J. Boer, "Mass Flow Controlled Evaporation System," *Journal de Physique IV* , c5-961, c5-966 (1995).
- ⁹FIS, "ST31 semi-conductor gas sensor," .
- ¹⁰FIS, "FIS Gas sensors for flammable gas detection," .
- ¹¹FIS, "Products review," .
- ¹²Analog Devices catalogue, "AD590 temperature sensor," .
- ¹³Philips catalogue, "humidity sensor," .
- ¹⁴Data Translation, "Data acquisition product catalog," .
- ¹⁵J.W. Gardner H.V. Shurmer, H.T. Chan, "The the application of discrimination techniques to alcohols and tobaccos using tin oxide sensors," *Sensors and Actuators* **18**, 361-371 (1989).
- ¹⁶A. Ortega "Sistema de Test para Sensores de Gases" (1998)
- ¹⁷S. Marco, A. Pardo, F. Davide, C. DiNatale, A. D'Amico, A. Hierlemann, J.Mitrovics, M. Schweizer, U. Weimar, W. Göpel, "Different strategies for identification of gas sensing systems," *Sensors and Actuators B* **34**, 213-223 (1996).
- ¹⁸F.Davide, C. DiNatale, A. D'Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, W. Göpel, S. Marco, A. Pardo,...., "Dynamic calibration of QMB polymer-coated sensors by Wiener kernel estimation," *Sensors and Actuators B* **26-27**, 275-285 (1995).
- ¹⁹E. Llobet, J. Brezmes and X. Vilanova, "Qualitative and quantitative analysis of volatile compounds using transient and stady-state responses of thick-film tin oxide gas sensor array", *Sensor and actuators B*, **41**, 13-21, 1997
- ²⁰B. Yea, T. Osaki, and R. Konishi, "The concentration-estimation of inflamabel gases with semiconductor gas sensor utilizing neural networks and fuzzy interference", *Sensor and actuatros B*, **41**, 121-129, 1997
- ²¹J. W. Gardner, T. Pearce and P. N. Bartlett, " A multisensor system for beer flavour monitoring using an array of conducting polymers and predictive classifiers", *Sensors and actuators B*, **18-19**, 240-243, 1994
- ²²A. Hierlemann, U. Weimar, G. Kraus, M. Schweizer and W. Göpel, "Polymer based sensor arrays and multicomponent analysis for the detection of hazardous organic vapours in the environment", *Sensor and Actuators B*, **26-27** 126-134, 1995
- ²³G.O. Nelson, "Gas Mixtures: Preparation and Control", Lewis Publishers, (1992) ISBN 0-87371-298-6

Capítulo 3

Metodología y diseño de experimentos: Resultados de calibración estática

- *Así pues, el general Kerebron, después de haber contemplado el problema con una metodología muy precisa, descubrió tres clases de dragones: a saber, los nulos, los imaginarios y los negativos. Como ya se ha dicho, los dragones no existen, pero cada uno tiene un tipo de inexistencia distinto.*

Stanislav Lem: Fábulas de robots

- *Los primeros intentos por medir cuantitativamente el tiempo de solución de la tiotimolina toparon con dificultades considerables debido a la propia naturaleza negativa de la estimación. El hecho de que el producto químico se disolviera antes de la adición de agua, hacía que el paso lógico y natural siguiente fuera el de retirar el agua después de producirse la disolución y antes de la adición. Lo cual, afortunadamente para la ley de conservación de la masa-energía, no ocurrió jamás, puesto que la disolución nunca se producía si después no había de verificarse la adición de agua.*

Isaac Asimov: Las propiedades endocrónicas de la tiotimolina resublimada

Prólogo del capítulo

La cita de Asimov presentada en la página anterior está extraída de un cuento corto aparecido a finales de los años 40. El cuento simula un artículo científico en el que se estudian las supuestas propiedades de un compuesto denominado tiotimolina, que se disuelve en agua antes de haber echado propiamente el agua. Curiosamente, si sólo se hace el ademán de aportar agua pero luego esta no se suministra, la tiotimolina no se disuelve. A imitación del estilo de los artículos científicos clásicos, se presentan en el cuento una serie de mediciones, gráficas y tablas que justifican las observaciones. Desde luego no existe ninguna sustancia con semejantes propiedades, pero lo que se pretende remarcar con esta cita es que la medición cuantitativa, por complicada que sea, permite comprender a un nivel muy profundo el comportamiento de los fenómenos físicos.

Además, si el fenómeno a medir presenta un comportamiento que puede aproximarse a una relación lineal, entonces, el tratamiento de los resultados no resulta, en principio, excesivamente complejo ya que la modelización lineal es un tema relativamente bien estudiado para la cual se dispone de una amplia batería de potentes herramientas técnicas y matemáticas. Sin embargo, ante un problema de modelización no lineal, la dificultad aumenta y las herramientas para el tratamiento de los datos no lineales son de considerable complejidad. Si a esta complejidad, se añade el hecho de tratar los datos no desde la concepción estática sino desde la dinámica, entonces el tratamiento de estos fenómenos se hace sumamente complejo, siendo este, además, un terreno relativamente poco investigado en comparación con el resto de casos. La tabla 3.1 pretende ilustrar con ejemplos esta discusión. Si se trata de estudiar modelos estáticos y lineales, la calibración (entendida como relación entrada-salida) del sistema (en particular, en la tabla se habla de un sensor de presión) suele ser facilitada por el fabricante y no entraña dificultad. El paso de la calibración estática lineal a la no lineal implica un cierto esfuerzo adicional pero factible de atacar con buen grado de aproximación; por ejemplo, la dependencia de la resistencia de una resistencia NTC con la temperatura es bien conocida y facilitada también por el fabricante. Al mismo nivel puede hablarse en el caso de calibración dinámica lineal donde, aunque el proceso sigue una vía distinta a la opción estática, tampoco entraña una excesiva dificultad, por ejemplo las funciones de transferencia de un sensor como el acelerómetro. En todos estos casos, los sistemas son relativamente fáciles de calibrar, de tal manera que el fabricante suele ofrecer datos, tablas, curvas o fórmulas relativas al problema. Sin embargo, el campo de la calibración dinámica no lineal pertenece al campo de la investigación, y ningún fabricante de sensores o sistemas ofrece información sobre calibraciones en este sentido. Es precisamente en este punto donde este trabajo pretende hacer una contribución.

Tabla 3. 1	Lineal	No Lineal
Estático	Sensor de presión: $V=S \cdot P + V_0$	Termistor $R=f(T)$
Dinámico	Acelerómetro $Y(s)=H(s)X(s)$?

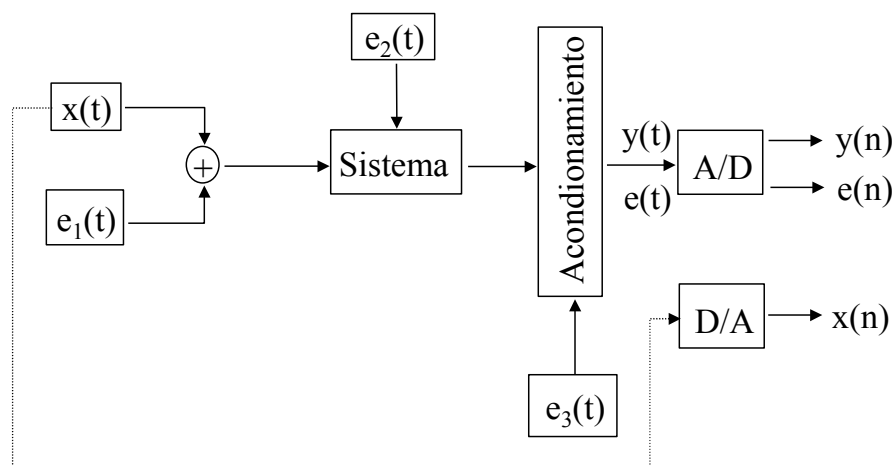
La complejidad en el tratamiento de estos fenómenos aún queda más enfatizada si se permite introducir en el problema un factor de multidimensionalidad. En el ámbito que ocupa a este trabajo, esta multidimensionalidad se introduce a través de experimentos que involucran a varios gases y varios sensores a la vez.

En este capítulo y los sucesivos se va tratar de reunir los conceptos teóricos esenciales relativos a la modelización de sistemas dinámicos así como su aplicación para los datos experimentales relativos al particular problema del sistema de test y medida expuesto en el capítulo 2. Sin embargo, ambos bloques no se han separado completamente sino que a medida que se avanza en la descripción teórica, se presentan los resultados, de tal manera que se obtiene, inmediatamente una visión completa teórico-práctica del problema.

3.1. Introducción a la teoría de sistemas: Conceptos básicos

El término “sistema” se emplea con abundancia en multitud de disciplinas: sistema de medida, sistemas de clases, sistema de gobierno, sistema nervioso, etc. Esencialmente, el concepto representa a un ente complejo (ya sea un procedimiento, un dispositivo, una organización,...) formado por elementos muy heterogéneos, pero cada disciplina que utiliza el término, incorpora matices diferentes a su significado. El resultado es que el sentido queda diluido y la palabra sistema se usa habitualmente como sustantivo comodín con el que calificar de forma general cualquier relación, simple o compleja, de variables. Desde el punto de vista de la teoría de identificación de sistemas, y sin entrar de momento en rigurosidad matemática, efectivamente un sistema es esencialmente una relación entre ciertas variables que están asociadas a algún procedimiento. El objetivo de la identificación es realizar estimaciones sobre dicha relación.

El sistema que se pretende estudiar en este trabajo es un equipo de test de sensores como el presentado en el capítulo 2. Un sistema de test como el presentado y una nariz electrónica presentan claras analogías en cuanto a la construcción del sistema. De hecho, esencialmente son iguales. La única diferencia entre ambos radica en la toma de muestras. Mientras en el sistema de test, la muestra viene determinada explícitamente por el experimento y el experimentador, la nariz electrónica, en teoría, debe tomar muestras activamente y por si misma (la forma habitual es a través de una bomba de succión) a partir de la atmósfera que la rodea. Hecha esta salvedad, narices electrónicas o sistemas de análisis de gas son sistemas análogos. Como ya se ha dicho, el esquema físico del sistema ya ha sido plasmado y descrito el capítulo 2. El siguiente esquema presenta dicho sistema en diagrama de bloques y desde el punto de vista del tratamiento de señal.



$x(t)$: Señal de concentración	$e_1(t)$: Ruido de establecimiento de la concentración
$y(t)$: Respuesta del sistema a $x(t)$	$e_2(t)$: Ruido en la medida
$x(n)$: $x(t)$ digital	$e_3(t)$: Ruido del acondicionamiento
$y(n)$: $y(t)$ digital	$e(t)$: Ruido total
$e(n)$: Ruido total digital	

Figura 3. 1 Diagrama de bloques de las señales del sistema

La señal de concentración establecida a través de los MFC se representa mediante $x(t)$. El establecimiento de dicha concentración, como se explica en el capítulo anterior (ecuación 2.2), se realiza a través de la relación de flujos provenientes de los MFC, ya que de hecho, las entradas del sistema son flujos. El sistema recibe la señal de concentración y ofrece una respuesta a dicha entrada, que debidamente acondicionada, se convierte en una señal $y(t)$. Esta señal es recogida finalmente a través de un convertidor A/D, que la convierte en una señal discreta $y(n)$. A lo largo de todo el proceso, diferentes entradas de ruido afectan la medida y aparece al final una cierta señal de ruido total $e(t)$, que el proceso A/D convierte en $e(n)$. La señal de concentración, de hecho se establece también de forma digital, es decir, en realidad es también una señal discreta $x(n)$. La identificación y modelización del sistema deben hacerse, pues, a partir de las señales $x(n)$, $y(n)$ y el conocimiento de ciertas propiedades de la señal $e(n)$; ya que aunque a través del experimento no se dispone de la propia señal $e(n)$ si que es posible acceder a sus propiedades a través de determinadas operaciones de tratamiento de la señal. El esquema de la modelización que se pretende llevar a cabo se presenta en la siguiente figura:

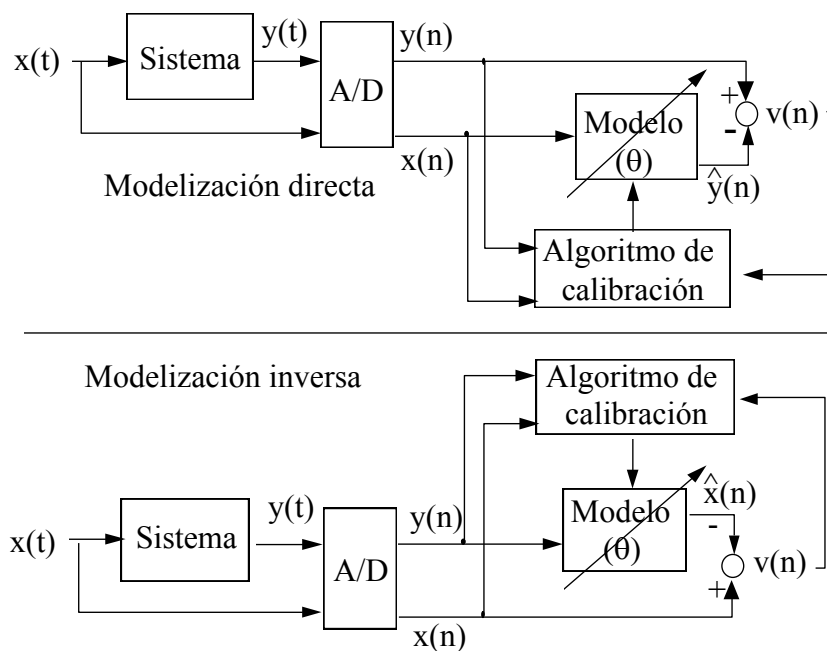


Figura 3. 2 Esquemas del sistema de modelado directo e inverso

Tal y como ya se ha dicho, el problema que se plantea con la identificación de sistemas es el de estimar la relación entre las señales $y(n)$ y $x(n)$. Este problema se puede abordar bajo dos filosofías diferentes: la identificación directa y la identificación inversa.

En la identificación y modelado directo se trata de encontrar un modelo con el cual poder predecir las salidas del sistema a partir de una entradas dadas. Para ello el modelo utiliza la entrada y una serie de parámetros para ofrecer una estimación de la salida. Los parámetros se modifican partir de una cierta señal de error (diferencia entra la salida del modelo y la salida real) a través de cierto algoritmo de calibración que busca, de alguna forma, minimizar ese error.

En la identificación y modelado inverso, el objetivo es conocer la señal de concentración a partir de la salida del sistema; es decir, es un sistema de medida de

concentraciones. Para ello, el modelo utiliza en este caso, la salida del sistema que operada con los parámetros (sean los que sean) del modelo inverso, ofrece una estimación de la señal de entrada. Igual que en el modelado directo, los parámetros se modifican y adaptan a partir de una cierta señal de error a través de un cierto algoritmo de calibración que busca el error mínimo.

Abundando en el tema de las características del sistema, es claro que a causa del convertidor A/D, el sistema con el que se está tratando es un sistema de señales discretas. Se dice entonces que el sistema es discreto; pero dependiendo del comportamiento de este en su entorno de trabajo puede hablarse de sistemas estáticos o dinámicos, lineales o no lineales, variantes o invariantes en el tiempo, etc^{11 12}.

Habitualmente, la teoría clásica de sistemas se centra en el estudio de sistemas lineales e invariantes en el tiempo (LTI), aunque, como se demostrará mas adelante, el orden lineal no es suficiente para explicar completamente el problema que se plantea. Por otro lado, y atendiendo al número de entradas y salidas que presentan los sistemas, estos se clasifican en:

- Sistemas SISO (Single Input Single Output)
- Sistemas MISO (Multiple Input Single Output)
- Sistemas SIMO (Single Input Multiple Output) y
- Sistemas MIMO (Multiple Input Multiple Output).

Las figuras 3.1 y 3.2 representan sistemas SISO, pero de hecho, la estación de test que se utiliza no es de una sola entrada y una solo salida. Cada sensor de la estación de medida recibe una serie de magnitudes o señales de entrada definidas a través del vector $\mathbf{x} = (x_1, x_2 \dots x_n)$ que describe las concentraciones de n gases. Además el sistema de medida ofrece una salida en forma de señal eléctrica y_i para cada sensor i. En este sentido, aunque el conjunto de matriz de sensores más cámara sensorial es un sistema MIMO, cada sensor de gas en el entorno de la cámara sensorial debe ser tratado como un sistema MISO causal, discreto, dinámico e invariante en el tiempo.

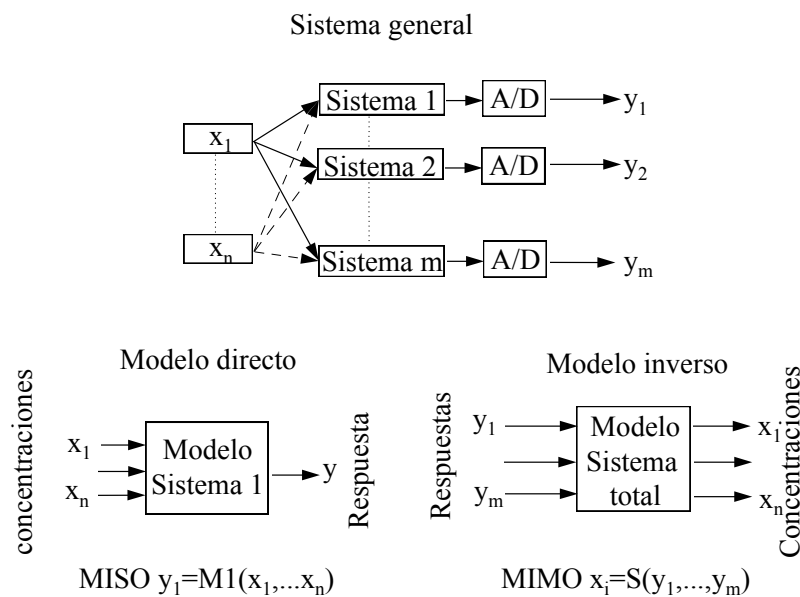


Figura 3.3 Esquema de las estructuras del sistema

Siguiendo un desarrollo desde lo sencillo hacia lo complejo, inicialmente se utilizarán técnicas lineales para generar modelos de primera aproximación y se considerará, a partir de su rendimiento, la conveniencia de utilizar estrategias no lineales.

Si se adopta un proceso identificativo de sentido inverso, planteando la estimación de las concentraciones x a partir del conocimiento del conjunto de señales $y=[y_1, \dots, y_m]$ de m sensores, la identificación debe centrarse en un sistema MIMO y acausal. Más adelante se desarrollarán estos aspectos y se profundizará en las dificultades del tratamiento de sistemas acausales y MIMO, por otra parte inevitables al tratar un sistema inverso a partir del directo.

En cualquier caso, en la pretensión de estudiar el comportamiento del sistema sensor ya sea cuantitativamente o cualitativamente, el propósito principal es generar modelos matemáticos útiles de los datos obtenidos a partir de una serie de experiencias realizadas sobre una cámara sensorial, entendiendo por cámara sensorial al conjunto de sensores más la cámara donde se alojan. El problema no es trivial y en función de la información que se desee extraer, los experimentos con el sistema deben enfocarse de diferentes formas. Por ejemplo, para obtener un buen modelo estático, se requiere un tipo de experimento muy específico y no del todo útil para generar modelos dinámicos. Por otra parte, la información que se extrae de un modelo estático no es la misma que la que se obtiene de un modelo dinámico. Con experimentos orientados hacia la extracción de información estática es necesario un tiempo mínimo de exposición del sistema a las concentraciones. Este tiempo depende de la velocidad de respuesta del sensor en el entorno de la cámara donde se aloja, es decir, del tiempo de respuesta de la cámara sensorial. Este tiempo puede ser más o menos largo, pero, por otro lado, un sistema de medición gaseosa, al igual que un sistema olfativo natural, no tiene que permanecer expuesto mucho tiempo a una atmósfera. En cuanto detecta la presencia de un cierto componente gaseoso, debe iniciar el proceso de reconocimiento y medida inmediatamente, obteniendo conclusiones en poco tiempo. En este sentido, todo sistema de medición gaseosa que, por ejemplo, intente imitar este comportamiento natural, debe ser sometido a un estudio dinámico.

La identificación de sistemas multivariables lineales dinámicos e invariantes en el tiempo se viene aplicando en diversos campos como biología, econometría, control, etc^{1 2 7}. Sin embargo, su aplicación al análisis de mezclas gaseosas se ha visto restringida y sólo en estos últimos años ha recibido mayor atención. Este hecho puede atribuirse a la complejidad, ya comentada, del problema y que es inherente a los sensores de gas y sus problemas de histéresis, derivas temporales y/o no linealidades, problemas estos que no suelen presentarse en sistemas más sencillos. La mayoría de trabajos centrados en mediciones cuantitativas de gas se basan en experimentos estáticos^{8 9 10 13 14 15}. Por ejemplo, algunos trabajos⁸ tratan de caracterizar el comportamiento del sensor, cuando éste ha sido sometido a un escalón de concentración. Para ello se ejecuta el experimento (habitualmente consiste en introducir muestra líquida en la cámara sensorial, evaporarla y homogeneizar la concentración mediante un ventilador) y se espera a llegar a un régimen permanente para tomar las medidas. Otros trabajos utilizan ciclos de temperatura, generados a través del heater, para establecer las características de los sensores. El sensor se introduce en la cámara sensorial donde se inyectan las muestras gaseosas, y de nuevo se ha de esperar a que se establezca la respuesta estacionaria. Todos estos estudios, aunque perfectamente válidos, no son capaces de comprender la dinámica del sistema, aunque tampoco es ese su propósito. En ellos, se plantean ciertas mediciones dinámicas pero desde el punto de vista de obtener información discriminatoria adicional en un proceso de reconocimiento de patrones, pero no para la

confección de una calibración dinámica. El objetivo que se persigue en este trabajo, tal y como ya se ha reiterado, es la calibración o modelización dinámica^{16,17} puesto que, a pesar de la dificultad que entraña, debe enfatizarse las ventajas que los modelos dinámicos proporcionan:

Posibilitan el diseño de controles óptimos que tengan en cuenta la dinámica: control de sistemas

Alcanzan muy buenas prestaciones en la simulación de las respuestas sensoriales ante entradas arbitrarias: descripción de sistemas

Consiguen predecir de respuestas futuras a partir del conocimiento de presente y pasado: predicción

Permiten estimar con precisión las señales de entrada del sistema a partir de las salidas del mismo: el problema de la medida

3.2. Modelos de sistemas

La palabra modelo proviene del latín y originalmente significaba molde o patrón. Efectivamente, los modelos son una especie de patrones que contienen una aproximación más o menos ajustada a las pautas de comportamiento de los sistemas. Por supuesto, el grado de sofisticación de los modelos es muy variable y puede abarcar desde un puro conocimiento cualitativo hasta un formalismo matemático sumamente sofisticado. Así, existe los modelos mentales, que engloba todo modelo de conocimiento intuitivo, basado en la experiencia y no expresado en forma matemática, o los modelos verbales que se articulan en sentencias del tipo “Si ocurre la causa ‘x’, entonces se produce el efecto ‘y’”. Abundando en la enumeración de los diferentes tipos de modelos, es, por otra parte, posible que un sistema pueda quedar perfectamente caracterizado con una tabla o una gráfica. Esto suele ser así para los sistemas lineales, que quedan perfectamente descritos mediante la respuesta del sistema a un impulso mientras que para aplicaciones más complejas, como es el caso del presente trabajo, puede ser necesario describir las relaciones entre las variables en términos de expresiones matemáticas más elaboradas que van más allá del orden lineal.

Hay que tener siempre presente en esta descripción que un modelo es una abstracción del comportamiento del sistema real. Un modelo representa al sistema real en la medida que explique el comportamiento del mismo; y será tanto más perfecto cuanto más se ajuste a los resultados reales del sistema. La validez del modelo se basa precisamente en esta verificación, comparando los resultados extraídos a partir del modelo con los resultados reales.

La generación de modelos puede seguir dos procesos diferentes aunque no excluyentes mutuamente. Una de estas vías de generación de modelos es la experimentación. A través del resultado de la experimentación y analizando adecuadamente los datos de entrada y salida del sistema, es posible realizar una estimación de las relaciones entre las magnitudes del mismo, obteniéndose un modelo identificativo. Sin embargo, en algunos casos no resulta adecuado o es totalmente imposible realizar experiencia alguna con un sistema. Por ejemplo, puede resultar económicamente inviable, o demasiado peligroso o

simplemente el sistema no existe más que a nivel de proyecto o prototipo. En las circunstancias en las que la experimentación no es válida, el conocer las relaciones entre las magnitudes del sistema se realiza a través de la modelización física.

Bajo esta construcción subyacen las leyes físicas o matemáticas conocidas que rigen el comportamiento de dicho sistema. Consiste en esencia el modelado físico en descomponer el sistema en subsistemas más simples y bien descritos. El modelo completo se genera engarzando los modelos de los múltiples subsistemas mediante la combinación matemática adecuada.

Ambos caminos no son, como ya se ha dicho, excluyentes sino que en muchos casos se combinan en el proceso de generación de un modelo adecuado para el sistema.

No siempre es sencillo realizar una aproximación al problema de generar modelos para sistemas a través del modelado físico. En muchos casos, los procesos que tienen lugar en los sistemas son de tal complejidad que hace imposible obtener modelos razonablemente adecuados (cualquiera que sea el criterio de “razonabilidad”) mediante leyes físicas. La identificación de sistemas sí permite obtener modelos razonablemente adecuados allá donde la complejidad o la falta de información hacen no factible el modelado físico. Sin embargo, esta cualidad de la identificación aparece a costa de obtener modelos con una validez limitada, en el sentido de que los

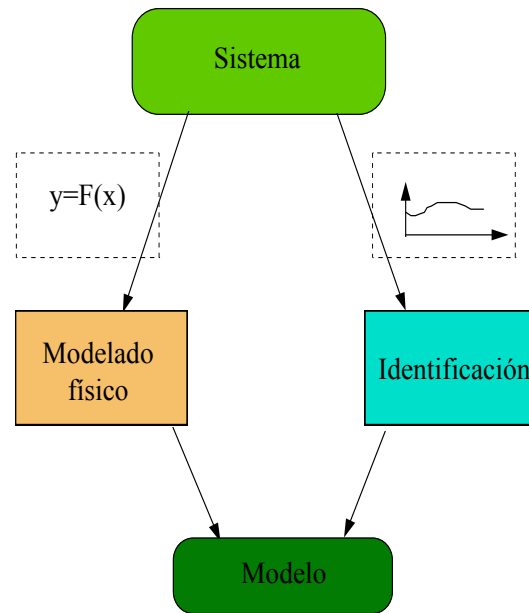


Figura 3.4 Procesos de generación de modelos

modelos generados mediante identificación adolecen de una cierta carencia de generalidad y de sentido físico. Efectivamente, los modelos generados mediante identificación contienen poca información acerca de la física del sistema y, en consecuencia, no ofrecen un significado físico claro. Además están limitados a las condiciones experimentales utilizadas y los resultados son de difícil extrapolación. Sin embargo, por otro lado, son relativamente fáciles de construir y usar.

La estrategia de este trabajo se centra en realizar estimaciones a través de modelos obtenidos por identificación. Por tanto, establecidos en el método identificativo, debe quedar claro que a pesar de su relativa sencillez, en comparación con el modelado físico, conseguir un modelo apropiado no siempre es fácil. Se pueden plantear varios problemas que dificulten la tarea identificativa:

- La dinámica del sistema puede ser no lineal;
- Los datos a analizar contienen ruido que, de alguna forma, debe tenerse en cuenta.
- Los procesos pueden derivar en el tiempo, y este hecho causa problemas si se intenta describir el sistema a través de modelos invariantes en el tiempo.

3.3. El proceso identificativo

La construcción de modelos mediante identificación de sistemas precisa de cuatro elementos básicos: los datos, un conjunto de posibles modelos, algún tipo de algoritmo

mediante la cual ajustar los modelos usando los datos y la validación del modelo generado. Asociados a estos elementos, en el proceso de la identificación se utilizan una serie de herramientas como el procesado de señal, diferentes rutinas de identificación para la generación de diferentes tipos de modelos, procedimientos de presentación (diagramas de Bode, representación de polos y ceros, etc.), diferentes fórmulas para el análisis y la comparación de distintos modelos, y algoritmos de cálculo de residuos y errores de la estimación.

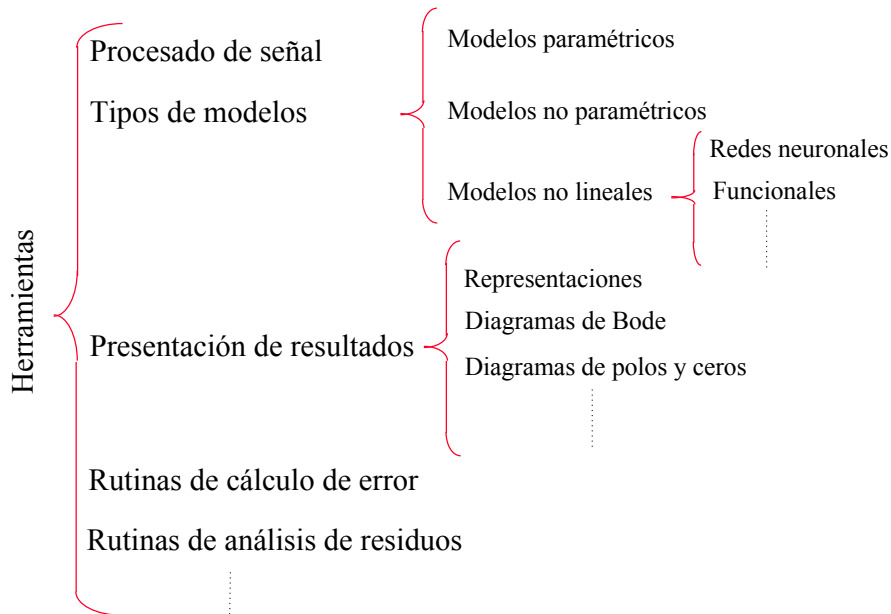


Figura 3. 5 Herramientas habituales en el proceso identificativo

Para completar el proceso identificativo que genera el modelo, los cuatro elementos mencionados deben adecuarse a un cierto proceso sistemático. Dicho proceso está esquematizado en la siguiente figura donde se pretende reflejar el conjunto de pasos a realizar en la identificación de un sistema.³

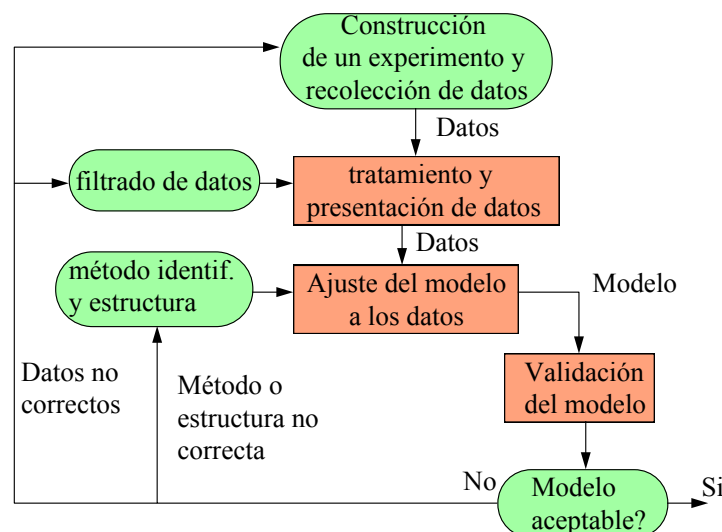


Figura 3. 6 Ciclo identificativo. Los rectángulos representan procesos de computación. Los óvalos representan elecciones del experimentador.

El primer paso de este esquema consiste en generar un conjunto de datos sobre el que realizar la identificación. Para ello es preciso diseñar un experimento.

Una etapa fundamental dentro del cuadro de los experimentos iniciales consiste en realizar las pruebas necesarias para determinar la linealidad del sistema. Establecido este punto, debe continuarse con el diseño del experimento tal que las condiciones sean lo más similares posibles a las condiciones en las cuales el modelo va a ser usado. Por otro lado, las señales de entrada del sistema deben excitar el sistema y deben excitarlo de manera que se pueda extraer la máxima información posible³. Por ejemplo, una señal sinusoidal de frecuencia w proporciona información sobre la respuesta del sistema para esa frecuencia; pero muchos sistemas responderían de manera similar a esta entrada. Es importante, por tanto que la señal o señales de entrada contengan, de alguna forma, la información necesaria y suficiente para que el sistema ofrezca una señal o señales de salida que describan al sistema lo más completamente posible. Si bien este razonamiento es más que obvio, llevarlo a la práctica no siempre resulta sencillo. Una buena elección como señal de entrada en general es la señal de ruido blanco porque este tipo de señales contiene todas las frecuencias posibles y por tanto testea al sistema en todas ellas. Para sistemas lineales basta con una señal aleatoria binaria, es decir que oscile aleatoriamente entre dos niveles. Para sistemas no lineales es conveniente disponer de mayor número de niveles^{4,7}.

Una vez los datos han sido recolectados, resulta útil realizar una primera inspección de los mismos a fin de tener una impresión general del resultado del experimento, de detectar la existencia o no de deficiencias en los datos y de determinar si resulta conveniente aplicar algún tratamiento a dichos datos antes de utilizarlos en procesos posteriores. Deficiencias típicas que se suelen encontrar son: derivas en los niveles de las señales, distorsiones en las señales debidas a ruido y pérdidas de datos o datos erróneos. La inspección inicial debe permitir escoger aquellos datos que parezcan “apropiados” para generar modelos, despreciando los datos defectuosos (“outliers”) reservándolos para su uso a posteriori siempre que algún tipo de tratamiento reduzca los efectos perniciosos de los mencionados defectos. Los datos que son claramente erróneos o las cadenas de datos que han perdido parte de los mismos suelen tener consecuencias desastrosas en la modelización, por ello lo mejor es depreciosarlos directamente.

A partir de los datos considerados como adecuados, se generan modelos para aproximar el comportamiento del sistema. El volumen de datos a utilizar para estimar modelos mediante identificación determina en buena medida el rendimiento del proceso a seguir. Es conveniente utilizar un número suficientemente grande para poder trabajar con buena precisión, pero no hay que utilizar tantos datos como para ralentizar todo el proceso de cálculo sin obtener apreciables mejoras. Esta limitación a la baja viene también apoyada por el hecho experimental ya que disponer de muchos datos es una cuestión de tiempo. Si el experimento es lo suficientemente largo, se dispondrá de una buena cantidad de datos, pero esta situación no es habitualmente factible. A menudo no es posible trabajar (por ser costoso en tiempo y dinero) con experimentos dilatados en el tiempo, sino que se precisa obtener una serie de resultados rápidos aunque fiables (por ejemplo en procesos de calibración). Así, la cantidad de datos que se van a utilizar habitualmente en los procesos de identificación que se siguen en este trabajo, va a estar limitada a una cifra alrededor de 300 datos (equivalente a un experimento de unas 5 horas), aunque el experimento completo contenga en general un volumen de datos unas cuatro veces superior.

Los métodos de modelización que se usan con los datos son muchos y muy variados, cada problema debe tratarse particularmente y estudiar qué tipo de modelo se adecua

mejor teniendo en cuenta no tan solo un criterio de error mínimo sino también un criterio de simplicidad clásica: de entre dos modelos de rendimiento similar siempre es mejor escoger como modelo válido el más sencillo. Los modelos generados deben, posteriormente, sufrir un proceso de confirmación de sus resultados. Esta prueba *a posteriori* que asegura la adecuación del modelo es lo que se denomina validación del modelo. La validación es el proceso que certifica que el modelo generado es básicamente correcto y que contiene un cierto grado de generalidad. Sin este paso final, del modelo sólo se podría decir que se ajusta a los datos estrictamente usados para generarlo.

El primer paso en la validación que se realiza en este trabajo es siempre una comparación entre los datos reales y los estimados. En función de esta comparación se puede extraer un residuo a partir de las diferencias entre ambas. Este dato, se puede además cuantificar, por ejemplo en forma de potencia del residuo en relación con la potencia de la señal real. Este dato da una indicación de cuan bueno es el modelo. En cualquier caso, es necesario realizar un segundo paso de validación en el cual se efectúa un estudio más profundo del residuo, es el denominado análisis del residuo donde se estiman las relaciones del propio residuo con las señales del sistema. No siempre el análisis del residuo resulta determinante en la decisión acerca del buen rendimiento de un modelo, pero junto con la comparación realizada en el primer paso de validación, ambos estudios combinados aportan una idea muy clarificadora sobre el asunto.

Hay que resaltar, sobre todo, que resulta fundamental para el desarrollo de los procesos de análisis de residuos y validación, que los datos utilizados sean diferentes a los usados en el proceso identificativo. De esta forma aseguramos que la validación tenga sentido general. Si para estos nuevos datos el resultado es satisfactorio (en los términos que se describa el grado de satisfacción), el modelo puede ser dado como válido; de otro modo, el modelo debe ser rechazado e investigar las causas del fracaso (principalmente, la arquitectura del modelo y los datos utilizados). La validez de un modelo, sin embargo, no debe aceptarse como la verdadera explicación del comportamiento del sistema (que por otra parte tampoco es el objetivo que se persigue), sino que es una buena aproximación (siempre susceptible de ser mejorada) a ciertos aspectos del sistema que son de interés.

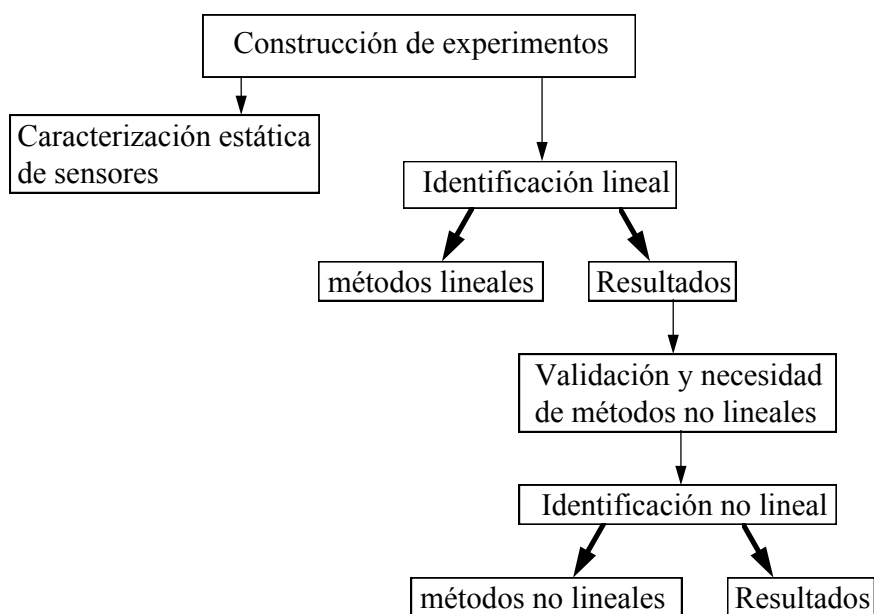


Figura 3. 7 Diagrama de desarrollo

A partir de este momento, se seguirá una evolución similar a la de la figura 3.6 para la identificación de un sistema sensor a partir de los datos de matrices de sensores de gas. El desarrollo del resto del capítulo y posteriores se esquematiza en el diagrama de la figura 3.17.

A continuación, por tanto, se va a proceder a la descripción de los experimentos usados para la identificación y a la presentación de los resultados. Acto seguido, se desarrollarán los diferentes aspectos teóricos utilizados así como el resultado de su aplicación.

3.4. Construcción de Experimentos

Tal como se ha apuntado, debe iniciarse el proceso identificativo con la generación de una serie de experimentos que permitan extraer las características interesantes de los sistemas sensores, entendidos estos, como siempre, como el sensor dentro de la cámara. El papel de la cámara en el proceso no es en absoluto despreciable, ya que influye en la dinámica de la respuesta irremediablemente, falseando, en cierta forma, la determinación de la respuesta debida estrictamente al sensor. Por otra parte, la presencia de una cámara (con ventilación forzada o no) donde se alojan los sensores es absolutamente necesaria, por lo que resulta no solo adecuado sino congruente, tenerla en cuenta en el proceso identificativo. Los experimentos que se presentan a continuación han sido diseñados para identificar un sistema de sensores QMB de frecuencia fundamental de oscilación 10 MHz. El funcionamiento fundamental de estos sensores ya se ha descrito en el capítulo 1 y una mayor información de las características específicas de los sensores utilizados puede encontrarse en la literatura ⁵.

Los gases que se utilizan son dos: octano y tolueno, pertenecientes al grupo de VOCs, el primero altamente explosivo, el segundo potencialmente nocivo para la salud y ambos presentes con frecuencia en el que hacer cotidiano; por ejemplo, el octano es uno de los componentes fundamentales de la gasolina y el tolueno es un disolvente de uso común en lacas y resinas y también es un componente habitual en la manufactura de colorantes.

Tabla 3. 2: Sensores del sistema de medida	
Notación	Polímero
Sensor 1	polydimethylsiloxaano (PDMS, Wacker-Chemie)
Sensor 2	poly(cyanopropyl)methylsiloxano (PCPMS, Hüls-Petrarch)
Sensor 3	Polyphenylmethylsiloxane (PPMS, Macherey-Nagel, OV-17)
Sensor 4	Poly(isopropylcarboxylic acid)methylsiloxane (PiPCMS, University of Tübingen)
Sensor 5	poly(aminopropyl)methylsiloxano (PAMPS 10% amino-groups, Universidad de Tübingen)
Sensor 6	poly[2-carboxy(D-valin-t-butylamide)propyl]methyl-siloxano (Chrasil-Val [®] , Universidad de Tübingen)

Los resultados experimentales de la aplicación de estos gases se van a expresar sobre cuatro sensores de una matriz de seis. Estos cuatro sensores, como veremos a continuación, son los que responden adecuadamente, de una u otra forma, a los

experimentos efectuados y descritos a continuación. La tabla 3.2 presenta el listado de los seis sensores de la matriz. En la primera columna, se encuentra la notación que se va a utilizar para denotar a cada sensor en el trabajo. La segunda columna especifica el polímero asociado a cada uno de ellos y entre paréntesis se encierra el acrónimo del material y el productor del mismo.

Así, siguiendo la discusión planteada al inicio de este capítulo, el sistema experimental se comporta, en cuanto a dimensionalidad, como un sistema MIMO de 2 entradas y 4 salidas.

Los experimentos esenciales realizados con estos sensores son cuatro, y en todos ellos, la frecuencia de muestreo es de una muestra por minuto:

Experimento 1:

Este es un experimento sencillo y que resulta extremadamente revelador para comprender el comportamiento del sistema. Consiste en someter los sensores a diferentes escalones de concentración. El estudio de la respuesta permite observar, fundamentalmente, la linealidad, el tiempo de respuesta, la histéresis y las derivas de cero. El experimento diseñado somete a la matriz de sensores a una batería de escalones de concentración crecientes para, una vez llegado un escalón máximo, descender de nuevo con algunos escalones de concentración más, de amplitud decreciente. La utilidad de estos últimos es comprobar la existencia de fenómenos de histéresis.

Para el gas octano se han generado, tal como muestra la figura, 15 escalones de concentración creciente desde 1000 ppm hasta 15000 ppm en saltos de 1000 en 1000 ppm. A partir del escalón de 15000 ppm se generan cuatro escalones más: de 10000, 7000, 5000 y 1000 ppm. Cada escalón tiene una duración en el tiempo de 30 minutos y entre escalón y escalón transcurre un tiempo también de 30 minutos durante el cual el sistema sólo recibe flujo de gas portador (aire sintético). Inicialmente hay un periodo de 120 minutos con el sistema en estado latente para establecer el nivel de base.

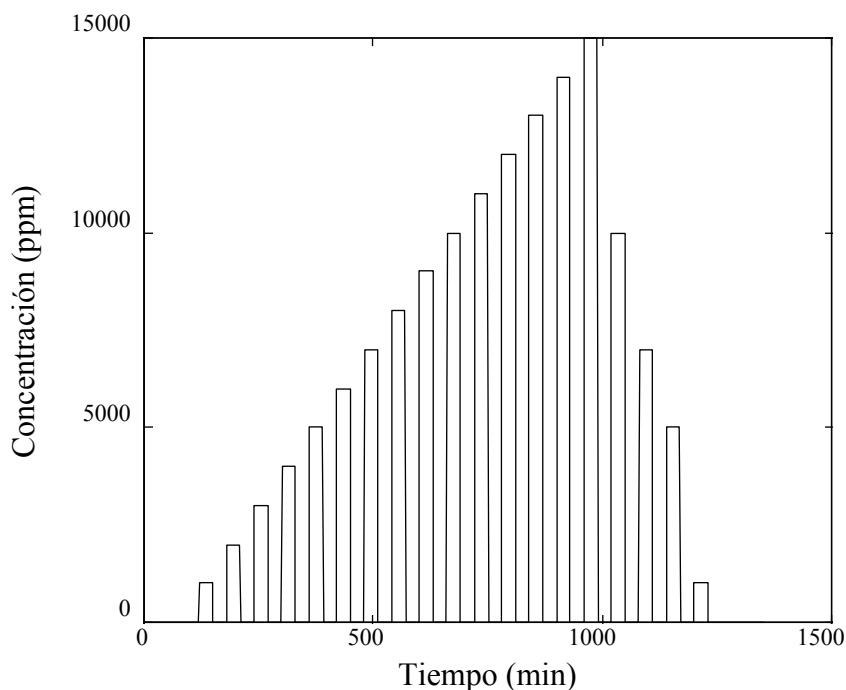


Figura 3. 8 Concentración de octano

Para gas tolueno se prepara un experimento similar. En este caso, después de llegar a 15000 ppm, se generan cinco escalones más, de concentración 17000, 20000, 22000, 25000 y 28000 ppm, descendiendo de nuevo con escalones de concentración 20000, 1000, 7000, 5000 y 1000 ppm. Los intervalos temporales se mantienen, igual que en el experimento con gas octano a 30 minutos. Se debe recordar que en este experimento y en los siguientes, la toma de datos se realiza a frecuencia de una muestra por minuto.

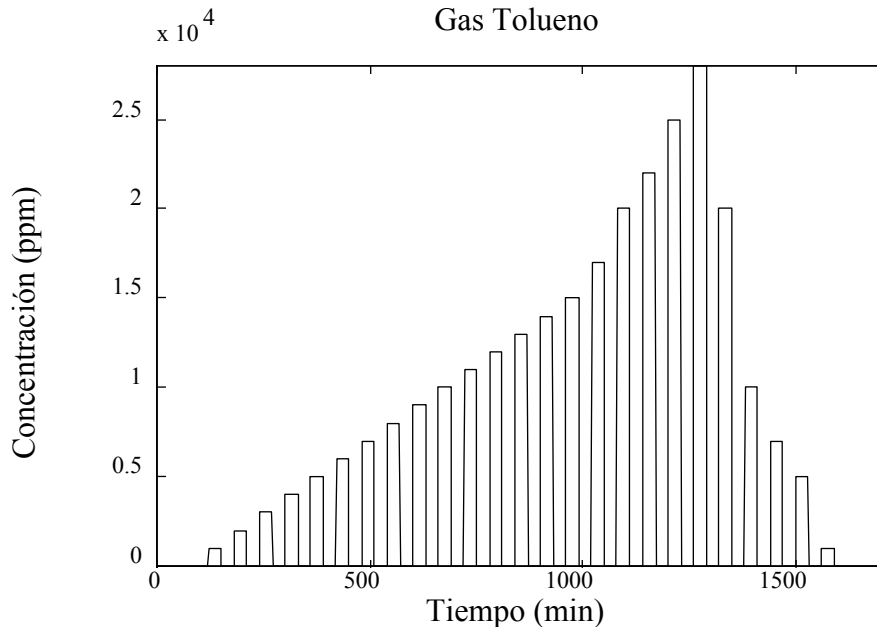


Figura 3. 9 Concentración de tolueno

Experimento 2:

Este experimento está enfocado hacia la aplicación de la teoría de identificación de sistemas. Para este propósito deben generarse señales de concentración que varíen aleatoriamente. En particular, resulta interesante obtener señales en forma de concentración que varíen como ruido blanco con función densidad de probabilidad gaussiana (White gaussian noise: WGN). Se planean entonces dos experimentos para cada uno de los gases. Indistintamente, para octano o tolueno, se generan concentraciones de gas que varían como una señal WGN de media 6500 ppm y desviación estándar 2800 ppm. Si el experimento se realiza para octano, entonces la señal de concentración de tolueno se mantiene constante a 6500 ppm. Y viceversa, si el experimento se planea para tolueno, es la concentración de octano la que se mantiene constante a un valor de 6500 ppm. La siguiente figura presenta el experimento para gas octano.

En el experimento existe un periodo inicial de 120 minutos en el que se mantiene la concentración a un valor constante de 6500 ppm con el objeto de realizar una limpieza de la cámara, y establecer la respuesta del sistema al valor medio de concentración que se va a usar. A continuación se inicia el proceso de variación de concentración según las especificaciones dadas anteriormente. Este proceso dura alrededor de 900 minutos. Para acabar el experimento, se repite la situación inicial y se finaliza con un periodo de relajación instantánea a concentración nula.

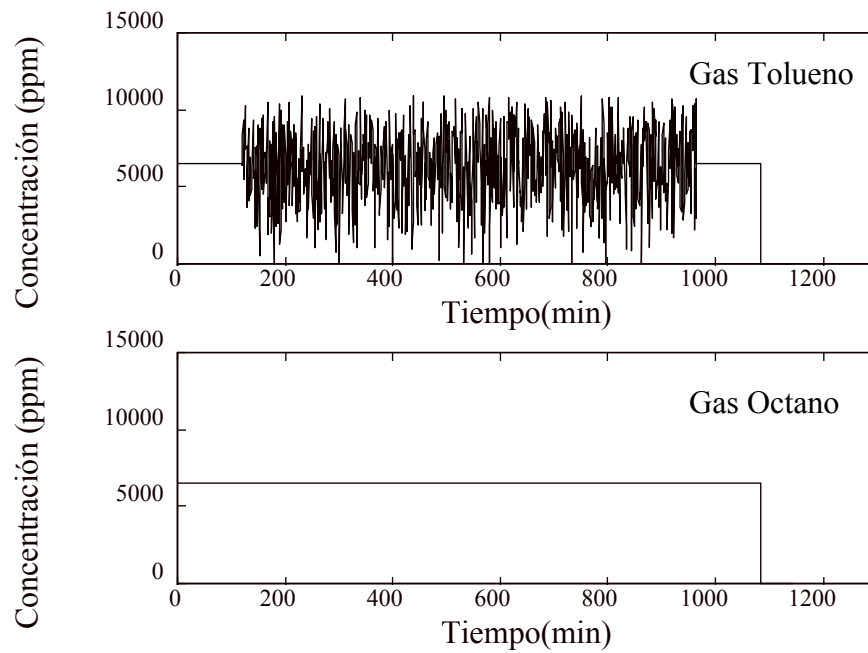


Figura 3.10 Experimento dos para gas tolueno

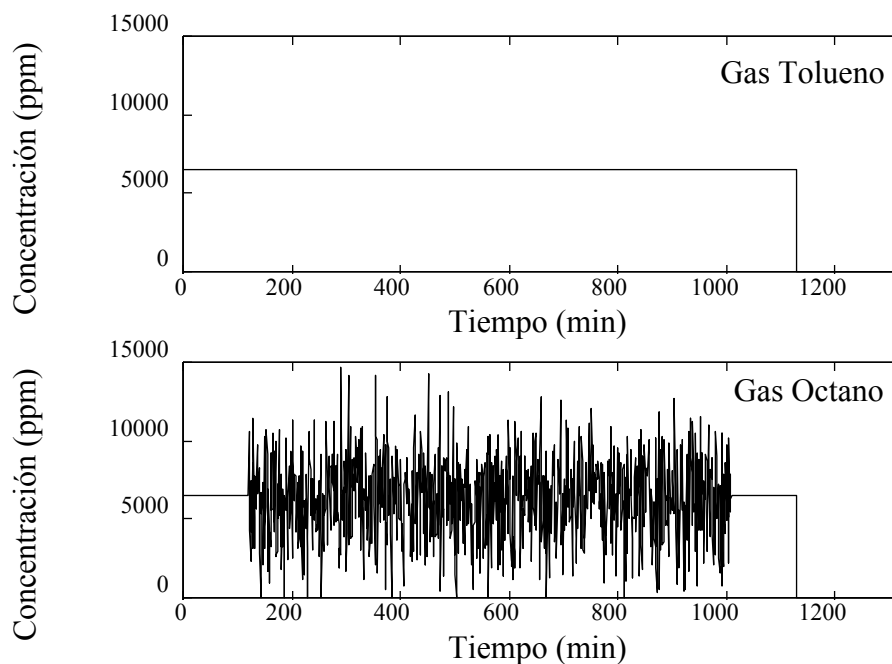


Figura 3.11 Experimento 2 para gas octano

Experimento 3:

Avanzando en complejidad, conviene generar experimentos donde se apliquen al sistema mezclas gaseosas en forma de señales de concentración WGN y no sólo experimentos de un solo gas. Este experimento somete al sistema de medida a mezclas binarias de los gases octano y tolueno variando aleatoriamente en forma de señales WGN de media 6500 ppm y desviación estándar 2800 ppm. La forma de la señal, representada en la siguiente figura, es similar a la descrita en el experimento 2.

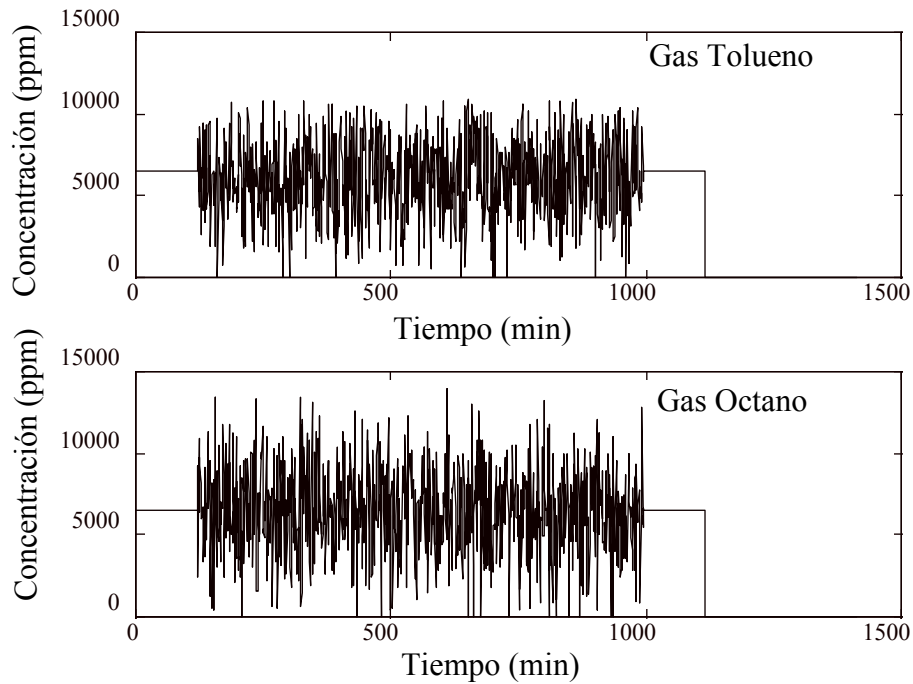


Figura 3.12 Concentraciones de tolueno y octano, experimento 3

Experimento 4:

Este experimento es similar al experimento 3. La diferencia se centra en los valores de concentración. El rango de las mismas está, en este caso, circunscrito las condiciones de valor medio 1000 ppm y desviación estándar 360 ppm. En la figura se presentan gráficas de los datos de concentración tanto para octano como para tolueno.

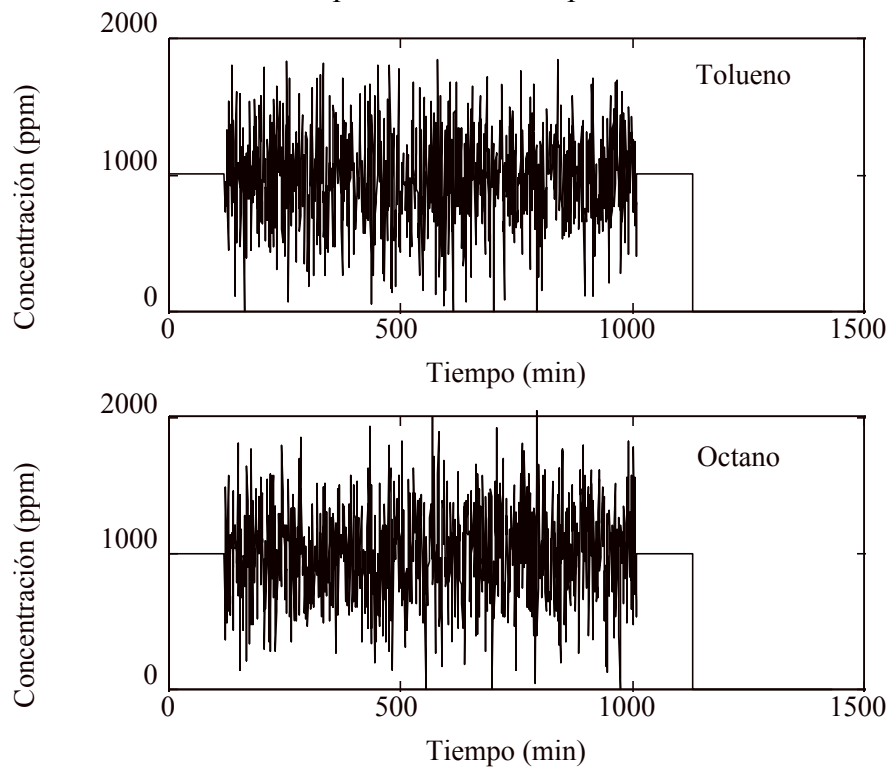


Figura 3.13 Concentraciones de tolueno y octano, experimento 4

Este experimento se va a emplear exclusivamente en el capítulo de modelización inversa, que se desarrollará más adelante. La presentación de los resultados del mismo, que para el resto de experimentos se presentan en el siguiente apartado, se pospone hasta el mencionado capítulo.

3.5. Presentación de resultados

Resultados experimento 1:

La siguiente figura presenta los datos de todos los sensores al experimento 1 para gas octano. De ella se puede observar el mal comportamiento del sensor 4 y el problema que aparece en la respuesta del sensor 5 con una serie de impulsos extraños de alto valor que no tienen significado físico.

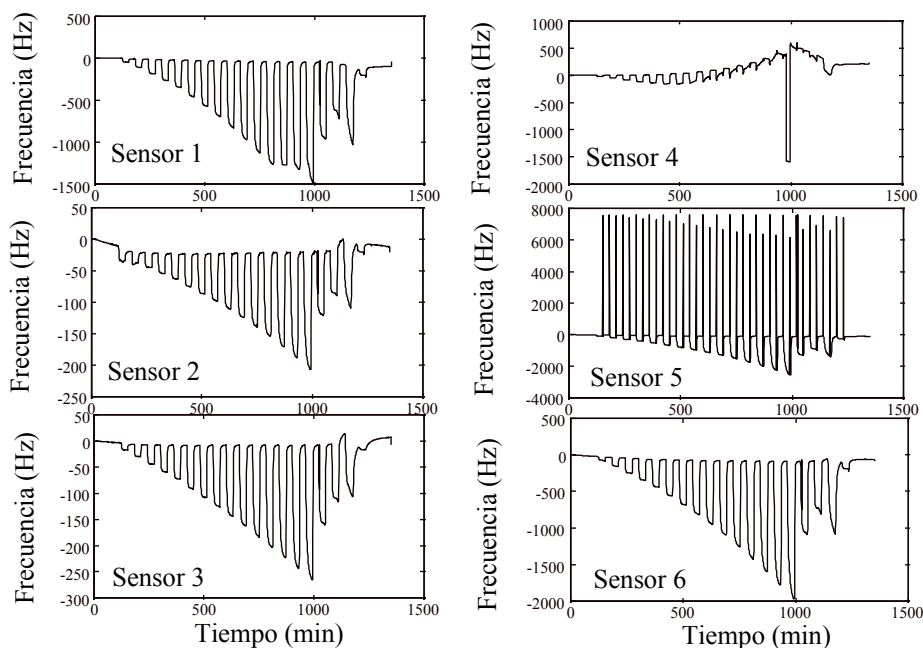


Figura 3. 14 Resultados los seis sensores ante el experimento 1

El sensor 4 es irrecuperable y no puede utilizarse en ningún caso para realizar un proceso identificativo. Sin embargo, sobre la respuesta del sensor 5 puede observarse que la anomalía es sistemática, produciéndose siempre sobre un determinado punto de la respuesta del sensor. Se ha realizado un filtro de dicha anomalía, eliminándose los puntos sin sentido físico tanto de la respuesta como de su correspondiente señal de entrada. La siguiente figura muestra la respuesta del sensor 5 restablecida.

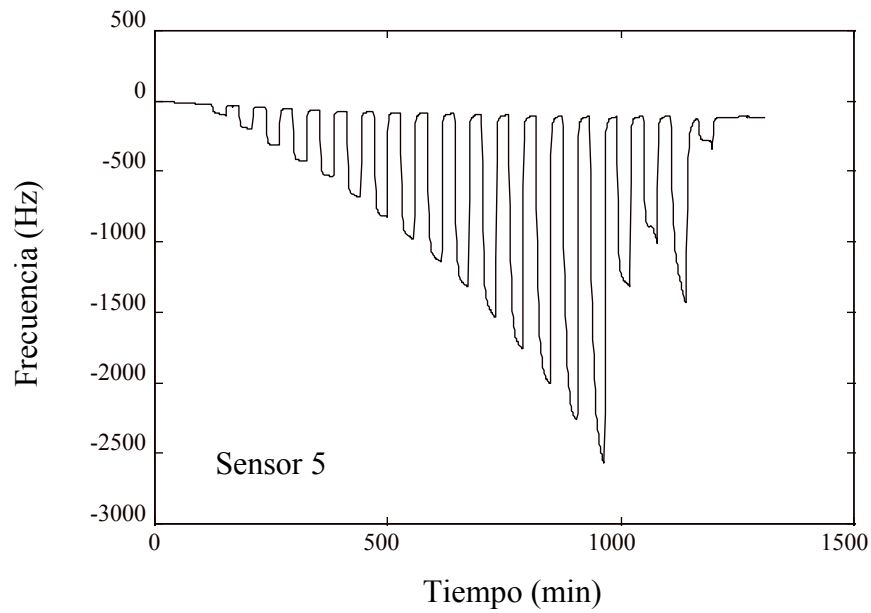


Figura 3.15 Respuesta del sensor 5 filtrada

Se observan también ciertas características como derivas de cero o histéresis sobre las cuales se hará mayor énfasis posteriormente.

Resultados experimento 2:

Para el experimento 2, los resultados son los siguientes:

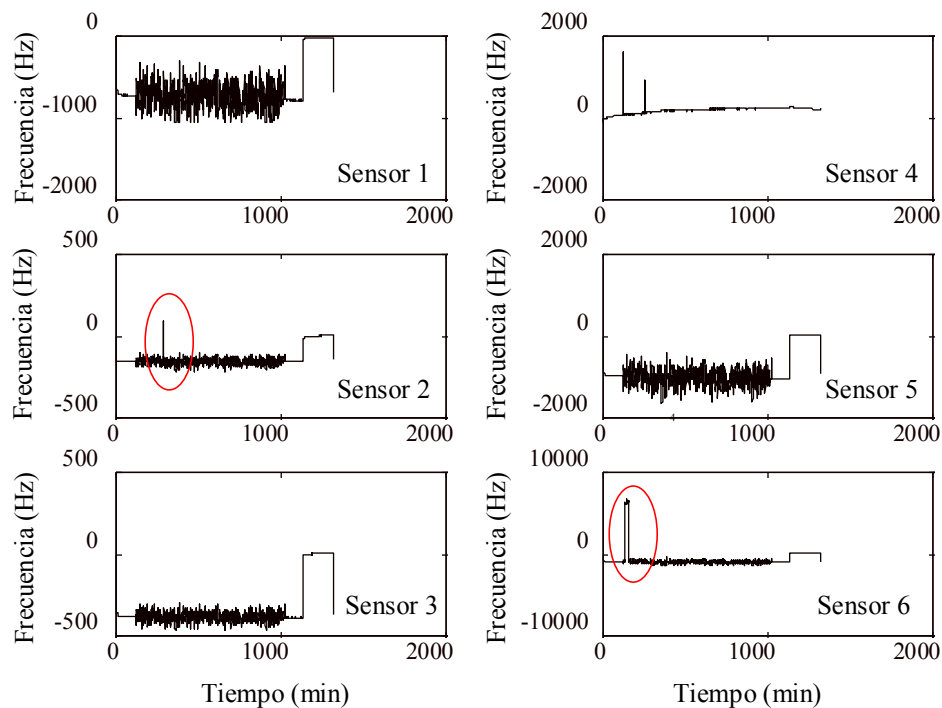


Figura 3.16 Resultados de los seis sensores ante el experimento 2

Sigue observándose a la luz de estos resultados que el sensor 4 no funciona bien. Por otro lado, el sensor 2 y el sensor 6 presenta ciertos comportamientos anómalos

(‘outliers’), marcados en rojo en la figura, en forma de picos de señal desviados del comportamiento general de la respuesta. Estos datos pueden resultar perjudiciales al usarlos en el proceso posterior de la identificación. Conviene por tanto, no utilizar ese conjunto de datos anómalos.

Resultados experimento 3:

Para el experimento 3, los resultados son:

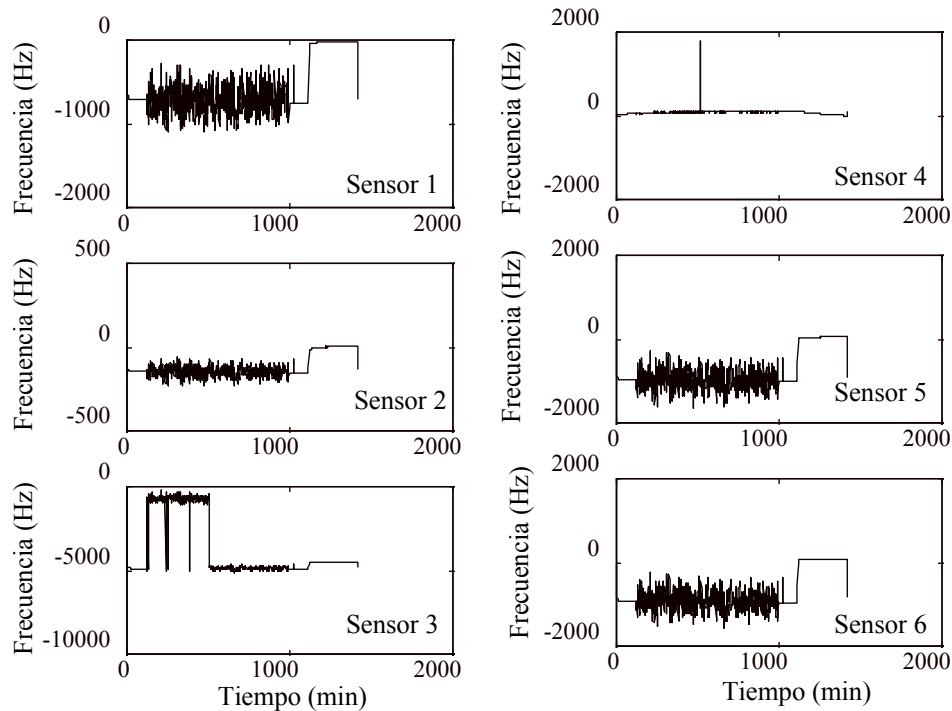


Figura 3. 17 Resultados de los seis sensores ante el experimento 3

De estos nuevos resultados sigue deduciéndose que el sensor 4 es inutilizable. Además el sensor 3 presenta una respuesta que no ofrece garantías de éxito para el proceso identificativo.

Como primer resumen del proceso experimental llevado a cabo, se debe concluir que el sensor 4 no debe utilizarse en ningún proceso identificativo y que, una vez corregidas o despreciadas las posibles anomalías, los sensores que básicamente se utilizarán en el estudio de identificación que sigue a continuación serán los sensores 1 2 5 y 6, despreciando el sensor 3 por su mal comportamiento en el experimento 3. Por otro lado, tal como ya se comentó anteriormente y a fin de observar el rendimiento de los diferentes métodos identificativos en relación con la cantidad de datos, e intentando que el número de datos a utilizar sea pequeño, se van a usar solamente 300 puntos de los experimentos 2 y 3 para realizar todos y cada uno de los procesos de identificación.

3.6. Caracterización estática de los sensores

Antes de entrar en el problema de identificación dinámica conviene conocer las características de la respuesta de los sensores, saber si responden linealmente o no, si derivan con el tiempo, si presentan histéresis, etc.

Estas características estáticas pueden conocerse a través del experimento 1. Si el sensor tiene tiempo de llegar al estado estacionario, del análisis de la respuesta pueden extraerse algunas características interesantes que conviene conocer previamente al estudio dinámico (por ejemplo las no linealidades del sistema o las derivas en el tiempo).

Las siguientes gráficas y tablas muestran las características estáticas y dinámicas del sistema sensor.

La siguiente figura presenta la respuesta del sensor 2 sometido al experimento 1 realizado para gas octano. En ella se representan solo las respuestas del sensor al conjunto de escalones de amplitud creciente.

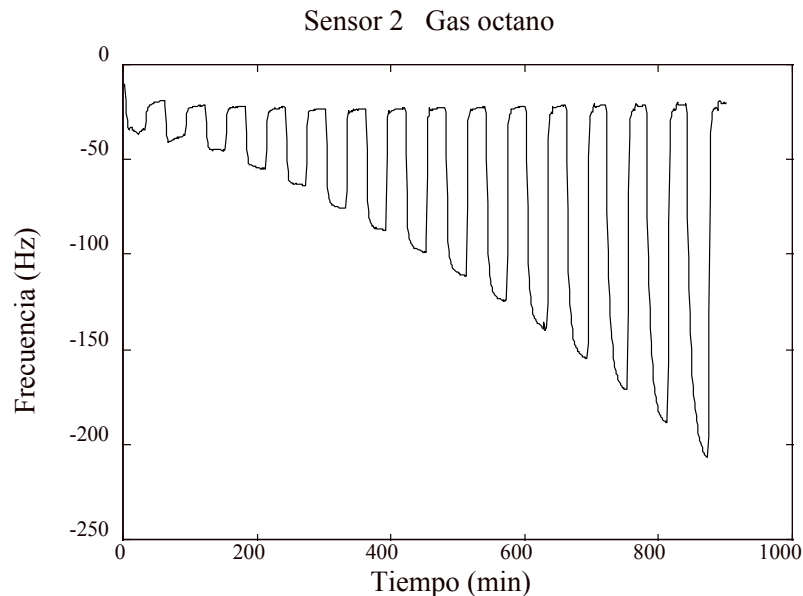


Figura 3. 18 Respuesta estática del sensor 2 para gas octano.

En esta gráfica se pueden observar varios detalles interesantes. En primer lugar, la respuesta estática del sensor no es lineal. Se puede observar, que la condición de establecimiento de una respuesta estacionaria no se cumple en todos los escalones, acrecentándose el problema para concentraciones altas. En cualquier caso, aun y trabajando con estos datos, la medida de no linealidad no queda invalidada, sino que constituye una cota mínima de la misma. Si se dejara al sistema llegar al estado estacionario, el valor de no linealidad sin duda sería superior. Se aprecia, además, que los tiempos que el sistema tarda en alcanzar el estado estacionario son diferentes y que los tiempos de subida y bajada son distintos y esta diferencia se acrecienta para concentraciones altas. En la siguiente figura, extraída a partir de la anterior, se aprecia con mayor claridad estos detalles.

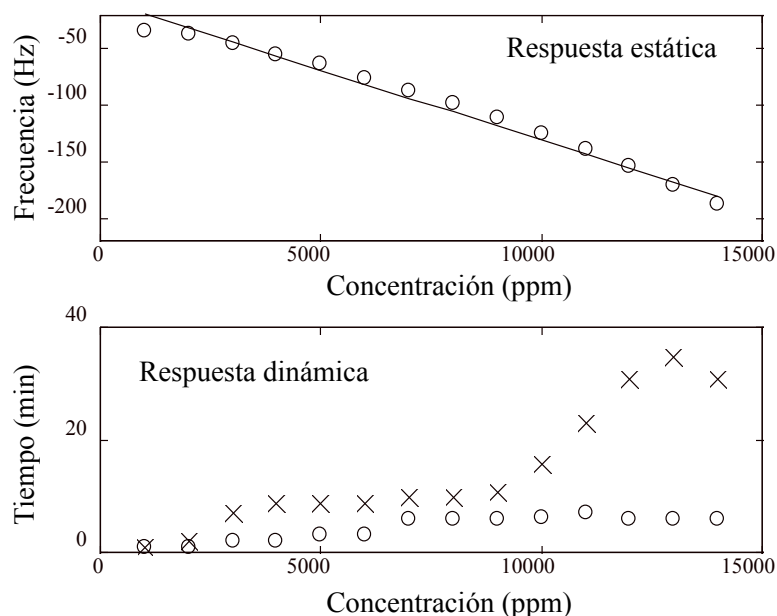


Figura 3.19 Sensor 2 gas octano, a) Característica estática del sensor 2 para gas octano
b) Tiempos de respuesta en función de la concentración para gas octano.

En la primera gráfica se dibuja la respuesta estática comparada con un ajuste lineal en función de la concentración. La segunda gráfica muestra los tiempos de excitación y relajación de la respuesta del sensor 2 para el experimento 1 con gas octano. Estos tiempos se definen como el tiempo que transcurre entre la transición de la señal desde un 10% a un 90 % del valor final para la excitación y al revés para la relajación. Se observa que para ciertos rangos de concentración, los tiempos se mantienen constantes, pero en el conjunto del experimento, los tiempos de excitación varían notablemente.

Las siguientes tablas presentan un resumen de las características estáticas de los sensores en el sistema de medida, extraídas a partir del comportamiento observado a través del experimento 1. La linealidad está calculada en referencia al mejor ajuste lineal independiente, en porcentaje respecto a la salida a fondo de escala, mientras que la histéresis se presenta como un tanto por ciento del valor de fondo de escala.

Tabla 3.3: Características estáticas para el gas octano				
Gas Octano	Sensor 1	Sensor 2	Sensor 5	Sensor 6
No linealidad mínima	7.1%	7.2%	9.8%	6.5%
Sensibilidad (Hz/ppm)	$-49.6 \cdot 10^{-3}$	$-5.6 \cdot 10^{-3}$	$-56 \cdot 10^{-3}$	$-67 \cdot 10^{-3}$
Histéresis (%fondo de escala)	10.6	6.1	7.5	5.8
Deriva de cero (Hz/1000 min)	47.7	13.6	93.4	67.1

Tabla 3.4: Características estáticas para el gas tolueno				
Gas Tolueno	Sensor 1	Sensor 2	Sensor 5	Sensor 6
No linealidad mínima	1.5%	1.2%	1.3%	2.5%
Sensibilidad (Hz/ppm)	$-2.76 \cdot 10^{-3}$	$-7.23 \cdot 10^{-3}$	$-5.12 \cdot 10^{-3}$	$-6.07 \cdot 10^{-3}$
Histéresis (% fondo de escala)	4.8	8.0	5.0	4.2
Deriva de cero (Hz/1000 min)	34.1	36.4	43.7	36.9

Aunque el tema de las derivas no va a ser tratado en este trabajo, es un tema interesante en el que diversos grupos de investigación trabajan en la actualidad^{18 19 20}.

Del conjunto de todos estos resultados, se extraen algunas conclusiones claras:

- a) Todos los sensores presentan un cierto grado de no linealidad
- b) El sistema presenta una dinámica variable dependiendo de la concentración
- c) La sensibilidad de los sensores es en general (excepto para el sensor 2) mayor para el octano, siendo la diferencia de un orden de magnitud. Este hecho, debería dificultar las medidas de tolueno en presencia de octano.
- d) La no linealidad de los sensores, sin excepción, también es mayor para el octano.
- e) Los sensores responden a ambos gases
- f) Existe histéresis en el sistema.
- g) Los sensores derivan

Conocidas las características estáticas del sistema de medida y como paso previo al proceso de identificación dinámica, conviene demostrar que un ajuste estático no proporciona resultados lo suficientemente precisos como para no justificar el salto hacia la modelización dinámica. Las características estáticas antes expuestas se han extraído a partir de los resultados del experimento 1. El ajuste estático que se va a realizar a continuación se basa en el cálculo mediante mínimos cuadrados de la relación entre la entrada y la salida de un sistema SISO formado por un sensor y un gas sometido al experimento 2. Según este modelo estático, el sistema se rige por la siguiente ecuación:

$$y_{\text{est}}(n) = b \cdot x(n-k) \quad (3.1)$$

donde el coeficiente b se ajusta mediante un proceso basado en cálculo de mínimos cuadrados, tal y como se ha dicho. El factor de retraso k se obtiene a partir de la mejor estimación de $y_{\text{est}}(n)$ respecto al valor real $y(n)$, habiéndose generado estimaciones para valores de k desde 1 hasta 20 ($k=1 \dots 20$), $x(n)$ es una señal, como ya se ha dicho, de ruido blanco gaussiano, de media cero y desviación estándar uno. Para el caso concreto de la matriz de sensores presentada aplicada sobre los gases octano y tolueno, los retrasos son: $k=1$ para el gas tolueno y $k=2$ para el gas octano.

La siguiente figura muestra el ajuste estático realizado sobre el sensor 1 y el gas octano mientras que la tabla presenta la relación en tanto por ciento de la parte de la señal no estimada respecto de la señal de salida para los sensores 1 2 5 y 6 sometidos a concentraciones de octano o tolueno. Esta relación se mide a través de una función de mérito denominada Potencia de señal no estimada (PSNE) definida según la siguiente ecuación:

$$\text{PSNE}(\%) = \frac{\sum_n (y(n) - \hat{y}(n))^2}{\sum_n (y(n))^2} \times 100 = \frac{\sum_n (v(n))^2}{\sum_n (y(n))^2} \times 100 \quad (3.2)$$

El error que se introduce a través de las estimaciones en la figura PSNE se puede evaluar fácilmente. Dicho error se define según la siguiente ecuación:

$$\Delta \text{PSNE}(\%) = \frac{\sqrt{2}}{2} \frac{\text{PSNE}(\%)}{\sqrt{N}} \quad (3.3)$$

Se distingue en la tabla entre estimación y validación, refiriéndose la columna de estimación a los resultados obtenidos con los mismos datos usados en la estimación del modelo y la columna de validación a los resultados obtenidos con datos del experimento no usados en la estimación del modelo.

Tabla 3. 5: Valor en tanto por ciento de la PSNE para una estimación estática

Sensores	Octano (k=2)		Tolueno (k=1)	
	Estimación	Validación	Estimación	Validación
Sensor 1	27.3 %±1.1%	42.0 %±3.3%	53.3 %±2.2%	49.8 %±3.9%
Sensor 2	42.8%±1.7%	31.5%±2.5%	13.7 %±0.5%	15.1 %±1.2%
Sensor 5	27.8 %±1.1	23.9 %±1.9%	13.8 %±0.5%	14.6 %±1,1%
Sensor 6	23.6 %±1.0	23.6 %±1.9%	7.4 %±0.3%	8.1 %±0.6%

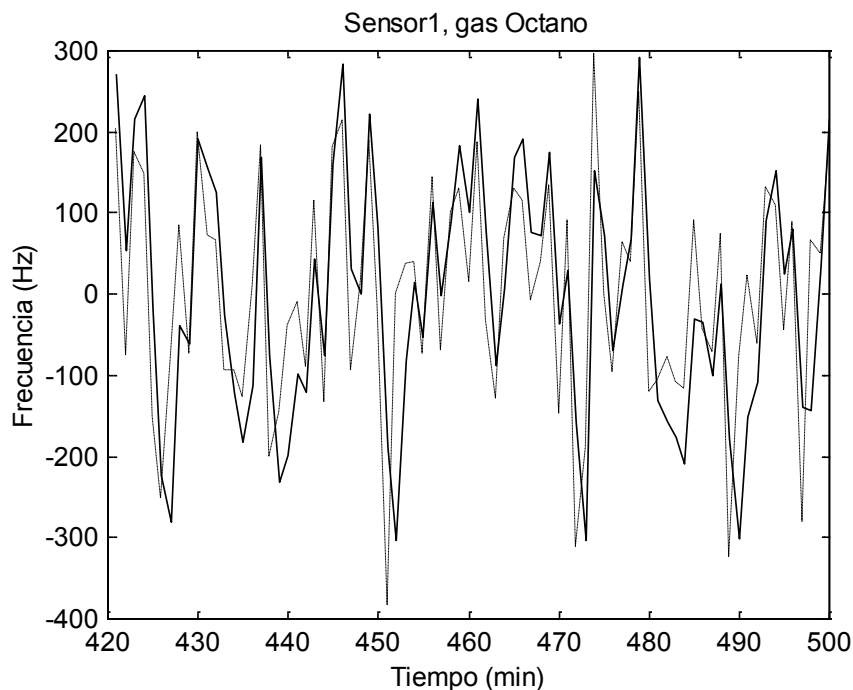


Figura 3. 20 Comparación entre la señal real y la estimación estática

En la figura se presenta en línea continua los datos reales del experimento, mientras que en línea punteada se representan los resultados de la modelización estática.

Aparte de estos resultados obtenidos, cabe preguntarse por el rendimiento de esta estimación estática frente al experimento 1. Las siguientes figuras muestran la comparación del resultado real con el ajuste estático para el sensor 1 y el gas octano o gas tolueno. Se observa que, por supuesto, el modelo estático aplicado al experimento 1 no es capaz de aproximar ni la dinámica del sistema ni su no linealidad. Este resultado ya era de esperar, aunque sí parece que, salvo diferencias debidas al uso de diferentes datos para realizar los cálculos, el modelo tiende a ajustar bien la sensibilidad a bajas concentraciones, siendo esto cierto tanto para el modelo estático con gas octano como con gas tolueno.

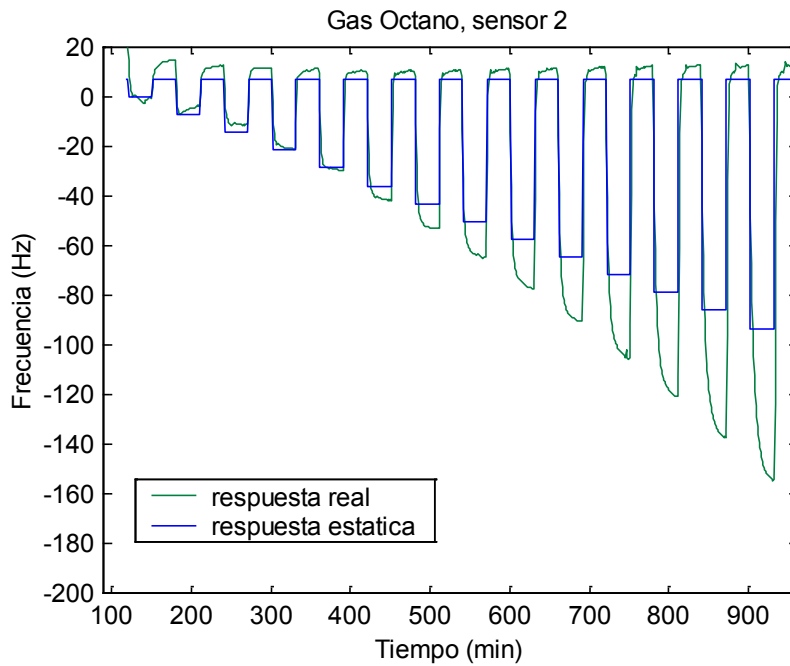


Figura 3. 21 Aproximación estática al experimento 1: Gas octano

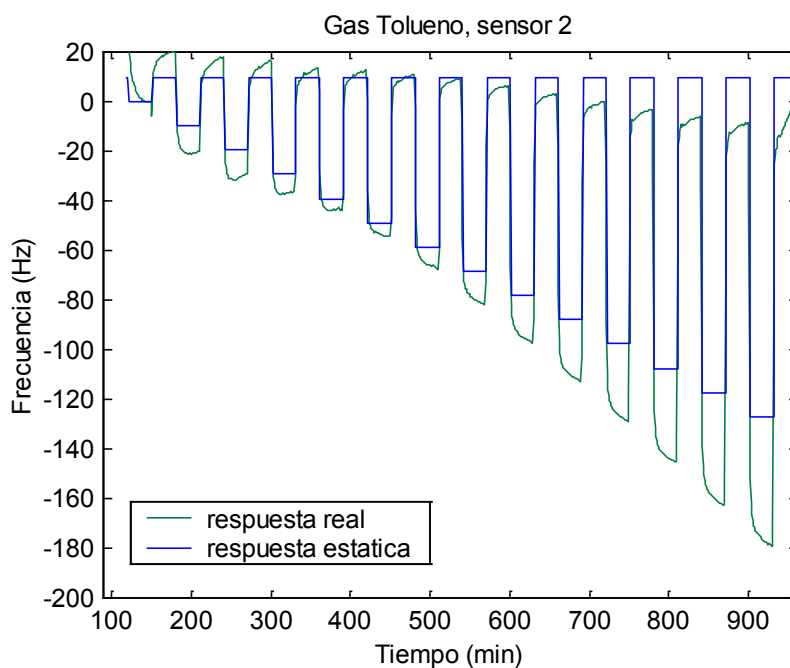


Figura 3. 22 Aproximación estática al experimento 1: Gas tolueno

3.7. Resumen y conclusiones del capítulo

En este capítulo se han introducido las nociones elementales y las bases de las estrategias que se van a seguir a lo largo de todos los procesos identificativos que se presentarán en los sucesivos capítulos. Siempre, por tanto, se hará referencia a este capítulo para interpretar el proceso que se lleve a cabo. Se han presentado, por otra parte, los datos que se utilizarán en los procesos y se han seleccionado un conjunto de sensores por su adecuado funcionamiento frente a los experimentos realizados sobre él. Además, analizados los experimentos y sus resultados, se ha demostrado que el sistema presenta un cierto grado de no linealidad, más acusado para el gas octano que para el

tolueno. En cuanto a la aplicación de la estimación estática, se ha demostrado que proporciona unos resultados cualitativamente aceptables, en el sentido que el modelo generado para el experimento 2 parece seguir las tendencias de la señal real tal y como muestra la figura 3.20, pero por el contrario y como era de esperar, el uso de este modelo para ajustar los datos del experimento 1 ofrece un rendimiento bajo. En cuanto al estudio de la estimación estática a nivel cuantitativo y a partir de los resultados expuestos en la tabla 3.5, es necesario abrir una reflexión más profunda; reflexión que hay que encaminar por varias vertientes. Por un lado, la aproximación estática para el gas octano presenta un rendimiento pobre por dos motivos fundamentales, por un lado existen dependencias dinámicas no contempladas en el modelo estático (como se demostrará en el capítulo siguiente) y por otro lado, el sistema presenta un grado de no linealidad que también afecta al rendimiento de la estimación. Es precisamente este factor el responsable de que el rendimiento de la estimación estática en el caso del tolueno sea mejor (a excepción del caso del sensor 1), ya que el grado de no linealidad para este gas es menor. El hecho de que el rendimiento del sensor 1 para gas tolueno sea tan malo es achacable exclusivamente al otro factor en juego, es decir a la dinámica no contemplada. Efectivamente, tal como muestra la figura 3.23 además del factor de retraso $k=1$, existe un coeficiente adicional para el retraso $k=2$ que es prácticamente del mismo valor que el usado en el proceso de estimación estática.

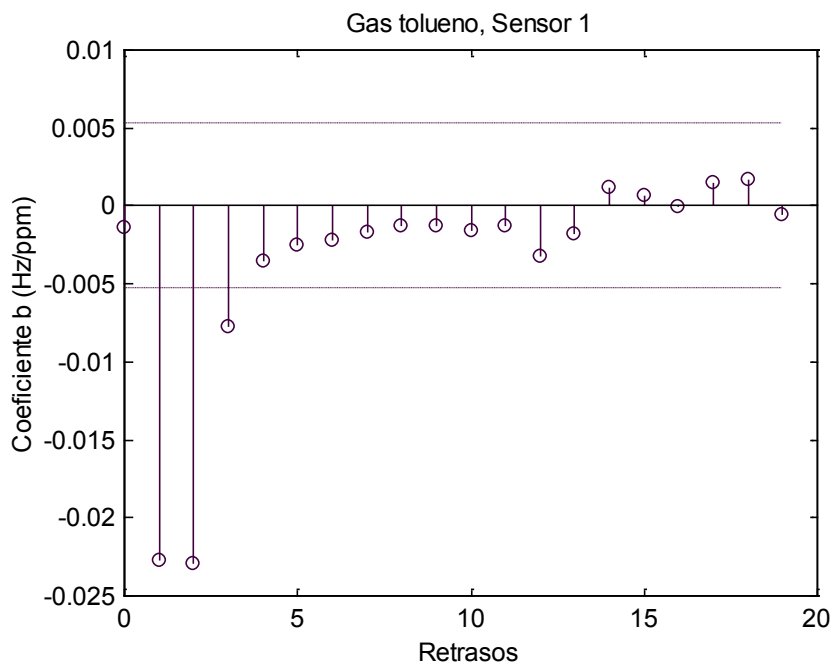


Figura 3. 23 Coeficientes de aproximación estática en función de los retrasos k

Este hecho no ocurre en ninguno de los otros casos contemplados, en los cuales, si bien sí que existen otros términos que contienen parte de la dinámica del sistema, en ningún caso llegan a ser son de igual o superior valor al contemplado en esta estimación estática. En cualquier caso, queda claro que los resultados son abiertamente mejorables y lo son, además en dos claras vías de actuación:

- a) la estimación de la dinámica del sistema
- b) la no linealidad

La primera vía es la que se desarrolla en el capítulo 4 mientras que la segunda se ha reservado exclusivamente al capítulo 5.

Bibliografía

- ¹Jerry M. Mendel, *Lessons in estimation theory for signal processing, communications and control* (Prentice Hall signal processing series, New Jersey, 1995).
- ²T. Glad L. Ljung, *Modeling of Dynamic Systems* (Prentice Hall, New Jersey, 1994).
- ³L. Ljung, *System identification. Theory for the user* (Prentice Hall, New Jersey, 1987).
- ⁴S. Marco, A. Pardo, F. Davide, C. DiNatale, A. D'Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, U. Weimar, W. Göpel, "Different strategies for identification of gas sensing systems," *Sensors and Actuators B* **34**, 213-223 (1996).
- ⁵F. Davide, C. DiNatale, A. D'Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, W. Göpel, S. Marco, A. Pardo. "Dynamic calibration of QMB polymer-coated sensors by Wiener kernel estimation," *Sensors and Actuators B* **26-27**, 275-285 (1995).
- ⁶Gwilym M. Jenkins George E.P. Box, *Time series analysis forecasting and control* (Holden-Day, Oakland, 1976).
- ⁷P.Z. Marmarelis, Ken-Ichi Naka, "Identification of Multi-Input Biological Systems" *IEEE Transactions on biomedical engineering* (1974)
- ⁸X. Vilanova, E. Llobet, J.R. López, X. Correig, "Nuevo método de identificación de gases basado en la respuesta transitoria de sensores de óxido de estaño" Seminario anual de Automática y Electrónica Industrial (1995)
- ⁹S. Wlodek, K. Clobow, F. Con sadori, "Signal Shape analysis of a thermally cycled tin oxide gas sensor" *Sensors and Actuators B* **3**, 63-68, (1991)
- ¹⁰K. Fukui "Detection and measurement of odor by sintered tin oxide gas sensors" *Sensors and Actuators B*, **5**, 27-32 (1991)
- ¹¹Jhon M. Parr Charles L. Phillips, *Signals, Systems and Transforms* (Prentice Hall, New Jersey, 1995).
- ¹²Ronald W. Schafer Alan V. Oppenheim, *Digital Signal Processing* (Prentice Hall, New Jersey, 1975).
- ¹³T. Nakamoto, N Okazaki, T. Mornzumi "High speed active gas odor sensing system using adaptig control theory", *Sensors and Actuators B*, **41**,183-188, (1997)
- ¹⁴M. Nakamura, I. Sujimoto, H. Kuwano, "Aplication of plasma polymer film coated sensors to gas identification using linear filters" *Sensors and Actuators B*, **33**,122-127, (1996)
- ¹⁵D. Wilson, S. DeWeerth, "signal processing for improving gas sensor response time" *Sensors and Actuators B*, **41**,63-70, (1997)
- ¹⁶J.W. Gardner, E. Llobet, E.Hines, "Pattern recognition techniques in chemical sensing: a review of dinamic modelling", Proc. 9th int. conf. on modern materials and technologies. Florence, Italy (1998)
- ¹⁷J.W. Gardner, "Recent advances in electronic noses: miniaturization and signal processing", *Actas de la conf. De dispositivos electrónicos*. 5-9.Madrid (1999)
- ¹⁸C. Di Natale, F. Davide, A. D'Amico, "A self organizing suystem for pattern classification: time variing statistics and sensor drift efects", *Sensors and Actuators B*, **26-27**,237-241, (1995)
- ¹⁹M.Holmberg, F. Winquist, I. Lundström, F. Davide, C. Di Natale, A.D'Amico, "Drift counteraction for an electronic nose", *Sensors and Actuators B*, **35-36**,528-535, (1996)
- ²⁰M. Roth, R. Hartinger, M. Schwaiger, M. Graber, H. E. Endress, "Drift reduction of gas sensors by Temepreature modulation and signal processing", 3th IFAC Symp. on Intelligent Components and Instruments for Control Aplicacions, 405-408, Annecy. France (1997)

Capítulo 4

Métodos de identificación dinámica: Aproximación lineal

- *Claro, claro, hemos de buscar otra cosa. Vamos a utilizar pues una variante estratégica basada en las aproximaciones sucesivas. Ve, señor, y telegráfale al dragón que le cederás el trono a condición que ejecute tres operaciones matemáticas muy sencillas...*

Stanislaw Lem: Fábulas de robots

- *Basta aquí con decir que cuando deseamos que un movimiento siga una pauta determinada, se utiliza la diferencia entre esta pauta y el movimiento realmente efectuado como nueva señal de entrada para hacer que la parte regulada se mueva de tal forma que su movimiento siga fielmente el efectuado por el modelo.*

Norbert Wiener: Cibernética

Prólogo del capítulo

Acabada la presentación de los experimentos y realizada la caracterización estática, y demostrada además la baja eficiencia de los métodos estáticos, hay que avanzar hacia la caracterización dinámica. En una primera aproximación, según se plasma en el esquema de la figura 4.1, se inicia la caracterización dinámica desde el punto de vista lineal. El desarrollo de esta aproximación lineal se va a realizar según el siguiente esquema:

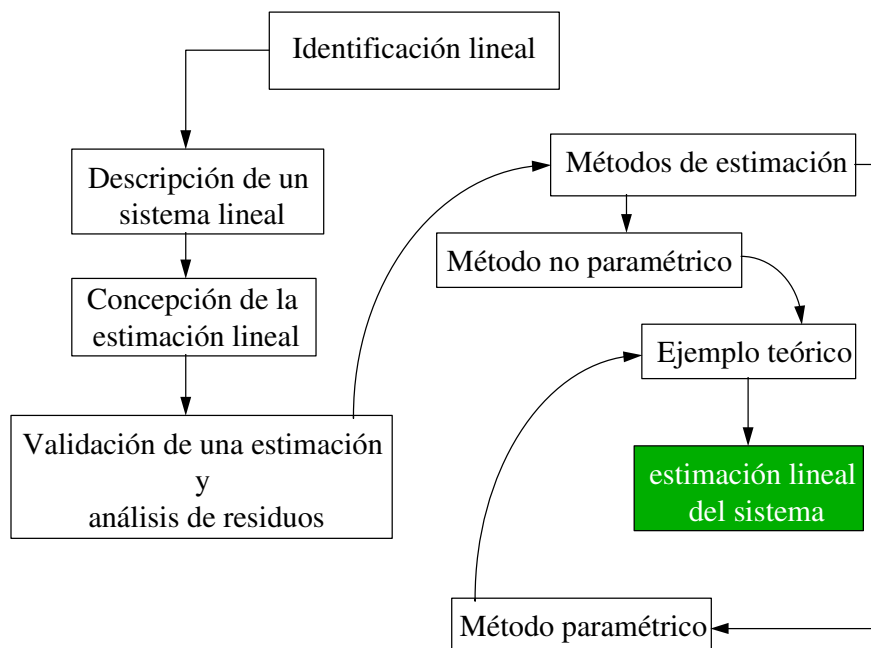


Figura 4. 1 Esquema del desarrollo del capítulo

4.1. Introducción a la descripción matemática de un sistema Lineal e Invariante en el tiempo (LTI)

Como es bien sabido, funciones y conceptos como respuesta impulsiva, ecuación de diferencias, función de transferencia o respuesta en frecuencia son diferentes desarrollos que reflejan la descripción matemática de una misma cosa: un sistema.

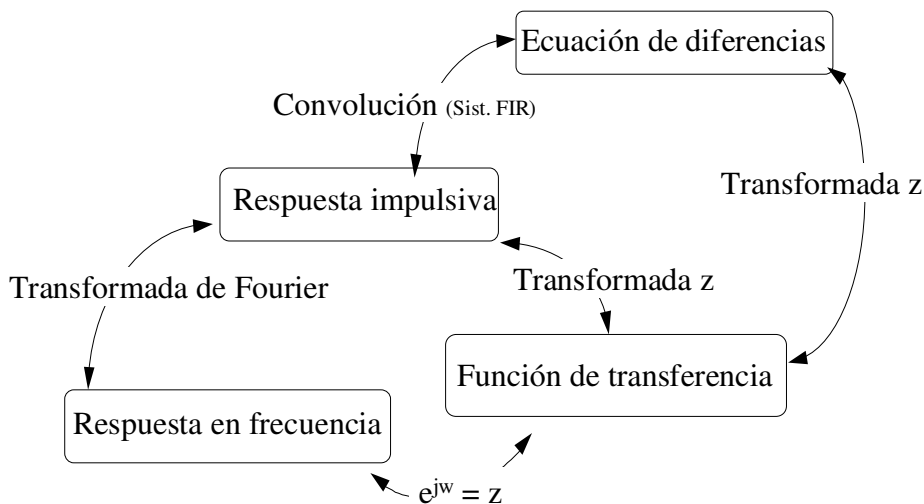


Figura 4. 2 Relaciones entre las distintas descripciones de un sistema LTI

Aunque en la literatura específica referida a identificación de sistemas, procesamiento de señal o control de sistemas pueden hallarse descripciones extensas y rigurosas de estos conceptos conviene introducir brevemente una definición de los mismos^{3 4 5}.

En el presente trabajo, tal y como ya se ha comentado varias veces, la concepción del sistema hace que las señales con las que se trabaja sean discretas, y bajo esta concepción hay que desarrollar las descripciones. Sea, por tanto, un sistema SISO discreto lineal e invariante en el tiempo (LTI) y sean $x(n)$ e $y(n)$ las señales de entrada y salida respectivamente del sistema. Supóngase además que el sistema es causal; entonces, debido al carácter lineal del sistema, la salida $y(n)$ viene descrita por una ecuación de diferencias en forma de combinación lineal de las entrada $x(n)$.

$$y(n) = \sum_{j=0}^{\infty} g(j)x(n-j) \quad (4.1)$$

El conjunto de valores $g(j)$ con $j=0,1,2,\dots,\infty$ que se combina con la entrada $u(n)$ forma la respuesta impulsiva del sistema. El conocimiento completo de la respuesta impulsiva $g(j)$ caracteriza completamente a los sistemas LTI.

Para un sistema MIMO, la descripción del sistema viene dada según la expresión:

$$y_i(n) = \sum_{k=0}^{\infty} g_{ij}(k)x_j(n-k) \quad (4.2)$$

con $i=1,\dots,n$ y $j=1,\dots,m$.

En definitiva, la respuesta impulsiva representa la relación de las variables del sistema en el dominio temporal, y define al sistema completamente. Pero, como es conocido^{1 6} y según muestra la figura 4.2 la representación de la propia respuesta impulsiva en espacios de variables diferentes genera nuevas funciones que, obviamente, también describen las relaciones entre las variables del sistema completamente, pero presentan la información desde otro punto de vista. La función de transferencia $G(z)$ es la representación de la respuesta impulsiva en el espacio de la variable z y se define según la siguiente expresión:

$$G(z) = \sum_{j=0}^{\infty} g(j)z^{-j} \quad (4.3)$$

Esta función describe el comportamiento de un sistema lineal en el espacio de la transformada z según la ecuación:

$$Y(z)=G(z)X(z) \quad (4.4)$$

donde la salida del sistema queda completamente determinada por el producto algebraico de la función de transferencia y las señales de entrada en el espacio z .

La función $G(z)$ suele representarse para sistemas LTI, sin perder generalidad, en forma de cociente de polinomios de la forma:

$$G(z) = z^{-n_k} \frac{b_1 z^{-1} + b_2 z^{-2} + \dots + b_{n_b} z^{-n_b}}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_{n_a} z^{-n_a}} \quad (4.5)$$

El valor de n_k está relacionado con el retraso temporal puro entre la entrada de la señal y la respuesta del sistema. Por otro lado, n_b y n_a determinan el número de ceros y polos de

la función racional $G(z)$. A través de ellos es posible establecer ciertas características del sistema como por ejemplo la estabilidad del mismo. De alguna manera, esta forma racional de describir los sistemas entronca con las ecuaciones de diferencias, y es posible, a través de las propiedades de la transformada z , reescribir la descripción del sistema en el espacio temporal según la ecuación 4.1 substituyendo un sumatorio de carácter infinito por dos sumatorios de carácter finito:

$$\sum_{i=0}^{n_a} a_i y(n-i) = \sum_{j=1}^{n_b} b_j x(n-n_k-j) \quad (4.6)$$

Por otra parte, la respuesta en frecuencia $G(\omega)$ de un sistema representa la relación entre las variables del sistema en el dominio de frecuencias y puede obtenerse a partir de la transformada de Fourier de la respuesta impulsiva. Por supuesto, al igual que anteriormente, la descripción de un sistema LTI en base a la respuesta en frecuencia describe el comportamiento de dicho sistema completamente desde el espacio de las frecuencias. Gracias a la equivalencia de las tres representaciones, la respuesta en frecuencia también se puede obtener a partir de la función de transferencia substituyendo z por $\exp(-j\omega)$ siempre que la región de convergencia de la transformada z incluya el círculo unidad^{13 14}.

4.2. Estimación de un sistema LTI

Introducidas las descripciones básicas de un sistema LTI, hay que continuar con la estimación de las mismas a partir de los datos experimentales.

Sea entonces un sistema SISO, LTI y causal con respuesta impulsiva $g(k)$. Mediante el experimento adecuado, se dispone del conocimiento experimental de una cierta señal de entrada $x(n)$ y su correspondiente señal de salida $y(n)$. El hecho de trabajar con señales experimentales hace ineludible la presencia de cierta señal de ruido $v(n)$ que, por sus características aleatorias, es difícil de estimar y de tal manera que $y(n)$ puede representarse según la siguiente ecuación:

$$y(n) = \sum_{k=0}^{+\infty} g(k)x(n-k) + v(n) \quad (4.7)$$

La identificación del sistema consiste en hallar una respuesta impulsiva $\hat{g}(n)$, de tal manera que:

$$\hat{y}(n) = \sum_{k=0}^{+\infty} \hat{g}(k)x(n-k) \quad (4.8)$$

donde $\hat{g}(k)$ representa una estimación de la respuesta impulsiva e $\hat{y}(n)$ es la respuesta del modelo generado a partir de la estimación de la respuesta impulsiva y la señal de entrada. Debe remarcarse la diferencia entre la respuesta impulsiva real del sistema $g(k)$ y la estimación $\hat{g}(k)$, así como entre la salida real $y(n)$ y la salida estimada a través de $\hat{g}(k)$: $\hat{y}(n)$.

La diferencia entre la salida real $y(n)$ y la estimación $\hat{y}(n)$ es el error de la estimación o residuo.

$$\varepsilon(n) = y(n) - \hat{y}(n) \quad (4.9)$$

En definitiva, las señales de un sistema estimado a través de un modelo lineal están relacionadas según la ecuación:

$$y(n) = \sum_{l=-\infty}^{+\infty} \hat{g}(l)x(n-l) + \varepsilon(n) \quad (4.10)$$

donde, si la estimación fuese perfecta, se cumpliría que $\hat{g}(n) = g(n)$ y $\varepsilon(n) = v(n)$ para todo n .

La cuestión fundamental radica en hallar un método eficiente para calcular $\hat{g}(n)$ minimizando las diferencias con $g(n)$. A esta cuestión no es ajeno el residuo $\varepsilon(n)$, en el sentido que los diferentes métodos de estimación de modelos, como se verá mas adelante, exigen que el residuo tenga características blancas. Sin embargo, la señal de ruido $v(n)$ no tiene porque ser estrictamente blanca. Esto puede suponer una dificultad puesto que ese caso no sería posible realizar la equiparación directa entre el residuo blanco y el ruido no blanco.

Algunos métodos de estimación (los paramétricos, por ejemplo)^{1 2} incluyen una cierta modelización de la señal de ruido $v(n)$ de tal manera que $v(n)$ y una cierta perturbación blanca $e(n)$ están relacionados a través de una cierta función de transferencia. En esos casos, el objetivo de la identificación es, a la vez, la función de transferencia entre la entrada y la salida del sistema y la función de transferencia entre la perturbación y el ruido. Por el contrario, otros métodos^{15 16} (los no paramétricos, por ejemplo) no realizan ninguna reflexión sobre el ruido $v(n)$. En esos casos, el residuo $\varepsilon(n)$ y el ruido $v(n)$ sólo pueden ser equivalentes si el propio ruido ya es blanco de origen.

4.3. Validación y análisis de residuos

La comprobación de si la estimación $\hat{g}(n)$ es aceptablemente buena es tarea de la validación y del análisis de residuos. Obviamente, si se conociesen los valores de $g(n)$, esta tarea no sería necesaria, y bastaría con una simple comparación de los valores entre $g(n)$ y $\hat{g}(n)$. Esto no puede ser así ya que las señales que son accesibles experimentalmente son $x(n)$ e $y(n)$, y es a partir de estas que se debe evaluar la adecuación de $\hat{g}(n)$.

El primer paso obvio en la comprobación de la bondad de una estimación consiste en la simple comparación entre los resultados aportados por el modelo identificado y los resultados reales. Este proceso recibe el nombre de validación y a través de él puede calcularse el error de la estimación como la diferencia entre ambos resultados. Una buena manera de cuantificar la bondad de la estimación es precisamente a través del error obtenido por la validación y consiste en calcular el tanto por ciento que representa la potencia del error respecto de la potencia de la señal real de salida. De esa manera se obtiene el tanto por ciento de señal no estimada por el modelo.

A partir de todo esto, se ha generado una figura de mérito PSNE (Potencia de señal no estimada), que ya se utilizó en el capítulo anterior, y que evalúa la potencia de la señal no estimada por el modelo según la siguiente ecuación:

$$\text{PSNE}(\%) = \frac{\sum_n (y(n) - \hat{y}(n))^2}{\sum_n (y(n))^2} \times 100 = \frac{\sum_n (\varepsilon(n))^2}{\sum_n (y(n))^2} \times 100 \quad (4.11)$$

donde $\varepsilon(n)$ es el residuo que el modelo no estima e $y(n)$ es la salida real del sistema. La evaluación del error que se comete en este cálculo ya se presentó en la ecuación 3.3 del capítulo anterior.

Obviamente, cuanto menor sea PSNE mejor será la estimación, pero en este punto hay que volver a remarcar la diferencia, también ya explicada en el capítulo anterior, entre la validación con datos usados en el proceso de identificación y la validación con datos nunca usados en el proceso de identificación. El primer caso es una validación en zona de estimación, pero el verdadero sentido de la validación está en el segundo caso, que en esta trabajo se denominará validación en zona de validación. Habitualmente, las tablas de resultados sobre PSNE, presentarán columnas bajo la denominación estimación o validación, haciendo referencia precisamente a la validación en zona de estimación o en zona de validación respectivamente.

El análisis de residuos, es un estudio adicional para comprobar la calidad del modelo generado. Consiste en estudiar las relaciones entre el residuo $\varepsilon(n)$ y las señales del sistema. Para que el modelo tenga plena validez, no tan sólo es necesario que el error sea pequeño, sino que, si el modelo es correcto, el residuo $\varepsilon(n)$ debe ser independiente de las entradas del sistema^{9 17}. Si no fuese así, el residuo contendría algún tipo de información que el modelo no ha sido capaz de extraer y por tanto, éste no puede considerarse correcto.

Basándose en este criterio, el análisis de residuos para estimaciones lineales se basa en la evaluación de la correlación cruzada entre $\varepsilon(n)$ y $x(n)$

$$\hat{R}_{\varepsilon x}(m) = \frac{1}{N} \sum_{n=1}^N \varepsilon(n+m)x(n) \quad (4.12)$$

Si no existe relación entre los residuos y la entrada, esta correlación debe dar una señal aleatoria de media cero y varianza:

$$P_r = \frac{1}{N} \sum_{k=-\infty}^{k=\infty} R_{\varepsilon}(k)R_x(k) \quad (4.13)$$

donde R_{ε} y R_x representan las covarianzas de las señales residuo y entrada.

En general, se suele representar la expresión $\hat{R}_{\varepsilon x}(m)$ entre dos líneas de valor $\pm 3(P_r)^{1/2}$, que representa un margen de confianza de, aproximadamente, 95%.

Si algún punto de $\hat{R}_{\varepsilon x}(m)$ sobrepasa este margen, es una indicación de que puede haber una dependencia entre el residuo y la señal de entrada que el modelo no ha podido identificar.

Por otro lado, el residuo debe ser una señal estrictamente blanca, y por tanto sus elementos deben ser independientes unos de otros. Para probar este punto se utiliza la autocorrelación de los residuos:

$$\widehat{R}_\varepsilon(m) = \frac{1}{N} \sum_{n=1}^N \varepsilon(n+m)\varepsilon(n) \quad (4.14)$$

cuyo resultado debe ser 0 para todo $m \neq 0$ y debe coincidir con la varianza de $\varepsilon(n)$ cuando $m=0$.

4.4. Determinación de no linealidades a partir del análisis de residuos

Los tests anteriores se refieren exclusivamente al análisis de las dependencias lineales de las estimaciones, pero el sistema real no tiene porque responder dentro de los términos exclusivamente lineales. En el capítulo anterior quedo patente la no linealidad de la respuesta de los sensores ante la presencia de escalones de concentración. El objetivo que se persigue en esta parte del trabajo con los siguientes test de análisis de residuos que se presentarán a continuación, es evidenciar también el comportamiento no lineal de los sensores frente a las entradas de concentración en forma de ruido blanco.

El hecho que los análisis de residuos lineales no detecten anomalías, sólo es indicio de que la parte lineal está bien estimada dentro de los límites que marque el experimento, pero no indica si en el residuo quedan o no componentes no lineales sin resolver. Hay que introducir por tanto algunos tests de no linealidad que permitan detectar este tipo de relaciones complejas. De entrada, debe quedar claro que, tal como se conciben estos análisis, resulta imposible generar un test que compruebe cualquier relación no lineal entre las señales del sistema. Un análisis así, requeriría una gran multitud de ensayos en un esfuerzo imposible de abarcar bajo el concepto exclusivo del análisis de residuos. Por tanto, el análisis de residuos no lineal se resume a algunos tests que tienen la capacidad de detectar las relaciones no lineales más comunes.

Así, al igual que para la estimación lineal, los tests de análisis de residuos no lineales se definen para detectar si los residuos de la estimación contienen información no lineal relativa al sistema que el modelo no ha sido capaz de extraer. El problema no lineal es más delicado que el lineal. Por definición de no linealidad, el ruido que entra en el sistema no puede ser considerado como un término aditivo en la salida. De hecho, en general para un sistema no lineal se cumple que:

$$y(n) = G^x[x(n)] + G^{x\varepsilon}[x(n), \varepsilon(n)] + G^\varepsilon[\varepsilon(n)] \quad (4.15)$$

donde $G^x[x(n)]$ y $G^\varepsilon[\varepsilon(n)]$ son funciones solamente de $x(n)$ y $\varepsilon(n)$ respectivamente, y donde $G^{x\varepsilon}[x(n), \varepsilon(n)]$ contiene productos cruzados de $x(n)$ y $\varepsilon(n)$. En general no se puede resolver la ecuación 4.15 para $\varepsilon(n)$ y por tanto en el caso no lineal, aun y cuando el modelo este bien estimado, el error $e(n)$ y los residuos $\varepsilon(n)$ no serán iguales.

Existen muchos tests que tratan de realizar esta tarea, pero no todos son válidos^{9 17}, en el sentido de que muchos de ellos no aportan información o sugieren resultados que a posteriori se demuestran falsos. Por ejemplo, la correlación entrada-ruido que era perfectamente válida para sistemas lineales no puede ser tenida en cuenta de la misma manera ya que no puede detectar términos de orden par en $G^x[x(n)]$.

Uno de los tests fundamentales para analizar los residuos desde el punto de vista no lineal es la autocorrelación de los mismos:

$$R_{\varepsilon\varepsilon}(m) = E[\varepsilon(n+m)\varepsilon(n)] \quad (4.16)$$

Este test debería dar como resultado un impulso en la posición m del cálculo $R_{\varepsilon\varepsilon}(m)=\sigma^2_{\varepsilon}\delta(m)$. Sin embargo, este test falla bajo dos condiciones:

a) cuando:

$$\begin{aligned} \varepsilon(n) &= x(n-k) + v(n) \\ &\quad \text{o} \\ \varepsilon(n) &= x(n-k)e^{2(n-j)} + v(n) \text{ para todo } k, j \end{aligned} \quad (4.17)$$

b) cuando

$$\varepsilon(n) = x^q(n-k)e(n-j) + v(n) \text{ para todo } k, j \text{ y siendo } q \text{ impar.} \quad (4.18)$$

Estas deficiencias en el test $R_{\varepsilon\varepsilon}(m)$ se pueden corregir añadiendo dos test más:

$$R_{x\varepsilon}(m) = E[x(n)\varepsilon(n-m)] \quad (4.19)$$

que complementa a $R_{\varepsilon\varepsilon}(m)$ y subsana sus carencias para el caso a) y

$$R_{\varepsilon\varepsilon x}(m) = E[\varepsilon(n)\varepsilon(n-1-m)x(n-1-m)] \quad (4.20)$$

que complementa a $R_{\varepsilon\varepsilon}(m)$ y subsana sus carencias en el caso b).

En resumen, para un modelo no lineal, los residuos resultan impredecibles si y solo si se cumple que:

$$\begin{aligned} R_{\varepsilon\varepsilon}(m) &= \sigma^2_{\varepsilon}\delta(m) \\ R_{x\varepsilon}(m) &= 0 \text{ para todo } m \\ R_{\varepsilon\varepsilon x}(m) &= 0 \text{ para todo } m \end{aligned} \quad (4.21)$$

Si este test no diese el resultado deseado, todo lo que se puede decir del modelo es que de alguna manera no es adecuado.

Existen otros tests interesantes a realizar y que determinan ciertas correlaciones típicas son:

$$\begin{aligned} R_{\varepsilon xx}(m) &= E[\varepsilon(n)x^2(n-m)] = 0 \text{ para todo } m \\ R_{\varepsilon\varepsilon xx}(m) &= E[\varepsilon^2(n)x^2(n-m)] = 0 \text{ para todo } m \end{aligned} \quad (4.22)$$

Ambas tratan de determinar correlaciones del residuo con la entrada al cuadrado, y ambas deben dar cero si el modelo está correctamente estimado en su no linealidad. Más adelante, introducido ya algún método de estimación, se presentarán resultados de estos tests sobre un ejemplo sintético.

El conjunto de análisis de residuos más validación ofrece una visión muy completa acerca de la adecuación del modelo estimado a las señales experimentales del sistema. Se debe recordar de nuevo, que para que dicha versión no resulte distorsionada, el análisis de residuos y la validación deben realizarse con datos no empleados en la estimación.

4.5. Estimación no paramétrica mediante análisis de correlación

Introducidos los conceptos teóricos, a continuación se presenta la primera de las aproximaciones que se va a estudiar: es la denominada estimación no paramétrica mediante análisis de correlación.

Es sencillo demostrar^{1 2 16} que la relación entre señales de entrada y salida mediante la estimación de la respuesta impulsiva expresada en la ecuación 4.10 se puede transformar en relaciones entre señales correlacionadas de tal manera que se cumple que:

$$R_{yx}(k) = \sum_{l=-\infty}^{\infty} \hat{g}(l)R_{xx}(k-l) + R_{\varepsilon x}(k) \quad (4.23)$$

Donde R_{yx} representa la correlación entre las señales $y(n)$ $x(n)$, R_{xx} la autocorrelación de la señal $x(n)$ y $R_{\varepsilon x}$ la correlación entre la perturbación $\varepsilon(n)$ y la señal $x(n)$ respectivamente.

La ecuación 4.23 es especialmente interesante si se cumplen dos condiciones:

- Que la perturbación no guarde ninguna relación con la señal de entrada $x(k)$
- Que $x(k)$ sea una señal de ruido blanco de valor medio nulo.

En este caso, se cumple que:

$$\begin{aligned} R_{\varepsilon x}(k) &= 0 \text{ para todo } k \\ R_{xx}(k-l) &= 0 \text{ para } l \neq k \\ R_{xx}(k-l) &= \sigma_x^2 \text{ para } l=k \end{aligned} \quad (4.24)$$

Siendo σ_x^2 la varianza del ruido blanco que representa x .

Por tanto, a través de 4.24 la ecuación 4.23 se transforma en la expresión:

$$R_{yx}(k) = \hat{g}(k)\sigma_x^2 \quad (4.25)$$

que permite una estimación directa de la respuesta impulsiva según la fórmula:

$$\hat{g}(k) = \frac{R_{yx}(k)}{\sigma_x^2} \quad (4.26)$$

donde $R_{yx}(k)$ y σ_x^2 se pueden calcular fácilmente ya que son funciones relativas a las señales experimentales del sistema.

A partir de esta estimación, se puede realizar, tal y como se propone anteriormente, una modelización de la salida del sistema usando la convolución de la respuesta impulsiva estimada con las entradas:

$$\hat{y}(k) = \sum_{j=0}^{\infty} \hat{g}(j)x(k-j) \quad (4.27)$$

Esta ecuación representa una estimación lineal del comportamiento del sistema y como tal estimación debe entenderse, presentando una cierta desviación respecto de la realidad que queda definida por la señal $\varepsilon(n)$ de la ecuación 4.10.

4.6. Aplicación del análisis de correlación a un problema sintético

Antes de abordar el problema de la modelización del sistema sensor a través del análisis de correlación, se ha generado un sistema artificial como problema sintético sobre el que ilustrar el comportamiento del análisis empleado. La importancia de este sistema es relativa, en el sentido que existen infinitos sistemas posibles a ensayar. No se trata por tanto de demostrar las propiedades del análisis, sino de comprobar el rendimiento de sus estimaciones sobre un sistema conocido y que presenta características similares a los sistemas reales estudiados en este trabajo. En este último sentido, no es importante la dinámica concreta que introduce en el problema, sino la existencia de dicha dinámica tanto en el orden lineal como en el no lineal. Tampoco es importante la potencia establecida para las diferentes dependencias dinámicas, siempre que se respete el hecho que la parte no lineal no sea de mayor potencia que la parte lineal, puesto que esta es una característica del sistema real, como se verá a posteriori. Si que se ha restringido la no linealidad al orden dos puesto que, también como se verá a posteriori, este orden de no linealidad es suficiente para explicar el comportamiento del sistema real con muy buena aproximación.

El mismo problema sintético se utilizará como test previo a la implementación de los nuevos algoritmos de estimación que se vayan presentando.

Supóngase, entonces, un sistema definido por una ecuación como la siguiente:

$$y(n)=2 \cdot x(n-1)+x(n-2)+0.5 \cdot x(n-2) \cdot x(n-3)+ \\ +0.9 \cdot x(n-4) \cdot x(n-7)+0.7 \cdot x(n-5) \cdot x(n-5)+v(n) \quad (4.28)$$

La relación entre la entrada y la salida contiene una parte lineal, representada por el primer y el segundo término de los sumandos de la ecuación, y una parte no lineal, representada por el resto de sumandos. Existe además un término de ruido $e(n)$ que se ha definido en este caso y por simplicidad, como una señal de ruido blanco de media cero, desviación estándar 0.1 y distribución gaussiana.

Puesto que las ecuaciones para el análisis de correlación precisan de ruido blanco en la señal de entrada, se han generado una señal de entrada $x(n)$ de ruido blanco con media cero y desviación estándar uno y distribución gaussiana. Hay que ser consciente que el hecho de disponer de una señal $x(n)$ de media cero es un caso particular especialmente cómodo a la hora de realizar las estimaciones, pero no es el caso general. En el caso particular presentado, el hecho de disponer de señales de entrada de media cero, permite una total desconexión entre la parte lineal del sistema y la no lineal con la consecuente sencillez en la interpretación de los resultados.

Aplicando la definición del sistema según la expresión 4.29 se ha obtenido la señal de salida. Ambas señales se muestran en la figura 4.3:

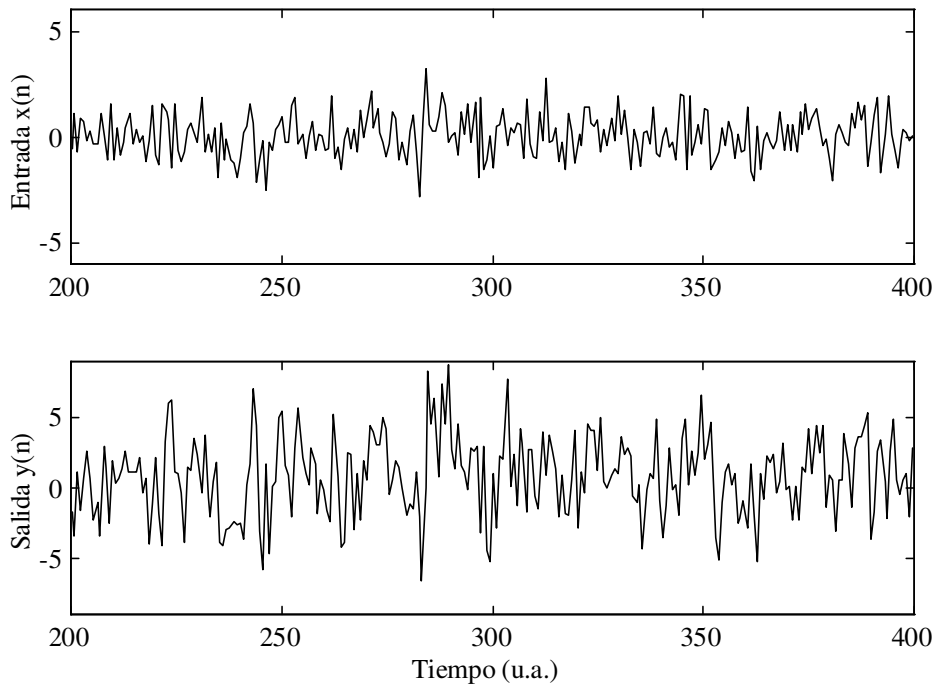


Figura 4.3 señales de entrada $x(n)$ y salida $y(n)$ del sistema 4.29

Antes de realizar los desarrollos para tratar de identificar el sistema, debe substraerse la media sobre la salida para obtener una señal de valor medio cero. Esta operación resulta crucial para el desarrollo de los cálculos posteriores, y sólo supone una corrección de nivel respecto de las señales del sistema. En base a las nuevas señales de media cero y al análisis de correlación, puede estimarse la respuesta impulsiva.

La figura 4.4 muestra la respuesta impulsiva calculada a través del método de correlación con 1000 puntos. Se observa que sólo aparecen fuera de los límites de incertidumbre los valores correspondientes a la parte lineal del sistema. De nuevo hay que remarcar en este punto que el resultado obtenido es consecuencia de disponer a la entrada de una señal de media cero y que, dentro de este caso particular, no debe aparecer en la respuesta impulsiva ninguna componente procedente de la contribución no lineal de orden dos. Sin embargo se debe tener claro que, en general, si que necesariamente han de aparecer en la respuesta impulsiva componentes de retraso procedentes de las relaciones de no linealidad.

Calculada la respuesta impulsiva, la estimación de la salida del sistema se realiza a través de la convolución de dicha respuesta con la entrada del sistema. Sobre los valores obtenidos tanto en zona de estimación como en zona de validación, cabe considerar varias fuentes de error. En primer lugar, obviamente, la señal de ruido $v(n)$ sobre la que concurren dos circunstancias:

- a) el modelo no realiza ninguna hipótesis de comportamiento sobre él (no será así en las estimaciones paramétricas) y por tanto ninguna información del mismo está presente en la estimación.
- b) contiene un cierto factor de impredecibilidad (en los modelos paramétricos recibe el nombre de perturbación) que nunca puede ser incluido en las estimaciones.

Otra fuente de error radica en la inexacta estimación de los coeficientes que definen al modelo. Esta inexactitud puede ser debida a varias causas. Una de ellas es inherente a la propia estadística al disponer de un número finito de muestras. Este es, de hecho, uno de

los problemas principales de la estimación mediante correlación: la gran dependencia con el número de datos utilizados. Sin embargo, puede también ser posible, como nueva fuente de error, que el modelo utilizado no sea adecuado para el problema que se trata. Este tercer factor es evidente para este problema donde la propia ecuación 4.28 muestra la existencia de dependencias no lineales mientras que el método de estimación es exclusivamente lineal.

Valorada la discrepancia total según la función PSNE, esta ha resultado ser de, aproximadamente, un 40% (variando ligeramente según la realización concreta de los datos aleatorios). El valor es alto, pero hay que tener en cuenta que el peso de los factores no lineales en el sistema es muy elevado y que por tanto la no adecuación del modelo lineal es un factor substancial de error. La información no lineal que el modelo no es capaz de ver, tiene que estar almacenada en el residuo de la estimación y mediante un análisis cuidadoso de dicho residuo ha de poder plasmarse claramente.

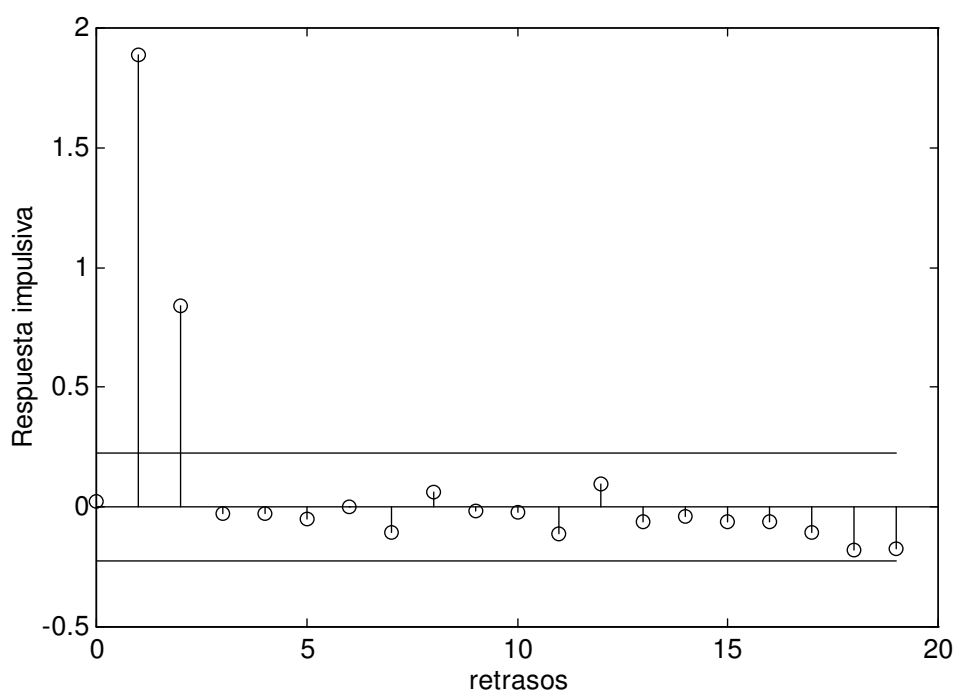


Figura 4. 4 Respuesta impulsiva del sistema 4.29

Las siguientes figuras muestran algunos de los test más significativos realizados para analizar el residuo. Las dos primeras gráficas hacen referencia a un análisis estrictamente lineal, en el que se realiza la autocorrelación del error y la correlación del error con la entrada. Se demuestra que los resultados concuerdan con los valores teóricos: la autocorrelación debe ofrecer una delta en la posición cero mientras que la correlación entrada residuo no debe aportar, si la estimación lineal es correcta, valores fuera del margen de significado estadístico. Los márgenes representados en las figuras muestran un límite de confianza del 95%.

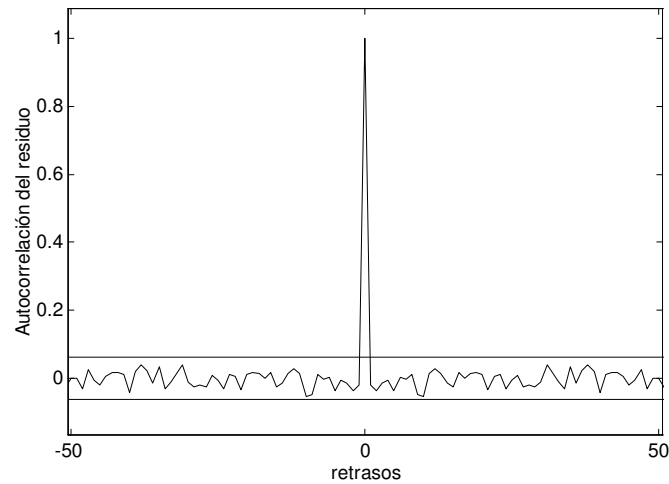


Figura 4.5 Análisis de residuos: correlación del residuo con el residuo

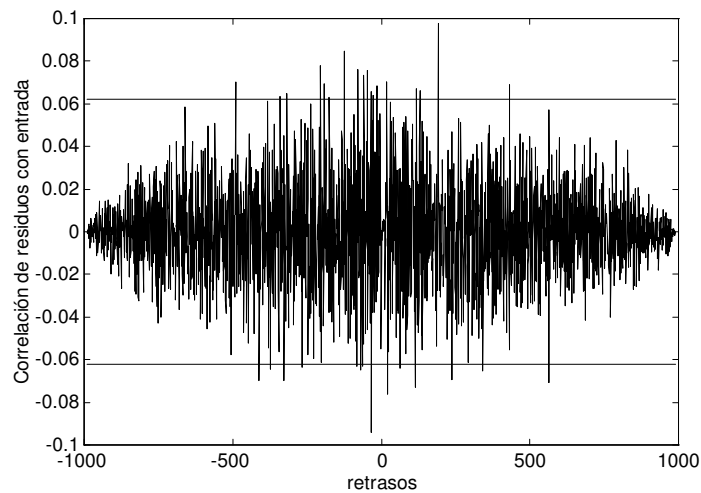


Figura 4.6 Análisis de residuos: correlación del residuo con la entrada

La efectividad del análisis de residuos lineal sólo demuestra que la estimación lineal es correcta, pero no dice nada respecto a la existencia o no de información no lineal oculta. Uno de los tests descritos anteriormente y que va a indicar la existencia de esta información oculta es el $R_{\epsilon x}(n)$. Este test deberá demostrar la presencia de información del tipo $x(n-j)x(n-j)$ para $j=1,2,\dots$. Ciertamente, el sistema presenta un término de este estilo para $j=5$, y en esa posición de retraso, el análisis debe presentar un valor fuera de los márgenes de significado. En la figura 4.7 se demuestra que efectivamente, existe en el residuo información relativa al término $0.5 \cdot x(n-5) \cdot x(n-5)$:

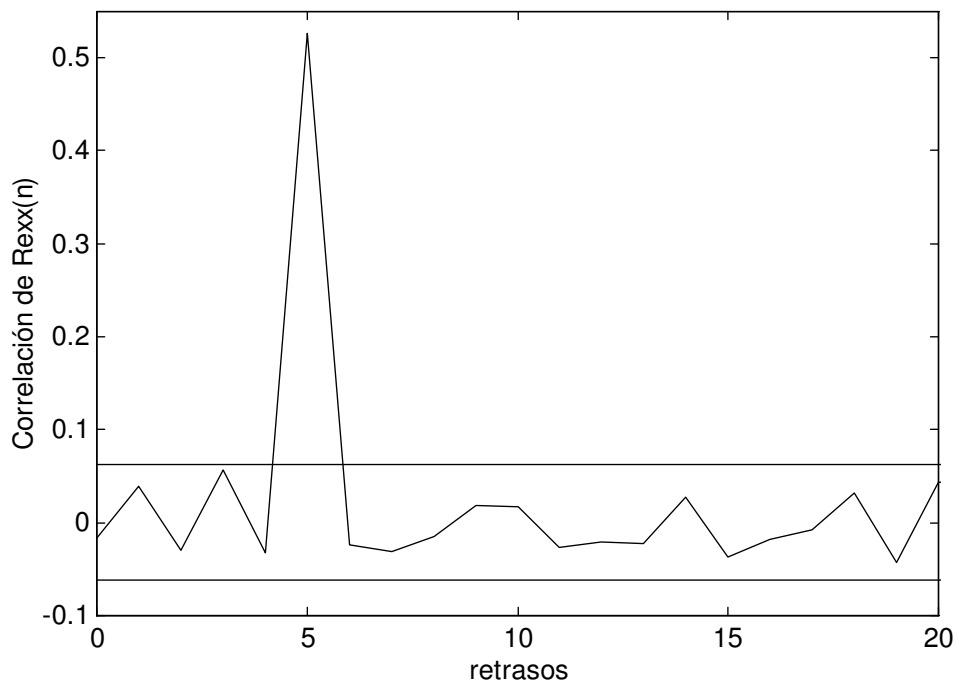


Figura 4.7 Residuo $R_{\text{exx}}(n)$

No es posible encontrar un test que encuentre en general, cualquier tipo de relación no lineal. De hecho los otros términos no lineales quedarían sólo patentes en el residuo a través de nuevos test del estilo $R_{\text{ex}(n)\text{x}(n-j)}$ con $j=1$ y $j=3$. Ante la imposibilidad de tratar generar la infinidad de test necesarios para este análisis, la estrategia por la que se opta, como ya se comentó anteriormente, es la de analizar el residuo sólo a través de una serie de test típicos. El hecho de que el resultado de estos tests sea negativo, no indica que no haya cierta información no lineal que convenga explorar. En todo caso, siempre se debe valorar la globalidad de las pruebas realizadas, es decir, la validación, los tests de análisis de residuos lineales y los no lineales. Es el conjunto de todo ello lo que da un cierto criterio para estimar si es interesante explorar grados de estimación superiores al lineal. En el ejemplo expuesto ha quedado demostrada la existencia de información no lineal y además, aun y siendo la estimación lineal esencialmente correcta, la validación ofrece valores de PSNE demasiado altos como para no intentar una estimación de grado superior. Veamos a continuación la aplicación del análisis de correlación al problema real del sistema de análisis.

4.7. Aplicación del análisis de correlación al sistema

4.7.1. Identificación de un sistema SISO

Antes de iniciar la estimación dinámica para cada sensor desde el punto de vista de un sistema MISO, resulta conveniente evaluar el proceso identificativo desde una perspectiva de un sistema SISO más sencillo. Mediante este procedimiento, se podrán evaluar mejor las características de los sistemas sensores así como el rendimiento del proceso identificativo aplicado al problema. El ejemplo anterior versaba exclusivamente sobre un sistema SISO, así que los desarrollos efectuados tienen plena aplicación en este problema real.

Construcción de un experimento

Siguiendo el esquema identificativo de la figura 3.6, comienza el proceso con la construcción de un experimento. Se han descrito anteriormente una serie de experimentos aplicados al sistema de sensores de gas. Para la descripción de cada sensor en la cámara sensorial como un sistema SISO, el proceso identificativo está garantizado a través del experimento 2.

Tratamiento y presentación de datos

Se han presentado anteriormente los resultados de cada uno de los experimentos para todos los gases y todos los sensores. Además, se han efectuado las modificaciones oportunas para poder utilizar los datos de los sensores 1 2 5 y 6 en los diversos procesos identificativos que se van a utilizar. En éste apartado que se inicia se utilizan los primeros 300 puntos de la entrada y de la respuesta de los sensores generadas por el experimento 2, tanto para octano como para tolueno. La figura 4.8 presenta 100 puntos de la señal de entrada en forma de concentración de gas tolueno a la que se somete al sensor 1, así como los 100 puntos que se corresponden de la respuesta de dicho sensor para ese tramo de concentraciones.

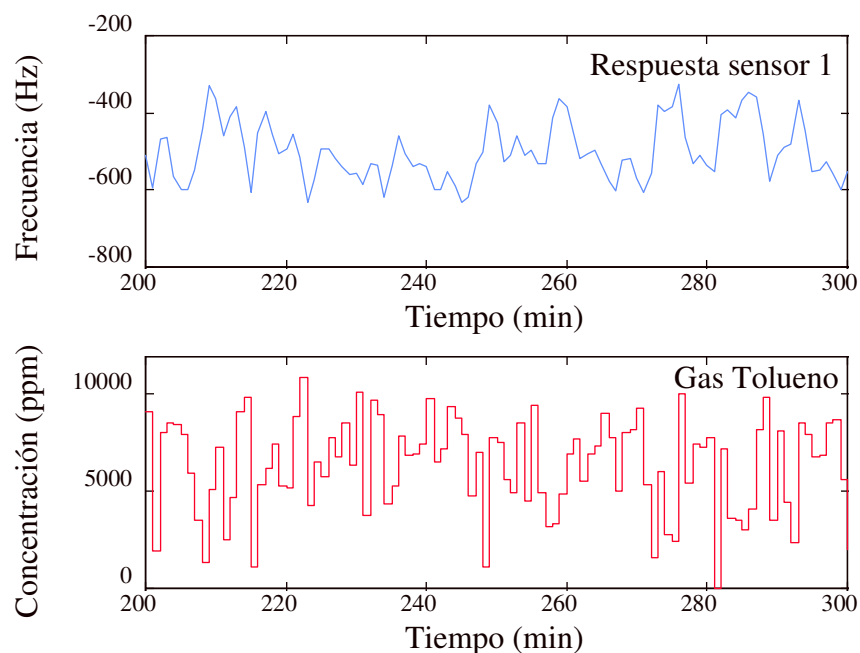


Figura 4. 8 Señales del sistema: a) señal de salida del sistema: sensor 1
b) entrada del sistema: gas tolueno

Ajuste del modelo a los datos

Aplicando la expresión 4.27 para ajustar un modelo con las 300 muestras de señales de entrada y salida, se genera una estimación de la respuesta impulsiva asociada a los datos reales del experimento. En la figura 4.9 se presenta el resultado de este cálculo. La gráfica muestra la estimación de la respuesta impulsiva para el sensor 1 sometido a una

concentración de tolueno que varía aleatoriamente alrededor de un valor medio de 6500 ppm con una desviación estándar de 2800 ppm.

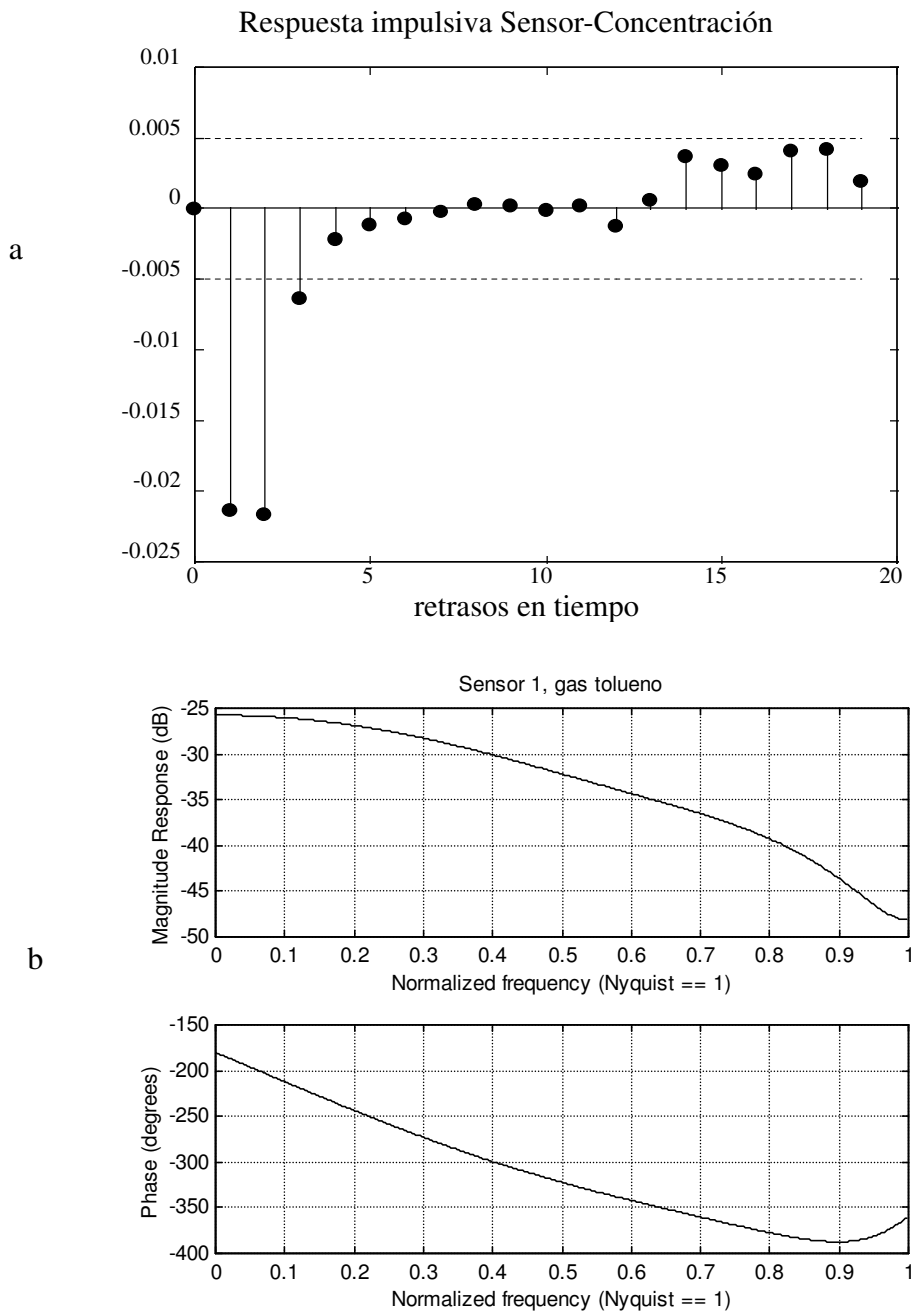


Figura 4. 9 Respuesta impulsiva estimada para el sensor 1 y gas tolueno
a: dominio temporal, b: diagrama de bode

De esta figura se puede observar que el sensor 1, no responde inmediatamente a la presencia de concentración. Por otra parte, se descubre que la salida depende solamente de las tres siguientes muestras, aunque la tercera casi es despreciable. El resto de muestras quedan dentro de la región de confianza cuyos límites marcan el nivel de significación del 95%. Obsérvese que para retrasos alrededor de 15 parece haber un repunte de los valores de la respuesta impulsiva, sin embargo, su sentido es contrario al de los valores principales y no superan el límite de confianza. Por todo ello, no deberían tenerse en

cuenta como verdaderos valores de la respuesta impulsiva. Según esta estimación lineal, el sensor 1, como sistema LTI SISO causal, responde al tolueno según la fórmula:

$$y(n)=g(1)x(n-1)+g(2)x(n-2)+g(3)x(n-3) \quad (4. 29)$$

con $g(1) = - 0.0214$ Hz/ppm, $g(2) = - 0.0217$ Hz/ppm y $g(3) = -0.0064$ Hz/ppm.

Esta estimación de la respuesta impulsiva constituye un modelo para el sistema sensor 1 sometido al gas tolueno que define completamente a dicho sistema dentro del concepto de sistemas LTI SISO bajo el que está sometido el sensor en el sistema de medida.

Validación del modelo

Para completar el ciclo identificativo debe realizarse la validación del modelo y el análisis de los residuos. Para la validación se va a comparar la señal real con la simulada a través del modelo lineal generado. Para ello se utiliza la convolución de la respuesta impulsiva con los datos de entrada. Esta operación se realiza sobre un conjunto de datos (unos 80 datos) no utilizados para la estimación y escogidos a continuación de los 300 usados para la estimación.

El tanto por ciento de señal no estimada (PSNE) para esta aproximación lineal y para el sensor 1 sometido al experimento 2 para gas tolueno es de un 7% en la zona de datos de validación. Este porcentaje baja hasta un 5 % si se trata de la zona de estimación.

La tabla 4.1 presenta resultados de las zonas de datos de estimación y validación para cada uno de los sensores sometidos al experimento 2 y para cada gas del experimento :

Sensores	Octano		Tolueno	
	Estimación	Validación	Estimación	Validación
S1 (PSNE)	4% ± 0.2%	8% ± 0.6%	5% ± 0.2%	4% ± 0.3%
S2 (PSNE)	5% ± 0.2%	8% ± 0.6%	4% ± 0.2%	3% ± 0.2%
S5 (PSNE)	7% ± 0.3%	12% ± 0.9%	4% ± 0.2%	3% ± 0.2%
S6 (PSNE)	7% ± 0.3	12% ± 0.9%	4% ± 0.2%	3% ± 0.2%

Cada valor de la figura de mérito PSNE está sometido a una cierta variabilidad dependiendo de los puntos que se tomen para realizar el proceso identificativo. A pesar de esta variabilidad, las tendencias que muestra esta tabla son correctas.

Para el análisis de residuos se utilizan, en principio, los tests de autocorrelación del error y la correlación entrada error. Los resultados para el sensor 2 y gas octano se muestran en las dos siguientes figuras:

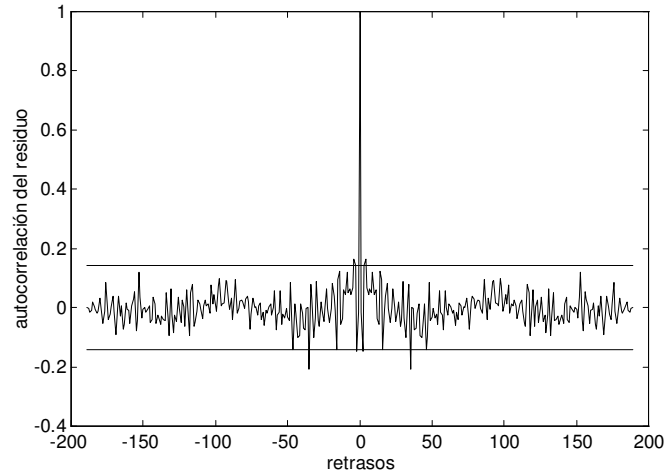


Figura 4.10 Análisis de residuos: test de autocorrelación de residuos

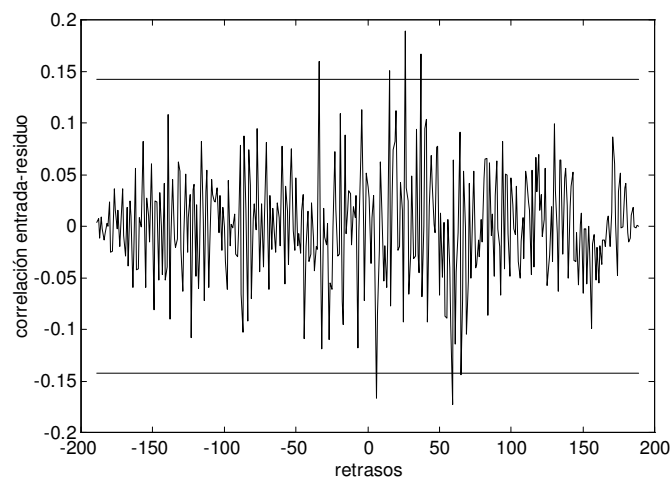


Figura 4.11 Análisis de residuos: test de correlación de residuo con entrada

El análisis para el resto de sensores y para gas octano o tolueno ofrecen resultados similares. Se puede concluir que los resultados de estos tests cumplen con las condiciones para aceptar el residuo como correcto desde el punto de vista lineal. La autocorrelación del residuo está siempre dentro de los márgenes de confianza que se supone no significativo, excepto en el retraso 0. El resultado de la correlación entrada residuo, que tiene en cuenta si en el residuo queda información lineal que el modelo no ha extraído, queda siempre dentro del margen de confianza.

La adecuación de los resultados de estos tests indican que la parte lineal está correctamente estimada, pero no ofrecen información acerca de la posible información no lineal que puede quedar. De hecho, si se realiza el test de correlación de la entrada al cuadrado con el error $R_{\epsilon\epsilon}(n)$, se puede observar como en el residuo queda información no lineal (que al modelo lineal le es imposible determinar) por extraer.

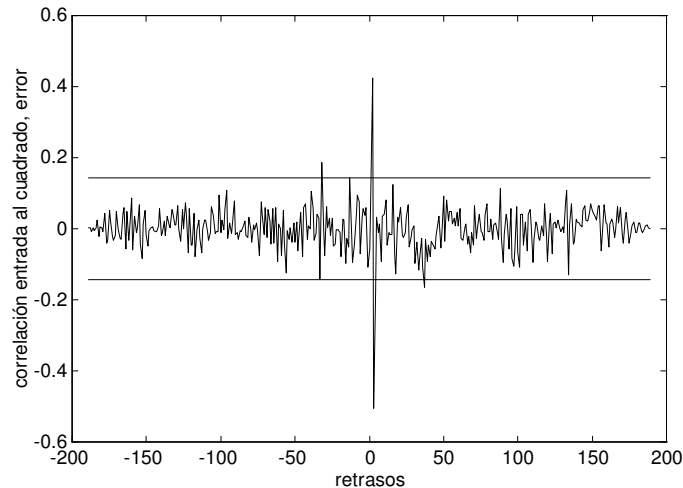


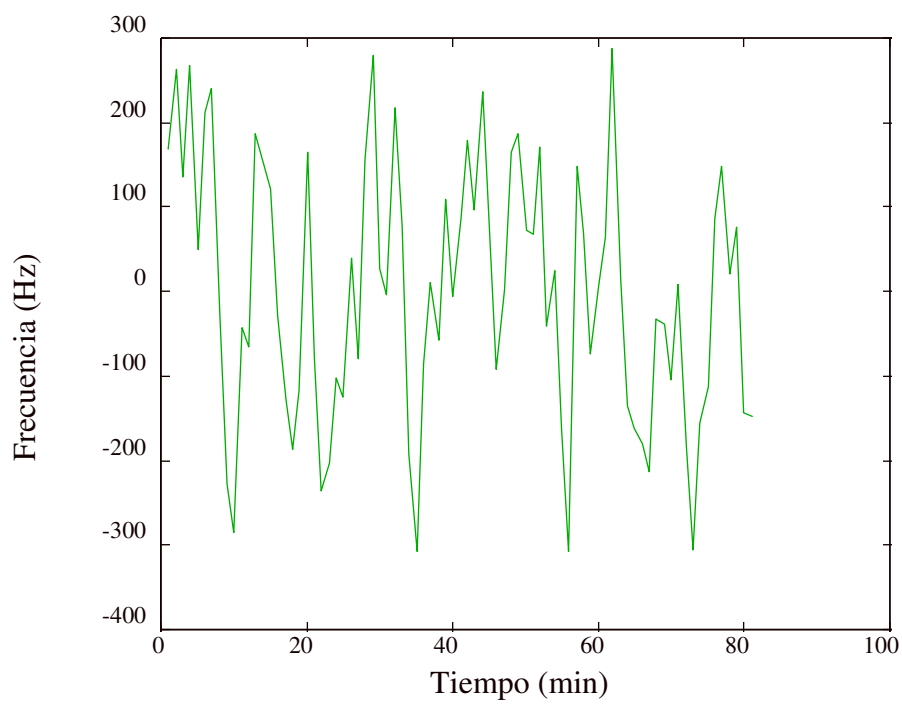
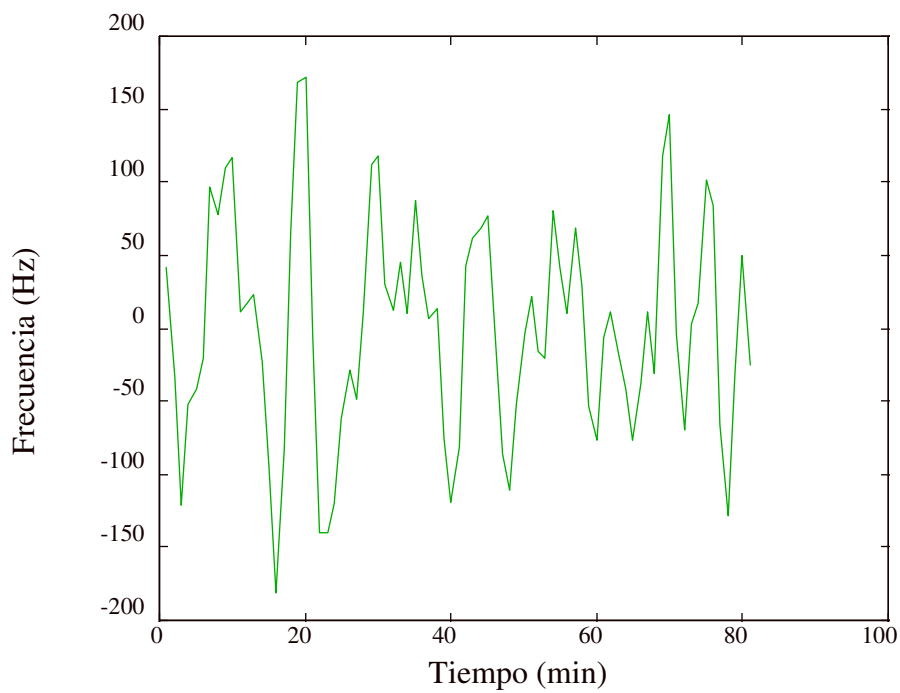
Figura 4.12 Análisis de residuos: test no lineal

Aceptación del modelo

Se observa pues que para un solo gas, los modelos lineales generados a través de análisis de correlación permiten estimar entre un 88% y un 96% aproximadamente de la señal, dependiendo del sensor y del gas. Siendo un resultado aceptablemente bueno, hay que considerar que es un sistema sencillo de tipo SISO y que el análisis de residuos parece indicar que falta por extraer cierta parte de información no lineal que el modelo lineal, obviamente, no puede ver. Debe resaltarse de los resultados expuestos en la tabla, el hecho de obtener un mayor error para el gas que presenta, precisamente, más coeficiente de no linealidad. Es conveniente evaluar la importancia de la validación y del análisis de residuos dentro de los límites del experimento. Así, los orígenes de la potencia no estimada pueden provenir de una no adecuación del modelo (problemas de no-linealidad, histéresis o derivas) o proviene simplemente del ruido inherente al sistema. Si el motivo puede atribuirse exclusivamente al ruido del sistema, se ha llegado prácticamente al límite de la estimación. Si por contra los motivos son una no adecuación del modelo, debe investigarse nuevas líneas de identificación.

Para disponer de un criterio que permita discriminar entre ambas opciones, se debe conocer la precisión del sistema experimental, que viene determinada principalmente por el error en la generación de las concentraciones a través de los MFC. Ya quedó establecido en el capítulo dos que esta precisión experimental se sitúa alrededor de los 200 ppm, lo que supone una PSNE que oscilará entre el 1% y el 2%. Por tanto, y a falta de corroboración empírica, parece que aun hay un cierto margen de maniobra para las técnicas identificativas, y que aun cabe la posibilidad de disminuir el tanto por ciento de error y acercarse al límite de la precisión. Se confirma por tanto el hecho indicado por el análisis de correlación: que el origen de la discrepancia entre el modelo y la realidad no proviene solamente del ruido sino que el modelo generado no es adecuado, pues no es capaz de representar toda la complejidad del sistema.

En las dos páginas siguientes se presentan resultados en zona de validación para el sensor 1 y para ambos gases. La figura superior de cada hoja se corresponde con el gas tolueno mientras que la figura inferior es el gas octano. Las gráficas en verde representan las señales reales mientras que las rojas representan las señales estimadas a través de la respuesta impulsiva calculada mediante el método de correlación descrito.



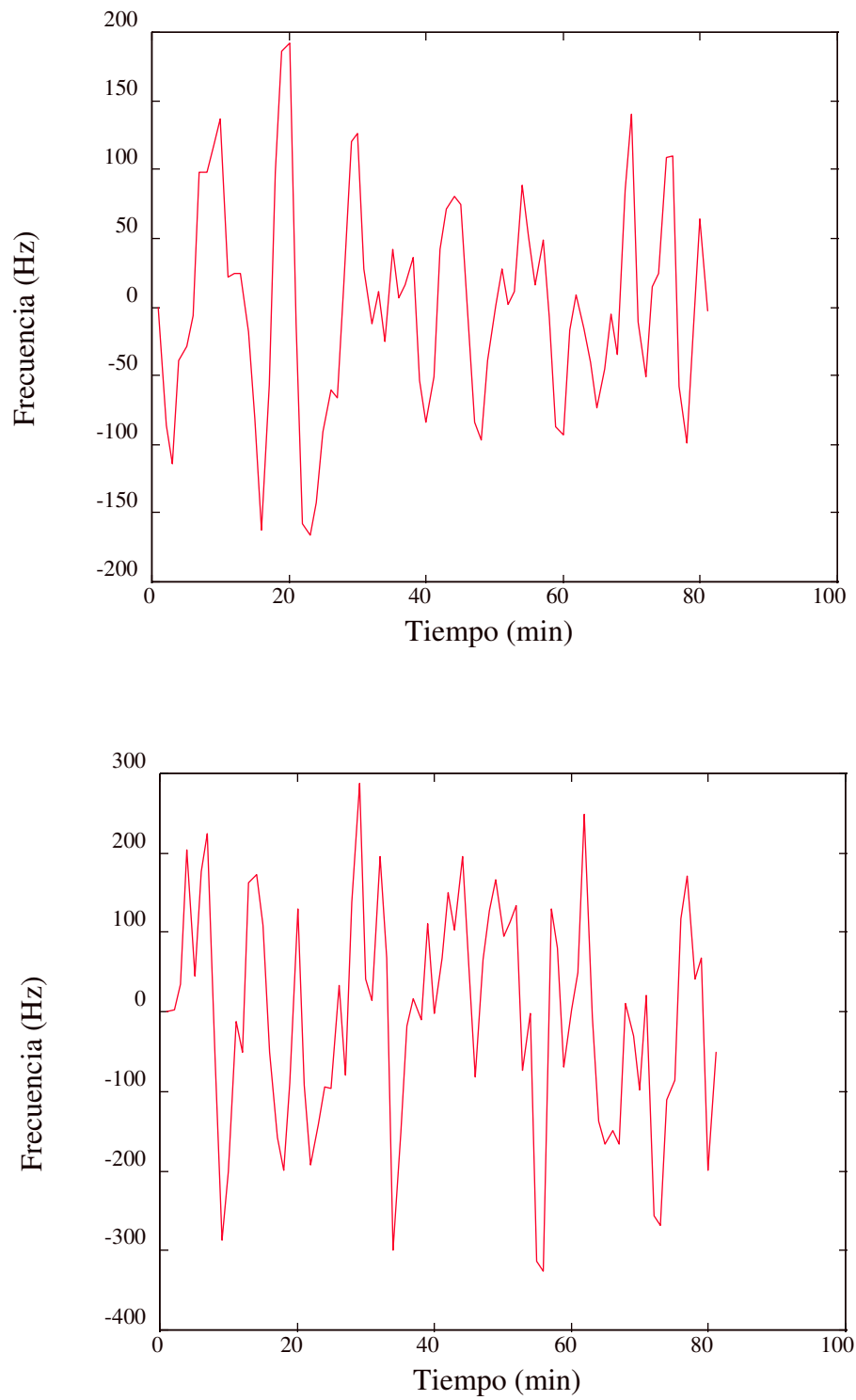


Figura 4.13 Resultados de la estimación mediante análisis de correlación

4.7.2. Identificación de un sistema MISO

Un paso adelante en complejidad, antes de pasar a la no linealidad, dentro de la tarea identificativa consiste en tratar al sistema sensor no como sistemas SISO sino como sistemas MISO, y en poder realizar predicciones cuantitativas bajo estas condiciones. El experimento 3 comprueba el rendimiento de la matriz de sensores para dos gases, octano y tolueno, cuyas concentraciones varían aleatoriamente. El valor medio de la concentración de ambos gases es 6500 ppm y la desviación estándar es 2800 ppm.

Construcción del experimento, tratamiento y presentación de datos

Las entradas y salidas para los sensores así como el tratamiento previo de los errores para el experimento 3 ya se han presentado en el capítulo anterior. En la figura 4.14 se presenta la zona de estimación para el sensor 2 que se va a utilizar para realizar el proceso de identificación MISO.

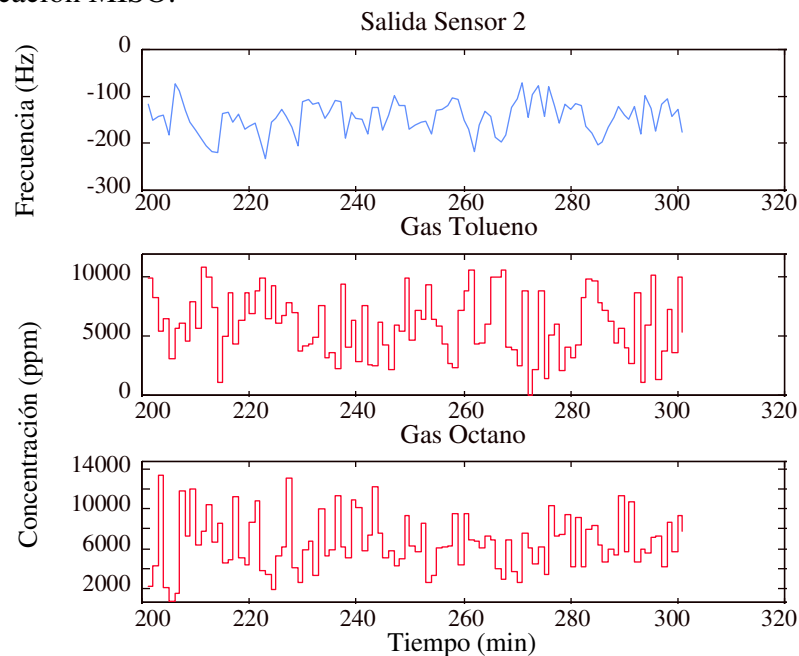


Figura 4. 14 Señales del sistema:

- a) señal de salida del sistema: sensor 2
- b) entrada del sistema: gas octano
- c) entrada del sistema: gas tolueno

Ajuste del modelo a los datos

Utilizando los métodos que se han desarrollado para los sistemas SISO, pero extendidos a un sistema MISO, es posible ajustar un nuevo modelo a la salida del sistema. Este, está definido, según indica el experimento anterior, como un sistema de dos entradas y una salida. Matemáticamente, este nuevo modelo dinámico, lineal, causal de tipo MISO ajustado a dos entradas y una salida, se expresa de la siguiente forma general:

$$y(n) = \sum_{j=0}^{\infty} h_1(j)x_1(n-j) + \sum_{k=0}^{\infty} h_2(k)x_2(n-k) \quad (4.30)$$

donde $h_1(n)$ y $h_2(n)$ representan las respuestas impulsivas para cada una de las entradas $x_1(n)$ y $x_2(n)$ respecto de la salida $y(n)$. Mediante el análisis de correlación es posible extraer de los datos del experimento una estimación para cada respuesta impulsiva del sistema MISO planteado.

La figura 4.15 presenta la respuesta impulsiva tanto para gas tolueno como octano del sensor 1 sometido al experimento 3 calculada según el método del análisis de correlación. Para el cálculo de todas y cada una de las respuestas impulsivas de los diferentes sensores se han utilizado, como antes, trescientos puntos (que corresponden a 300 minutos) de las entradas de concentración y sus correspondientes puntos de salida de cada sensor.

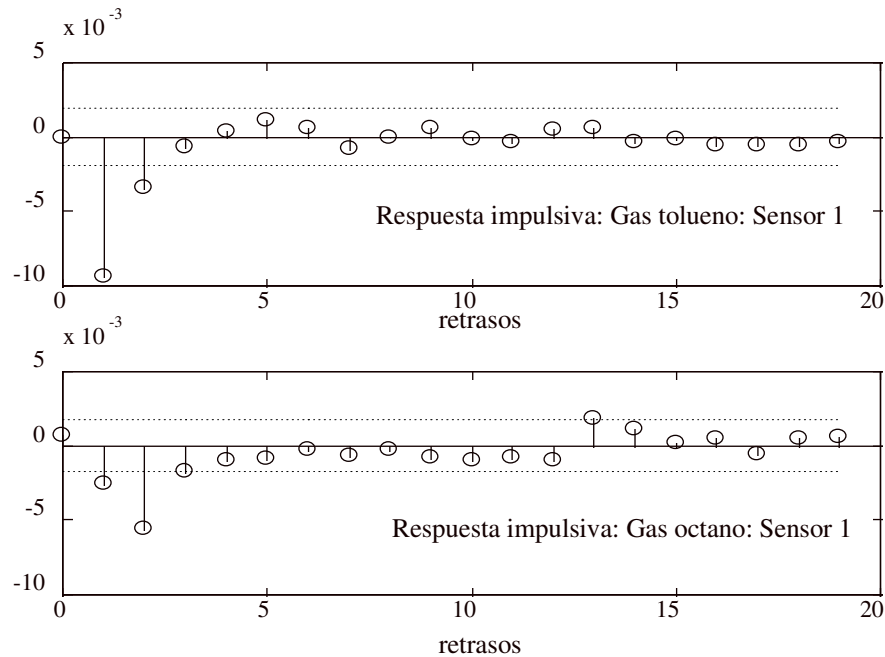


Figura 4.15 Respuestas impulsivas

De la observación de estos resultados se puede concluir que, tanto el tolueno como el octano presentan un cierto retraso en el inicio de la respuesta al estímulo. Sin embargo, mientras para el tolueno la influencia del primer valor de retraso es la más importante, para el octano la componente de la respuesta impulsiva que tiene mayor influencia aparece en el segundo término de retraso. Puede observarse que aunque las respuestas son cualitativamente iguales a las encontradas en el caso SISO, existen sustanciales diferencias cuantitativas. Estas diferencias pueden atribuirse o bien a poca eficiencia del algoritmo de estimación o bien a algún efecto cruzado entre gases. Esta circunstancia deberá ser explorada con más detalle más adelante.

Validación de los modelos

Para la validación del modelo se han utilizado los siguientes 80 datos a los utilizados para la estimación. El método de validación es similar al utilizado para el sistema SISO de tal manera que, para cada sistema sensor se compara la respuesta a partir de las entradas de gas tolueno y octano, con la estimación realizada a partir del análisis de correlación. Para obtener datos cuantitativos acerca de esta comparación se calcula la figura de mérito PSNE. La tabla 4.2 muestra los valores de la PSNE para este experimento, presentándose datos tanto para la zona de estimación como para la zona de validación y para todos los sensores.

Tabla 4. 2 :Resultados para el experimento 3

Sensores	Experimento 3 estimación	Experimento 3 validación
S1 (PSNE)	16.9% ± 0.7 %	31.9% ± 2.5%
S2 (PSNE)	20.2% ± 0.8%	34.8% ± 2.7%
S5 (PSNE)	15.8% ± 0.6%	28.2% ± 2.2%
S6 (PSNE)	18.4% ± 0.7%	35.6% ± 2.8%

Aceptación del modelo

Si el experimento bajo el punto de vista SISO presentaba unos resultados que pueden considerarse aceptables aunque mejorables, parece evidente que la complejidad que introduce un sistema MISO es demasiado elevada como para tratarla con el análisis de correlación lineal. Se debe comentar como dificultad, la influencia de cada gas sobre el otro en el cálculo del sistema MISO. Efectivamente, en el caso de estudiar el gas 1 (supóngase octano), la señal del gas 2 (tolueno) aparece como una señal de ruido de mucha potencia y no blanca, y viceversa. No es necesario hacer un análisis de los residuos; queda claro, a la vista de la tabla de resultados, que para un experimento MISO en los términos en los que se plantea en este trabajo, una estimación mediante análisis de correlación no es un método adecuado y el modelo extraído debe ser desestimado por su pobre rendimiento. Hay que buscar alternativas al análisis de correlación, y una de esas alternativas es la estimación paramétrica.

4.8. Estimación paramétrica de un sistema LTI

Después de la estimación no paramétrica mediante el análisis de correlación y vistos los rendimientos de los modelos sobre un ejemplo sintético y una realización práctica, se ha concluido que para un sistema tipo MISO se deben explorar estimaciones alternativas. La alternativa inmediata al método de correlación es la estimación paramétrica. Antes de llevar a la aplicación este tipo de estimación se debe desarrollar, igual que con el análisis de correlación, una breve introducción de la teoría sobre la que se apoya dicha estimación. Tal y como ya se ha establecido, un modelo general para un sistema SISO causal discreto, lineal e invariante en el tiempo se puede escribir como:

$$y(n)=w(n)+v(n) \quad (4. 31)$$

donde $v(n)$ representa el término que tiene en cuenta las perturbaciones del sistema y $w(n)$ se corresponde con el modelo del sistema libre de ruido.

En el análisis de correlación anterior se ha utilizado un estimador de la respuesta impulsiva en base a correlaciones cruzadas entre entrada y salida y la autocorrelación de las entradas. En este tipo de análisis la única suposición relativa a la perturbación es que esta y la señal de entrada no están correlacionadas. Esta suposición basta para realizar la estimación, pero en el propio cálculo de la estimación, los valores, pasados, presentes o futuros, de la perturbación no intervienen en ningún momento. Este proceso ha llevado, como ya se ha visto, a unos resultados que no son completamente satisfactorios y antes de atacar la estimación desde la no linealidad, conviene explorar si la introducción de una cierta modelización para las perturbaciones proporciona mejores resultados.

Sin embargo, como ya es bien sabido, las perturbaciones del sistema son intrínsecamente aleatorias y por tanto, resulta difícil hacer predicciones de sus valores futuros, ya que la característica principal de las perturbaciones, radica precisamente en su impredecibilidad. De todas formas, en todo momento es posible plantear el realizar suposiciones acerca de los valores de la perturbación basándose en los valores pasados ya conocidos¹. Al fin y al cabo, siempre será posible describir muestras de ruido futuras dentro del marco de la teoría de probabilidad. Una descripción completa del ruido en este marco consiste en obtener la función densidad de probabilidad para $v(n+k)$ con $k \geq 1$ conocidos los valores $v(s)$ con $s \leq n$. Suele, no obstante, utilizarse una aproximación no tan general pero útil y versátil para la gran mayoría de casos prácticos:

$$v(n) = \sum_{k=0}^{\infty} h(k)e(n-k) \quad (4.32)$$

En cierta manera, el ruido se modela como la salida de un sistema LTI excitado por perturbación blanca con una cierta densidad de probabilidad.

Esta aproximación a las perturbaciones como si de un sistema con entrada $e(n)$ y salida $v(n)$ se tratara es muy potente e incrementa la finura de los modelos de una forma inalcanzable para la estimación por análisis de correlación.

Se introduce un operador de retraso q definido como:

$$x(n-1) = q^{-1}x(n) \quad (4.33)$$

La introducción de este operador permite solventar los problemas formales que la transformada z presenta en el tratamiento de señales aleatorias.

En virtud de este operador, la señal $w(n)$ y la señal de entrada se relacionan mediante la función de transferencia definida para el operador q , según la siguiente expresión:

$$w(n) = G(q)x(n) \quad (4.34)$$

$$\text{con } G(q) = \sum_{k=1}^{\infty} g(k)q^{-k}$$

Por otro lado, y de la misma forma:

$$v(n) = H(q)e(n) \quad (4.35)$$

$$\text{con } H(q) = 1 + \sum_{k=1}^{\infty} h(k)q^{-k}$$

y siendo $e(n)$ ruido blanco con una cierta función densidad de probabilidad.

Introducida así esta notación, el sistema viene descrito por la siguiente ecuación:

$$y(n) = G(q)x(n) + H(q)e(n) \quad (4.36)$$

El propósito es, por tanto, determinar tres funciones: $G(q)$, $H(q)$ y la función densidad de probabilidad (FDP) para $e(n)$. Si un sistema resulta tener una respuesta impulsiva finita, la enumeración de $g(k)$ y $h(k)$ junto con la FDP de $e(n)$ resulta suficiente, pero en muchos casos esto no resulta práctico bien porque esta enumeración es tediosa o bien porque el sistema presenta respuestas impulsivas infinitas. Resultaría mucho más cómodo trabajar con estructuras que permitiesen la especificación de G y H en términos de un número

finito de parámetros. Una descripción típica que cumple esta condición son las funciones de transferencia racionales. Por otro lado la manera más común de definir la función densidad de probabilidad de $e(n)$ es a través de sus características numéricas. Usualmente, se asume que $e(t)$ es gaussiano, en cuyo caso, la función densidad de probabilidad queda completamente caracterizada por sus momentos de primer y segundo orden: la media y la varianza.

Hasta este punto, no se ha avanzado respecto a la descripción no paramétrica de un sistema LTI más que en el hecho de introducir una cierta modelización de la perturbación y en tratar al sistema desde el punto de vista de funciones de transferencia racionales en q . La perturbación $v(n)$ obtenida en el modelo no paramétrico mediante el análisis de correlación, que se correspondía a una señal aleatoria, queda ahora transformada y matizada en función de una señal aleatoria de ruido blanco convenientemente modificada por una función de transferencia $H(q)$. Evidentemente, la consecuencia es que la perturbación en la salida no puede considerarse blanca.

La diferencia radical y la esencia del método radica en la introducción en la descripción de un número finito de parámetros (los coeficientes de las funciones racionales) que definen el comportamiento de $G(q)$ y $H(q)$ y en la manera de calcular dichos parámetros. Con estos parámetros introducidos en la descripción del sistema, la ecuación 4.7 queda modificada de la siguiente forma:

$$y(n)=G(q,\theta)x(n) + H(q,\theta)e(n) \quad (4. 37)$$

siendo θ un vector de dimensión d que contiene en número total de coeficientes a determinar en el sistema.

Esta ecuación no es de hecho un modelo de sistema sino una generalización del conjunto de modelos posibles. En la tabla 4.3 se presentan descritos diferentes modelos ajustados a esta ecuación. Por ejemplo, para un modelo denominado autorregresivo exógeno (ARX), cuya estructura es:

$$A(q,\theta)y(n)=B(q,\theta)x(n-nk) + e(n) \quad (4. 38)$$

Las funciones de transferencia quedan definidas así:

$$G(q,\theta)= B(q,\theta)/A(q,\theta) \text{ y } H(q,\theta) = 1/A(q,\theta) \quad (4. 39)$$

Tabla 4. 3 :Diferentes modelos paramétricos

FIR	$y(n) = B(q)x(n-nk) + e(n)$
OE	$y(n) = [B(q)/F(q)]x(n-nk) + e(n)$
ARX	$A(q)y(n) = B(q)x(n-nk) + e(n)$
ARMAX	$A(q)y(n) = B(q)x(n-nk) + C(q)e(n)$
Box-Jenkins	$y(n) = [B(q)/F(q)]x(n-nk) + [C(q)/D(q)]e(n)$
General	$A(q)y(n) = [B(q)/F(q)]x(n-nk) + [C(q)/D(q)]e(n)$

4.9. Estimación paramétrica mediante un modelo ARX

A partir de la ecuación 4.38 se puede obtener una expresión general para predecir $y(n)$ a partir de $x(k)$ con $k \leq n$ e $y(k)$ con $k < n$. Es preciso imponer, sin embargo, que el modelo para el ruido, es decir la función de transferencia $H(q, \theta)$ sea invertible. Siendo esto así, y siendo $H^{-1}(q, \theta)$ el modelo inverso del ruido, la ecuación 4.28 queda modificada según la siguiente expresión:

$$H^{-1}(q, \theta)y(n) = H^{-1}(q, \theta)G(q, \theta)x(n) + e(n) \quad (4.40)$$

A partir de 4.40 se obtiene finalmente:

$$y(n) = (1 - H^{-1}(q, \theta))y(n) + H^{-1}(q, \theta)G(q, \theta)x(n) + e(n) \quad (4.41)$$

Evidentemente, la señal $e(n)$ que representa ruido blanco, no puede ser predicha ya que es absolutamente independiente de lo ocurrido anteriormente. La estimación se obtiene entonces, eliminando $e(n)$:

$$\hat{y}(n|\theta) = (1 - H^{-1}(q, \theta))y(n) + H^{-1}(q, \theta)G(q, \theta)x(n) \quad (4.42)$$

Esta estimación se denomina estimación un paso adelante y es fácil comprobar que, efectivamente, $\hat{y}(n|\theta)$ es función exclusivamente de valores $x(k)$ con $k \leq n$ e $y(k)$ con $k < n$. Aplicando esta expresión a un modelo ARX y sabiendo que $H(q, \theta) = 1/A$ y $G(q, \theta) = B/A$, la estimación toma la siguiente forma

$$\hat{y}(n, \theta) = B(q, \theta)x(n) + [1 - A(q, \theta)]y(n) \quad (4.43)$$

Obsérvese que la ecuación 4.44 se puede reescribir así:

$$\begin{aligned} \hat{y}(n) = & b_1 x(n - n_k) + b_2 x(n - n_k - 1) + \dots + b_{n_b} x(n - n_k - n_b + 1) \\ & - a_1 y(n - 1) - \dots - a_{n_a} y(n - n_a) \end{aligned} \quad (4.44)$$

Si existe un cierto retraso n_k entre la entrada $x(n)$ y la salida $y(n)$, debe quedar reflejado en la estimación de forma que los coeficientes iniciales de B , hasta n_k son nulos. Así, se puede reescribir la ecuación de forma que estos coeficientes no aparezcan, y reescalando la numeración tal como muestra la ecuación anterior.

Los modelos ARX se dice que pertenecen al grupo de modelos con predictores lineales. Ciertamente, la estimación expuesta en esta ecuación puede reescribirse y adaptarse a la expresión:

$$\hat{y}(n|\theta) = \theta^T \varphi(n) \quad (4.45)$$

siendo θ el vector de parámetros y $\varphi(n)$ un vector formado por entradas y salidas retrasadas:

$$\varphi(n) = [-y(n-1) \dots -y(n-n_a) \ x(n-n_k) \dots x(n-n_k-n_b+1)]^T \quad (4.46)$$

Los predictores de regresión lineal son especialmente importantes ya que son estimadores potentes y relativamente simples en su implementación¹. Esta sencillez en el cálculo y en la explicación de las relaciones entre entrada y salida es un factor importante a la hora de valorar las diferentes estructuras aplicadas al tratamiento de los datos del sistema de gases. Pero no es sólo la sencillez el criterio que prima para elegir un estimador ARX para trabajar en el problema que ocupa a este trabajo. Además, en un modelo ARX, la dinámica de la señal de ruido blanco es básicamente la misma que la de las señales de entrada. Es factible utilizar la aproximación ARX cuando puede suponerse que el ruido del sistema entra en él a la vez o bajo las mismas condiciones que la señal de entrada. En el caso del sistema de sensores de gas, este criterio es esencialmente correcto si se supone que la mayor fuente de error proviene de los MFC. Modelos más complejos que el propio ARX han sido también ensayados, pero sus resultados no mejoraron los del modelo ARX. Por este motivo, no son objeto de estudio en este trabajo.

4.10. Criterios para la estimación de los parámetros

Sea entonces una estructura de modelo M tipo ARX, y por tanto, en forma de regresión lineal dada por 4.46, para la cual existen varios modelos diferentes $M(\theta)$. Los distintos modelos difieren en el orden de los polinomios $A(q)$, $B(q)$ y en el factor de retrasos. Una correcta estimación trata de hallar aquel modelo $M(\theta)$ que mejor describa los datos observados en el sistema. Como siempre, la capacidad de descripción y predicción se juzga a través del error de predicción:

$$\varepsilon(n, \theta) = y(n) - \hat{y}(n, \theta) \quad (4.47)$$

Un buen modelo es aquel que predice bien y por tanto produce errores de predicción pequeños. La cuestión reside en definir lo que se entiende por “buen modelo” y en consecuencia, lo que se entiende por “errores de predicción pequeños”.

Habitualmente, y aunque existen formulaciones más generales², el criterio utilizado para cuantificar la ecuación 4.49 es el criterio cuadrático:

$$V_N(\theta) = \frac{1}{N} \sum_{n=1}^N \varepsilon^2(n, \theta) \quad (4.48)$$

Así, para cada vector de parámetros θ la ecuación 4.48 define perfectamente una función de pérdidas $V_N(\theta)$. La cuestión radica en encontrar el vector θ que minimiza la función $V_N(\theta)$:

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta) \quad (4.49)$$

Las maneras de realizar este proceso de minimización son bien conocidas y se suelen englobar, dentro del campo de la identificación de sistemas, bajo el término general de métodos de identificación por error de predicción (PEM)^{2 7 8}. Para el caso particular de criterio cuadrático, aplicado a modelos de regresión lineal, denominado también criterio de mínimos cuadrados, se cumple que:

$$\begin{aligned}
V_N(\theta) &= \frac{1}{N} \sum_{n=1}^N (y(n) - \theta^T \varphi(n))^2 = \frac{1}{N} \sum_{n=1}^N y^2(n) - \frac{1}{N} \sum_{n=1}^N 2\theta^T \varphi(n)y(n) + \\
&\frac{1}{N} \sum_{n=1}^N \theta^T \varphi(n)\varphi^T(n)\theta = \frac{1}{N} \sum_{n=1}^N y^2(n) - 2\theta^T f_N + \theta^T R_N \theta
\end{aligned} \tag{4.50}$$

donde

$$f_N = \frac{1}{N} \sum_{n=1}^N \varphi(n)y(n) \tag{4.51}$$

y

$$R_N = \frac{1}{N} \sum_{n=1}^N \varphi(n)\varphi^T(n) \tag{4.52}$$

f_N es un vector columna d dimensional mientras que R_N es una matriz $d \times d$ dimensional, siendo d la dimensión de θ . Si R_N es invertible entonces el mínimo de $V_N(\theta)$ se produce cuando:

$$\hat{\theta}_N = R_N^{-1} f_N = \left[\frac{1}{N} \sum_{n=1}^N \varphi(n)\varphi^T(n) \right]^{-1} \frac{1}{N} \sum_{n=1}^N \varphi(n)y(n) \tag{4.53}$$

Si el modelo de regresión lineal es una estructura ARX, el vector $\varphi(n)$ contiene, tal y como se ha definido, las funciones $y(k)$ y $x(k)$ siendo $k < n$. Por tanto, la estimación de los parámetros de un modelo de regresión lineal como el propuesto, se realiza a través de funciones de covarianza de las entradas y las salidas ya que la matriz R_N y f_N contiene términos del estilo:

$$\frac{1}{N} \sum_{k=1}^N y(k-j)x(k-i) \quad \frac{1}{N} \sum_{k=1}^N y(k-j)y(k-i) \quad \text{o} \quad \frac{1}{N} \sum_{k=1}^N x(k-j)x(k-i) \tag{4.54}$$

Los parámetros $\hat{\theta}_N$ así calculados minimizan el error de predicción en el sentido de mínimos cuadrados, y se aproximan a los valores reales de θ en ese sentido. Para N tendiendo a infinito, los parámetros $\hat{\theta}_N$ estimados y los reales (suponiendo una estructura adecuada), se hacen iguales ya que $V_N(\theta)$ tiende a cero.

En cualquier caso, aunque la estructura del modelo no fuese la adecuada para el sistema, al menos es seguro, como se ha demostrado, que el modelo estimado es el que minimiza el error de predicción de dicha estructura para los datos del sistema.

Aunque algunas otras estructuras pueden asociarse también a un modelo de regresión lineal, en general la mayoría de estructuras paramétricas son complejas y no es posible entonces hallar una expresión analítica de minimización de una función de pérdidas. El proceso de minimización para estos casos se lleva a cabo mediante optimización no-lineal.

4.11. Determinación del orden del modelo: métodos de regularización

La construcción de un modelo ARX no es obvia ya que el orden del modelo no es conocido a priori, y la determinación de dicho orden no es evidente. Existen varios

métodos que explotan diferentes vías de construcción del modelo. A continuación, se analizan algunas de estas vías.

a) Estrategias de estimación y validación

Una estrategia consiste en generar un modelo a partir de un conjunto de datos de estimación y validar el modelo a partir de otro conjunto de datos diferente o conjunto de validación. Los modelos a ensayar pueden entonces ir creciendo en complejidad, y su grado de ajuste a la señal real se mide en función de una cierta función de pérdidas o error. Lógicamente, el modelo que presenta una función de pérdidas mínima con los datos de validación, es el mejor modelo de los probados (aunque eso no quiere decir que sea el aconsejable, como se discutirá más adelante con el problema sintético). No es fundamental conocer, para la comprensión del funcionamiento de estos modelos, cómo se calcula esta función de pérdidas, pero hay que destacar que, fruto de realizar el estudio con una zona de validación, llega un momento, en el que una mayor complejidad del modelo no supone una menor función de pérdidas sino al contrario. Cuando el modelo ya ha explotado al límite las posibilidades de la estimación, añadir más complejidad sólo sirve para intentar modelar el comportamiento estadístico de la señal de error que subyace en los datos de estimación. Puesto que éste obedece a un comportamiento puramente aleatorio y que no se corresponde con el que existe en la zona de validación, la complejidad añadida al modelo redundará entonces en una pérdida de rendimiento del mismo. La zona de validación, es por tanto la referencia que marca el modelo óptimo, ya en la zona de estimación se cumple que a mayor complejidad del modelo, menor función de pérdidas.

b) estrategia con una sola zona de datos

Directamente relacionado con buscar la sencillez de los modelos, están las técnicas de penalización por número de parámetros que se aplican sobretodo con las estrategias de identificación paramétrica donde no es posible disponer de un conjunto de validación debido a un problema de escasez de datos. Hasta ahora se ha desarrollado la estrategia basada en utilizar una serie de datos para estimar el modelo y otra serie, diferente, para validarlo. Sin embargo, si no es posible disponer de datos para validar el modelo, la estrategia de identificación ha de basarse en estimar el modelo y validarlo sobre el mismo conjunto de datos. El problema que conlleva adoptar esta estrategia es, como ya se ha dicho, que los modelos siempre mejoran a medida que la estructura del modelo crece. Para compensar este efecto existen algunos criterios objetivos o métodos de regularización que permiten penalizar a los modelos que contengan estructuras complejas. Los métodos de regularización más conocidos son la técnica de error de predicción final de Akaike (FPE), el criterio de información teórica (AIC) y el criterio de Rissanen (MDL) cuyas descripciones se presentan a continuación^{10 11 12}:

$$\begin{aligned} \text{FPE} &= \frac{1 + n / N}{1 - n / N} \cdot V \\ \text{AIC} &\approx \log[(1 + 2n / N) \cdot V] \\ \text{MDL} &= (1 + c \cdot n / N) \cdot V \end{aligned} \quad (4.55)$$

donde n el número de parámetros de la estructura, N el número de puntos empleados en la zona de estimación, V la función de pérdidas y c un número dado, que por defecto está relacionado con el logaritmo neperiano de N .

Tanto una como otra estrategia (dos zonas de datos, para estimar y validar, o una sola zona de datos) pueden ser perfectamente válidas, pero en principio parece más apropiado, si es posible, utilizar la primera, generando conjuntos de datos de estimación y de validación. Por otro lado, la aplicación de los diferentes criterios da lugar a diferentes modelos con diferentes órdenes. Por lo tanto escoger entre un criterio u otro conlleva una cierta arbitrariedad no deseable. En definitiva, sólo en el caso de no disponer de suficientes datos se adoptaría la segunda estrategia. Este será el criterio que regirá para los ejemplos experimentales que se presentan en este trabajo.

4.12. Aplicación de la estimación paramétrica a un problema sintético

Antes de aplicar la teoría de la estimación paramétrica a los datos del sistema sensor, tal como se hizo con la estimación no paramétrica, se va a estudiar un problema sintético. De esta manera se podrán ilustrar algunos conceptos teóricos de las estrategias de identificación. El sistema a tratar se rige por la misma ecuación 4.29 que describía al sistema SISO ya utilizado para tratar la estimación no paramétrica.

La estimación del modelo paramétrico se va a realizar de dos formas diferentes. La primera tratará de generar un modelo de respuesta impulsiva finita (FIR). La segunda aproximación consistirá en utilizar la estructura ARX con toda su potencia.

4.12.1. Modelo FIR

Es fácil de comprobar que el modelo calculado a través del análisis de correlación es un modelo FIR que se puede asimilar a la estructura ARX sin más que definir el polinomio $A(q)$ como $A(q) = 1$. De ese modo, el modelo FIR queda definido como:

$$y(n) = B(q)x(n-nk) + e(n) \quad (4.56)$$

donde $B(q)$ puede entonces equipararse a una respuesta impulsiva finita.

A partir de esta equivalencia matemática, el paso lógico siguiente es comprobar el rendimiento de esta estructura en comparación con los modelos de análisis de correlación. La figura 4.16 muestra la respuesta impulsiva y el polinomio $B(q)$ para el sistema ejemplo 4.29

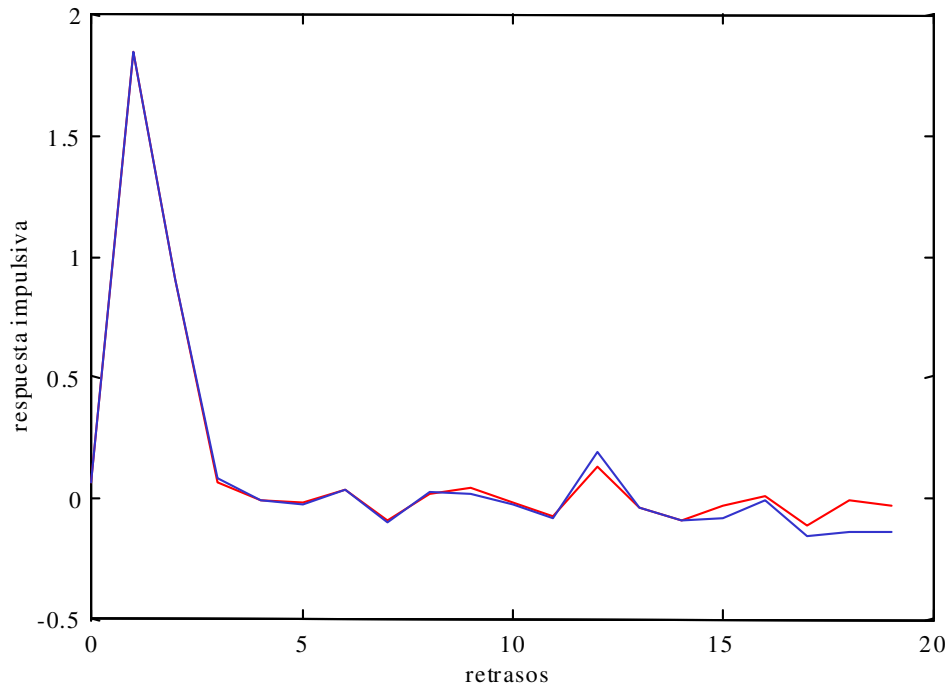


Figura 4.16 Respuestas impulsivas: método de correlación y cálculo FIR

En la figura se muestran las respuestas impulsivas calculadas según el método de correlación (línea roja) y según el sistema paramétrico FIR (línea azul). Del resultado de dicha comparación se deduce que, salvo pequeñas diferencias, los valores del polinomio $B(q)$ se corresponden con la respuesta impulsiva calculada a través del análisis de correlación.

El rendimiento del modelo es, por tanto, en este caso, similar al obtenido en el modelo de correlación.

4.12.2. Modelo ARX

Sin restringirse a la estructura FIR, se han probado diferentes modelos ARX ajustados al problema sintético definido por la ecuación 4.29, según las estrategias presentadas en el punto 4.11.

Siguiendo la estrategia de estimación y validación, la figura 4.17 presenta la función de pérdidas para una serie de modelos con un polinomio $A(q)$ fijo, un polinomio $B(q)$ creciente en complejidad y un número de retrasos fijo. Las dimensiones de los polinomios $A(q)$, $B(q)$ y la dimensión de los retrasos se definen como n_a , n_b y n_k respectivamente. El ensayo se ha confeccionado con $n_a=2, n_b=1, 2, \dots, 20$, y $n_k=1$, (aunque estos valores no tienen porque ser los óptimos).

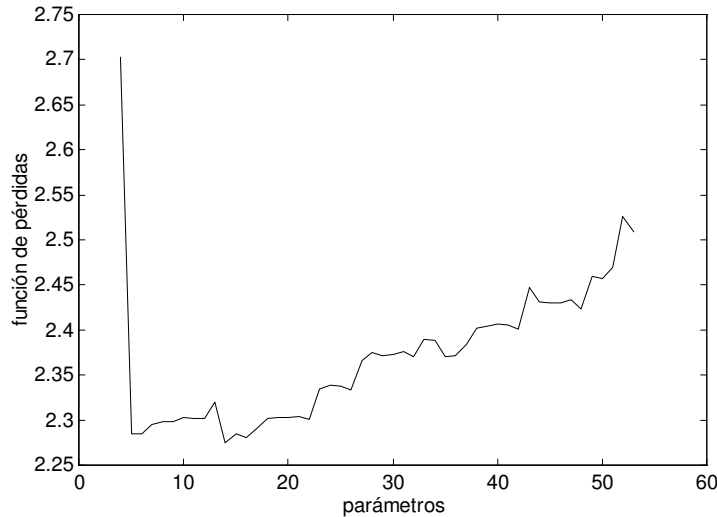


Figura 4.17 Evolución de la función de pérdidas en función del número de parámetros

Se observa como efectivamente hay una zona donde la función permanece en un cierto nivel pero cuando el número de parámetros sube por encima de, aproximadamente 20, la función de pérdidas comienza a crecer.

Queda aun una cuestión interesante a discutir, dentro de la estrategia estimación-validación. En el caso concreto presentado, el mínimo está situado en $n_a=2, n_b=11, n_k=1$. Sin embargo, la opción con $n_b=2$ o con $n_b=3$ no presenta una función de pérdidas significativamente mayor que $n_b=11$ y aportan la ventaja de ser modelos menos complejos. Quiere decirse con esto que no siempre el modelo que hace mínima la función de pérdidas es el óptimo en la práctica, y que, en general, es preferible sacrificar algo de precisión a cambio de sencillez. (aunque valorar estos aspectos siempre conlleva un cierto factor de subjetividad).

Puesto que no se considera la estrategia basada en métodos de regularización, los resultados que se obtienen para este sistema teórico a partir de la estrategia de estimación-validación, se presentan en la tabla 4.4:

Tabla 4.4 Resultados de la figura PSNE para el problema sintético		
nº parámetros	estructura	PSNE (%) (zona de validación)
3	2 1 1	47.1 ± 3.7
4	2 2 1	42.4 ± 3.3
5	2.3 1	42.6 ± 3.4
6	2 4 1	42.6 ± 3.4
10	2 8 1	42.2 ± 3.3
15	2 13 1	42.7 ± 3.4
20	2 18 1	43.9 ± 3.5
30	2 28 1	44.1 ± 3.5
40	2 38 1	46.6 ± 3.7

Queda clara la tendencia alcista del valor de PSNE a partir de una cierta cantidad de parámetros, tal y como mostraba la función de error. El valor de PSNE está sometido a una cierta variabilidad que depende de la realización concreta de la señal de error. Pero, parece claro que este valor es similar al hallado para el desarrollo FIR y para el análisis de correlación.

Así pues, se ha comprobado a través del ejemplo que el desarrollo FIR es similar en esencia al análisis de correlación. En este último caso, los coeficientes de la respuesta impulsiva se calculan a través de relaciones de correlación entre las señales del sistema mientras que en la estimación paramétrica de estructuras FIR o ARX se obtiene los coeficientes de la respuesta impulsiva a través de cálculos de mínimos cuadrados. En cuanto al rendimiento del método paramétrico en general, no puede concluirse que sea mejor que el análisis de correlación es su aplicación a sistemas SISO, pero falta comprobar el resultado aplicado a sistemas MISO.

Extraídas las conclusiones sobre el problema sintético, se presentan a continuación los resultados aplicados al sistema real de análisis de gases.

4.13. Aplicación de la estimación paramétrica al sistema

4.13.1 Identificación SISO: Estimación con estructura FIR

Igual que en el proceso de estimación por análisis de correlación, se inicia la identificación a partir de un sistema SISO.

Ajuste y validación

Aunque ya se ha comprobado la similitud entre los resultados del análisis de correlación y el modelo paramétrico con estructura FIR, se debe realizar el estudio aplicado al sistema de sensores de gas. Se presenta a continuación una figura comparando la respuesta impulsiva calculada mediante análisis de correlación para el sensor 6 sometido a concentraciones de gas tolueno con su equivalente modelo paramétrico con estructura FIR. Se omiten en este caso las descripciones del experimento y la presentación de los resultados ya que son estrictamente iguales a los descritos para el caso SISO en el análisis de correlación.

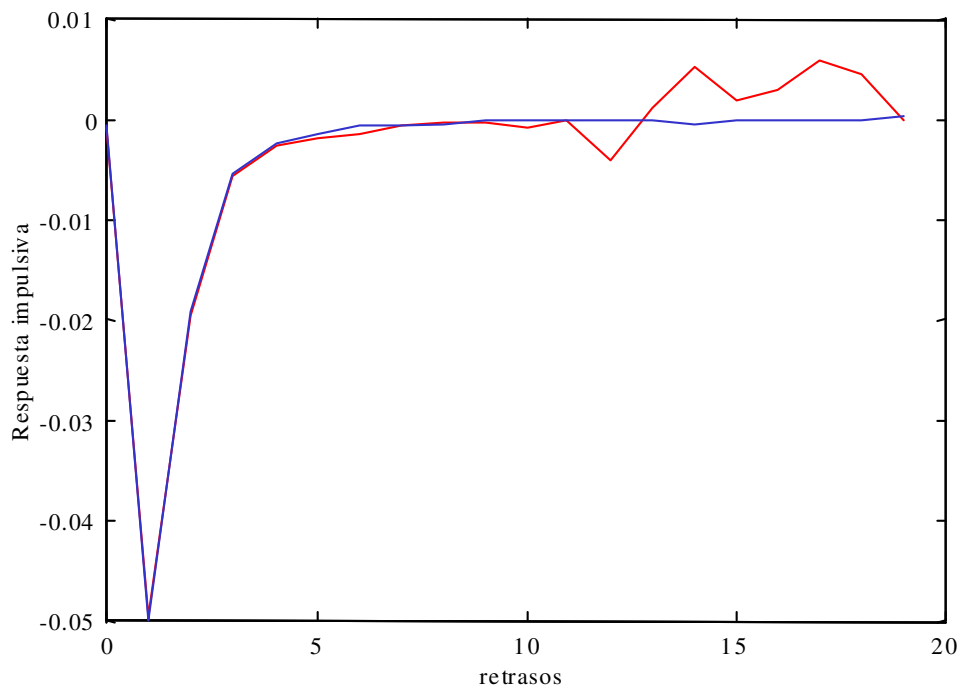


Figura 4.18 Comparación de respuestas impulsivas

En línea roja se muestra la respuesta impulsiva calculada mediante el método no paramétrico, mientras que en línea azul se presentan los valores de $B(q)$ calculados a través del modelo paramétrico de estructura FIR. Las diferencias en los valores fundamentales son mínimas y, las mayores diferencias se centran en datos irrelevantes y que quedan dentro de los márgenes de confianza del cálculo. Así ocurre también con el resto de sensores sometidos tanto a octano como tolueno. Cabe destacar sin embargo, la discrepancia entre ambas respuestas que aparece hacia retrasos de orden 12,13,14. Si bien en esos retrasos la respuesta impulsiva estimada por el método de correlación se mantiene por debajo del margen de significado, están muy cerca del límite que marca la confianza. Puede llegar a dar la impresión que existe una cierta relación entre la entrada y la salida en esa zona. Sin embargo una respuesta impulsiva así supone una relación de entrada-salida que como mínimo puede calificarse de extraña. No parece lógico que la memoria del sistema respecto de la señal de entrada aparezca y desaparezca de esta forma. Ciertamente esto no puede ser, y la respuesta impulsiva calculada a través de mínimos cuadrados deshace esta duda, presentando una evolución en la que no se observa esa discrepancia. En ese sentido parece más fiable la estimación mediante mínimos cuadrados que mediante la correlación, aunque el rendimiento sea similar.

En definitiva, las estimaciones SISO son, como no podía ser de otro modo y salvando las pequeñas discrepancias, esencialmente iguales a las ya presentadas mediante el análisis de correlación.

4.13.2. Identificación SISO: Modelo paramétrico con estructura ARX

Ajuste y validación.

El hecho de escoger una estructura ARX frente a otras estructuras paramétricas para ajustar un modelo a los datos experimentales se debe, como ya se ha comentado, a que es previsible que la componente de error va a ser pequeña comparada con la señal y a que además puede suponerse que está sometida a una dinámica similar a la de las entradas de concentración. Por otro lado, la estructura ARX es un modelo lineal en los parámetros y permite la identificación mediante mínimos cuadrados en un solo paso.

Si se revisa el modelo de correlación para el sensor 1 se comprueba que la respuesta impulsiva cuando el sensor está sometido a gas tolueno comienza a tener valores significativos a partir del primer minuto, mientras que para el gas octano, este hecho ocurre a los dos minutos. Siendo esto así, los valores de n_k para tolueno y octano han de ser 1 y 2 respectivamente. Efectivamente, si se comprueban diferentes modelos, computando cual ofrece un error mínimo cuando se fijan n_a y n_b , manteniendo variable n_k entonces los modelos que presentan error mínimo coinciden con los datos anteriores.

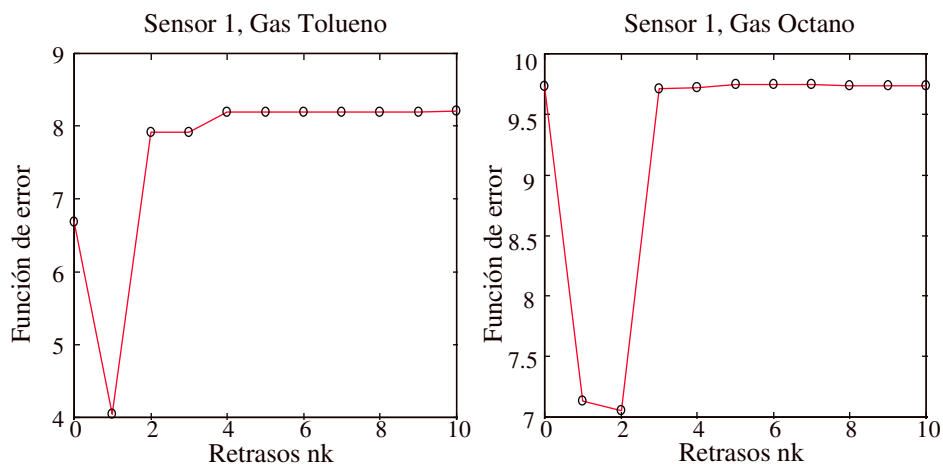
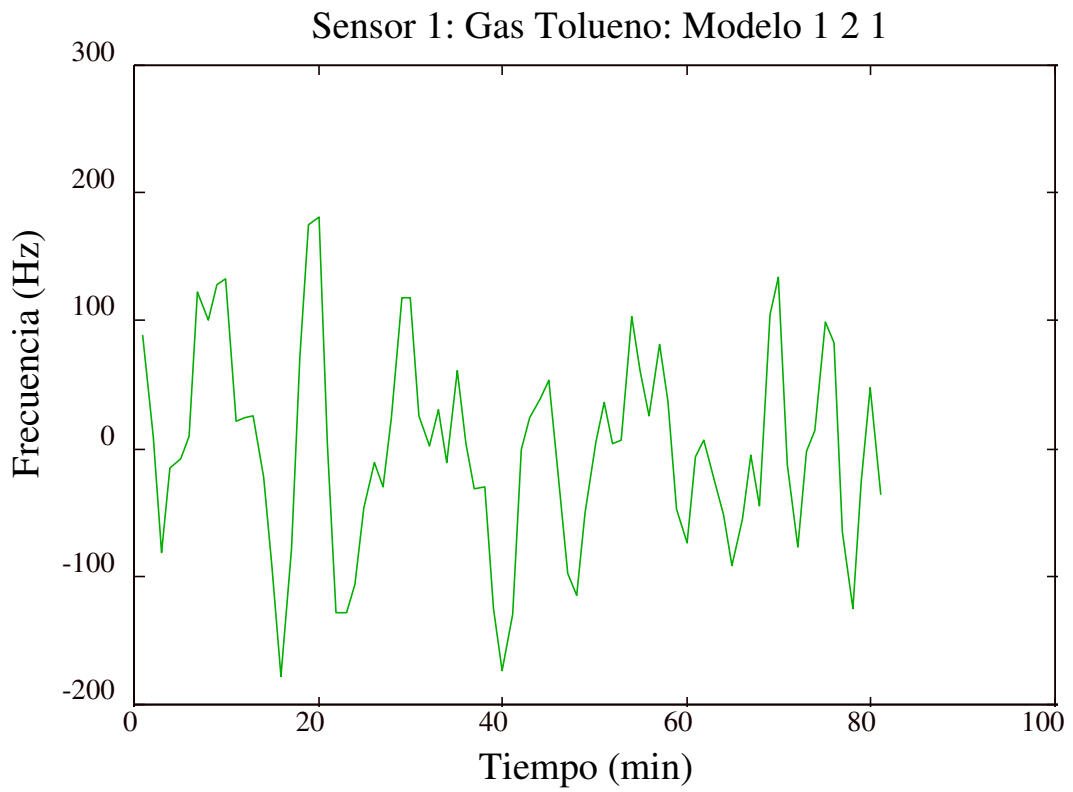
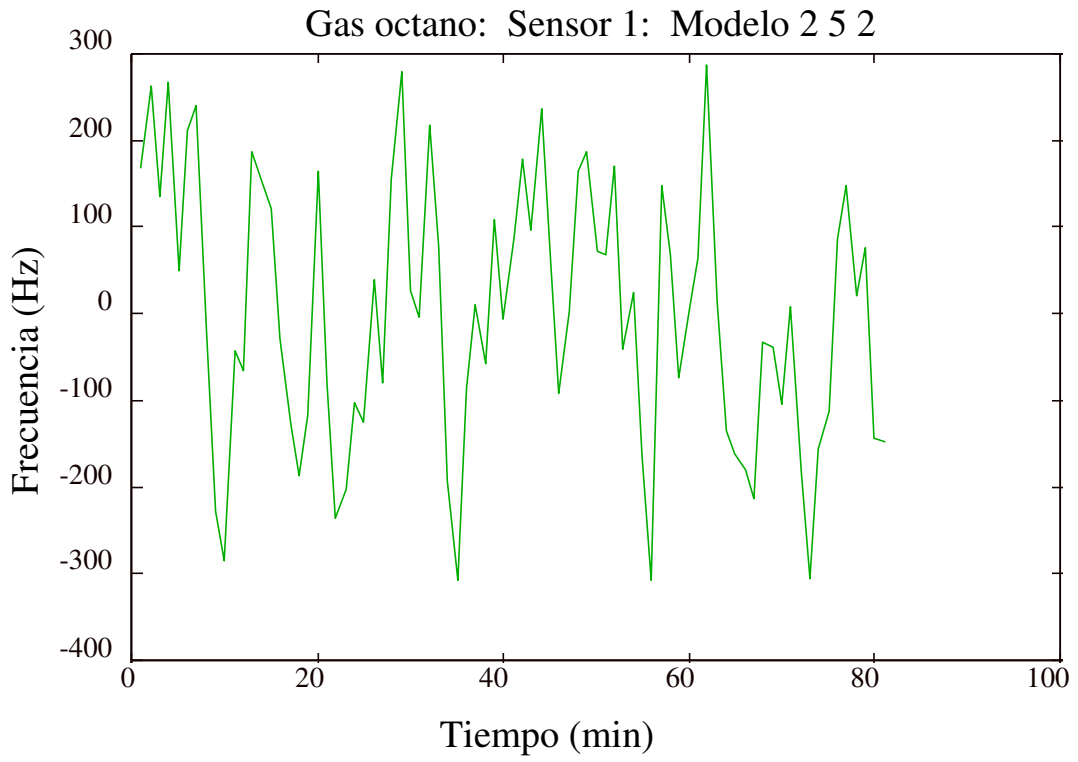


Figura 4.19 Función de error para n_a y n_b constantes, n_k variable

Fijados estos datos se pueden probar diferentes valores de n_a y n_b . En la tabla 4.5 se presentan diferentes estructuras, de tamaño creciente para el sensor 1 y para el experimento 2 con gas octano y tolueno por separado.

Tabla 4.5 Estructuras del modelo ARX para el sensor 1 y gas octano			
Estructura n_a n_b n_k	n° de parámetros	Estimación %	Validación %
1 1 2	2	4.73 ± 0.2	8.13 ± 0.6
1 2 2	3	4.48 ± 0.2	7.22 ± 0.6
1 3 2	4	4.53 ± 0.2	7.14 ± 0.6
1 4 2	5	4.36 ± 0.2	7.11 ± 0.6
2 5 2	7	4.25 ± 0.2	7.04 ± 0.5
5 5 2	10	4.24 ± 0.2	6.94 ± 0.5
8 7 2	15	4.16 ± 0.2	7.05 ± 0.5
4 16 2	20	4.04 ± 0.2	6.81 ± 0.5
7 23 2	30	4.06 ± 0.2	6.56 ± 0.5



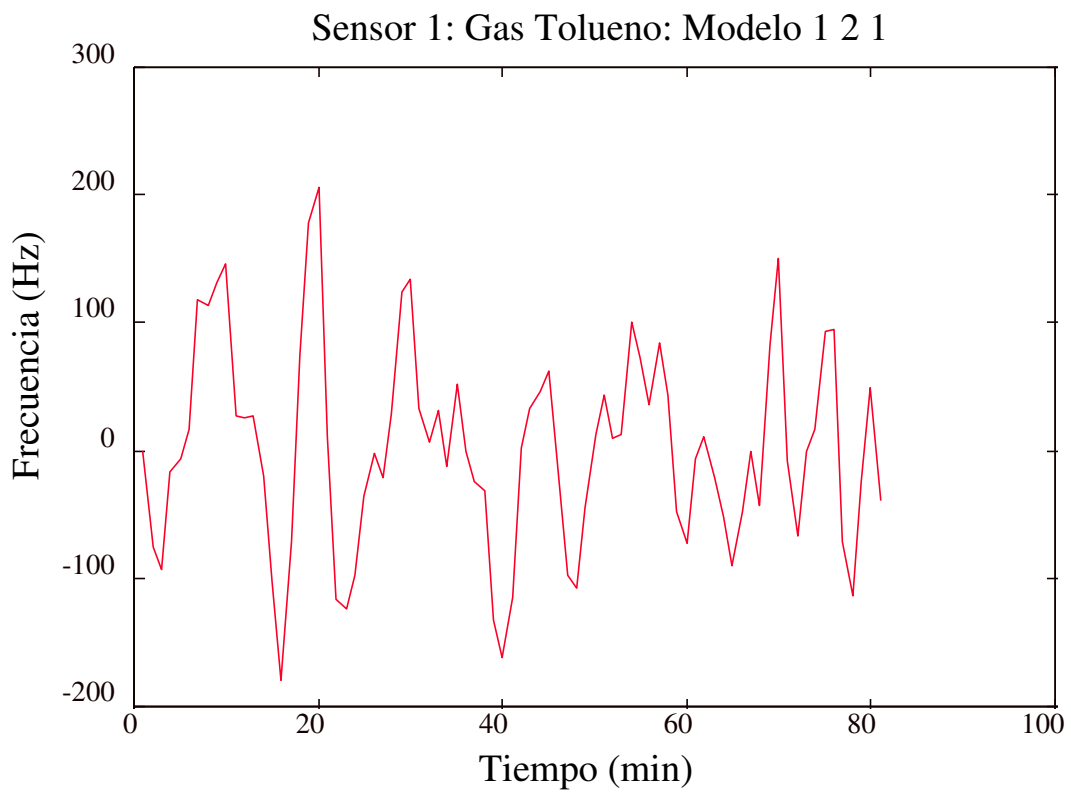
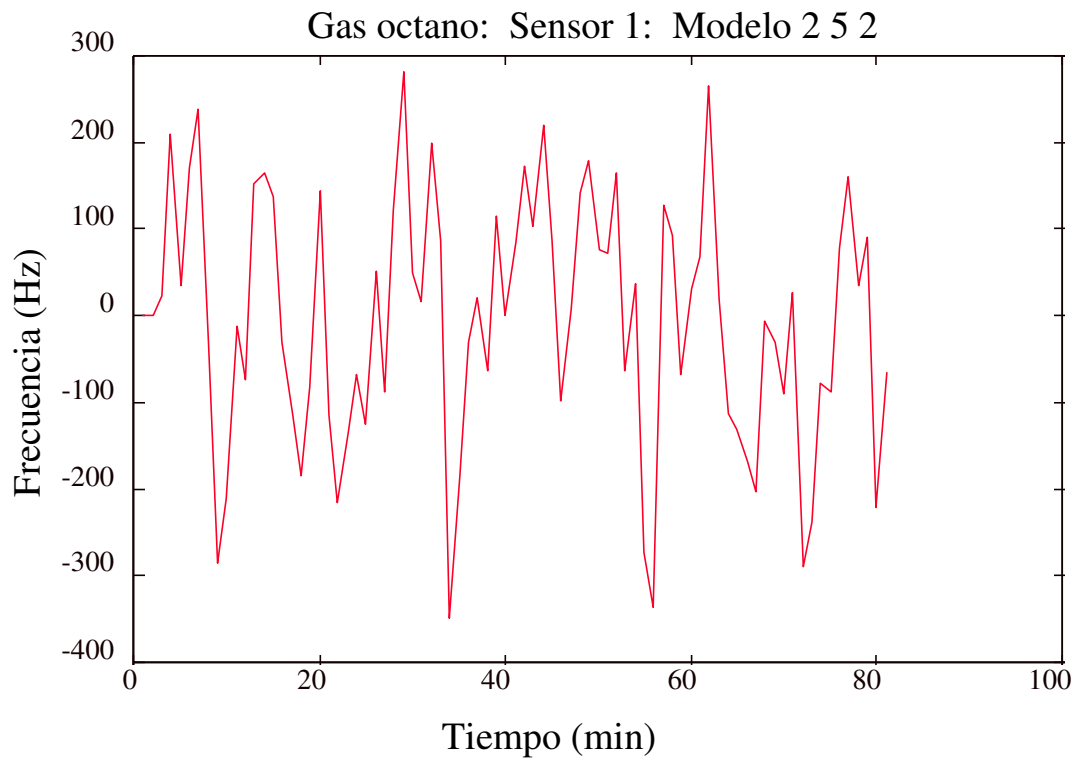


Figura 4. 20 Resultados para el sensor 1 de los modelos ARX óptimos para octano y tolueno

Estructura n_a n_b n_k	n° de parámetros	estimación %	validación %
1 1 1	2	10.5 ± 0.4	9.78 ± 0.7
1 2 1	3	1.22 ± 0.05	0.69 ± 0.05
2 2 1	4	1.16 ± 0.05	0.70 ± 0.06
3 2 1	5	1.12 ± 0.04	0.72 ± 0.06
2 5 1	7	1.10 ± 0.04	0.67 ± 0.05
4 6 1	10	1.12 ± 0.04	0.66 ± 0.05
7 8 1	15	1.11 ± 0.04	0.68 ± 0.05
10 10 1	20	1.13 ± 0.05	0.69 ± 0.05
15 15 1	30	1.12 ± 0.04	0.74 ± 0.06

La figura 4.21 muestra una comparación para el gas tolueno y el sensor 1 entre los dos tipos de estimación realizadas hasta el momento. Puede observarse en ella y comparando las tablas presentadas que el modelo paramétrico ofrece un mejor rendimiento que el no paramétrico.

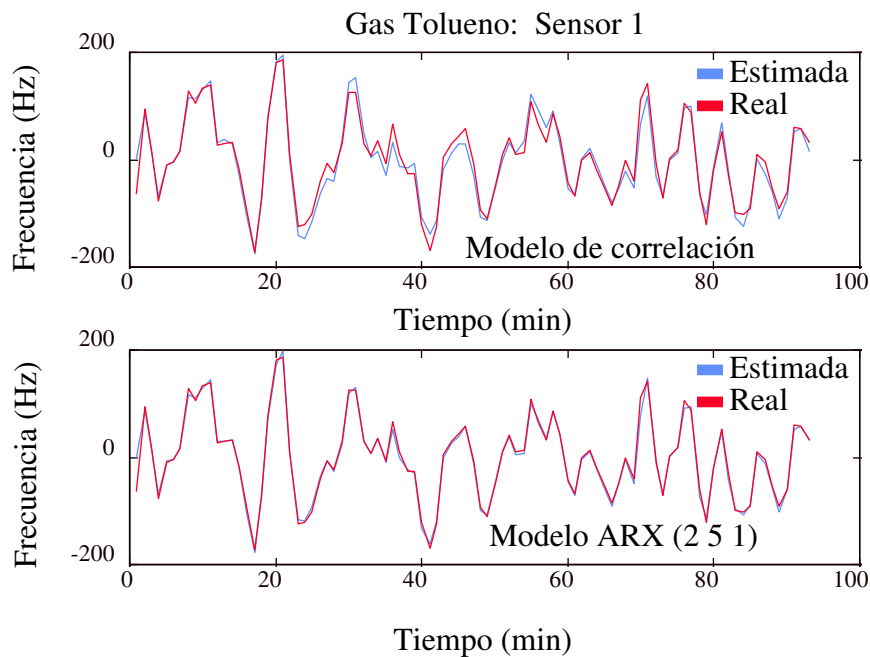


Figura 4.21 Comparación entre estimaciones

Aceptación de los modelos

De todas las estructuras probadas, se deben escoger como óptimas las que ofrecen un error pequeño en la zona de validación sin que la estructura sea excesivamente compleja. El grado de “complejidad” de una estructura siempre es un criterio subjetivo. Por ejemplo, de la tabla para el sensor 1 y gas octano, la estructura 5 5 2 ofrece el error en zona de validación mínimo, sin embargo la estructura 2 5 2, con tres parámetros menos ofrece un error similar; y aun la estructura 1 2 2 podría considerarse útil. Para no introducir un criterio subjetivo, se ha realizado una selección de estructura que, aun y utilizando dos conjuntos de datos, uno para estimar y otro para validar, penalice a las

estructuras excesivamente grandes mediante el criterio de Rissanen, que es el más restrictivo para la cantidad de datos con la que se trabaja. La tabla 4.7 muestra las mejores estructuras con sus respectivos resultados según el criterio de Rissanen:

Tabla 4.7 Estructuras ARX mínimas según el criterio de Rissanen						
	Gas Tolueno			Gas Octano		
	Estructura	PSNE (%) (zona est.)	PSNE (%) (zona val.)	Estructura	PSNE (%) (zona est.)	PSNE (%) (zona val.)
Sensor 1	1 2 1	1.22 ± 0.05	0.69 ± 0.05	1 4 2	4.36 ± 0.2	7.11 ± 0.6
Sensor 2	2 5 1	1.40 ± 0.05	0.42 ± 0.03	3 3 2	5.66 ± 0.2	6.90 ± 0.5
Sensor 5	6 5 1	0.55 ± 0.02	0.32 ± 0.02	2 5 2	18.10 ± 0.7	9.90 ± 0.8
Sensor 6	1 1 1	0.43 ± 0.02	0.24 ± 0.02	2 5 2	19.85 ± 0.8	10.6 ± 0.8

Probablemente, aplicando un criterio más estricto, podrían encontrarse modelos más sencillos (por ejemplo, para el gas tolueno y el sensor 2, el modelo 3 1 1 ofrece prestaciones similares con un parámetro menos). No obstante, la aplicación de dicho criterio es una cuestión sometida a subjetividad. Es por ello que se ha decidido finalmente, considerar los valores de la tabla como definitorios del sistema y definitivos.

En este sentido, y comentando los resultados, cabe destacar que para el octano, los valores obtenidos son similares a los del sistema SISO estimado por correlación, pero los valores para tolueno son significativamente mejores que los ofrecidos por aquella estimación. Ya había quedado claro que los sensores presentaban mayor grado de no linealidad para el octano que para el tolueno. Siendo esto así, es lógico que los ajustes mediante un modelo ARX lineal sean mucho más precisos para el tolueno que para el octano, y explican que, de alguna manera, una estimación por mínimos cuadrados rinda mejor que una estimación por correlación. De hecho, mientras que para el octano es justificable intentar atacar el problema desde modelos no lineales, para el tolueno, no supone ninguna mejora puesto que ya estamos en el límite de la precisión del sistema SISO.

En cuanto al análisis de residuos, los resultados no difieren a los obtenidos con el análisis no paramétrico. La figura 4.22 muestra un test lineal y uno no lineal para el gas octano y el sensor 1. De la figura y del resto de resultados del análisis de residuos se puede concluir que para el gas octano, la parte lineal está bien estimada, pero que queda información no lineal por desentrañar.

Sensor 1 Gas Octano

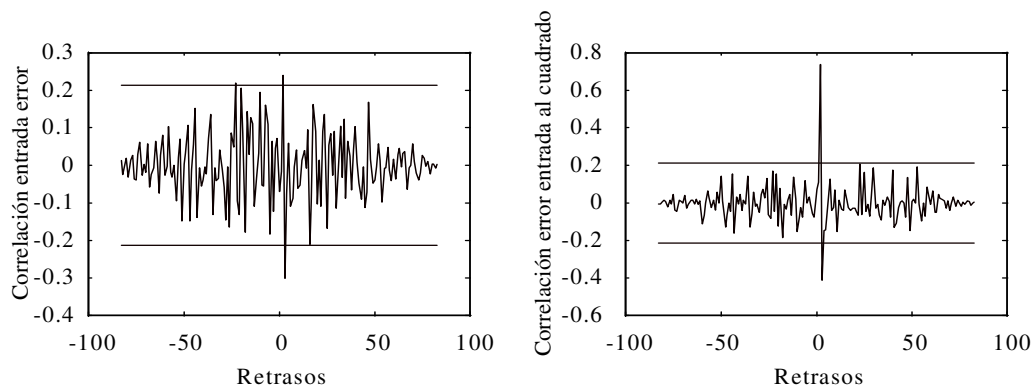


Figura 4.22 Análisis de residuos test lineal y test no lineal para el sensor 1 y gas octano

Para el gas tolueno, en ningún caso los tests no lineales ofrecen indicios que permitan justificar el ataque no lineal. Esto es lógico si se tiene en cuenta que el modelo paramétrico ha explotado prácticamente hasta el límite la información del sistema.

Como conclusión general para el sistema SISO estimado con métodos paramétricos, se puede decir que este método resulta más fiable en la estimación de ciertos aspectos de los modelos y que su rendimiento es similar o mejor al análisis de correlación. Además, para el gas tolueno dentro del marco de los sistemas SISO, se ha llegado al límite de explotación de los datos, mientras que para el octano todavía hay margen a explorar con métodos no lineales.

4.13.3. Estimación MISO: Estructura ARX

De los desarrollos generados con el modelo FIR y ARX parece claro que aquella que presenta un valor añadido al análisis de correlación es el desarrollo desde el punto de vista general. Las estructuras FIR, de hecho, no aportan una sustancial diferencia al método de correlación. Por tanto, para el estudio desde el punto de vista MISO, se trabajará ya exclusivamente con desarrollos generales.

Ajuste y validación

Para ver cómo se comportan estos métodos cuando el sistema es sometido a mezclas de ambos gases, se utiliza el experimento 3, cuyos resultados ya se han presentado anteriormente. El ajuste ARX para un problema MISO se ha de regir por una ecuación como la que sigue:

$$A(q)y(n) = q^{-nk_1} \cdot B_1(q)x_1(n) + q^{-nk_2} \cdot B_2(q)x_2(n) + e(n) \quad (4.57)$$

Donde $A(q)$ representa un polinomio que contiene la dinámica del sistema, incluida en ella, la propia dinámica de $e(n)$, y que es común a ambas entradas. Por otra parte, $B_1(q)$ y $B_2(q)$ contienen información relativa a la memoria del sistema, que en general, es diferente para cada entrada. Por último, los retrasos de la salida respecto de las entradas también pueden ser diferentes, y vienen representados en la ecuación por el factor q^{-nk_i} .

Una primera aproximación al problema MISO consiste en comprobar el rendimiento de la modelización ARX cuando se utilizan las mejores estructuras determinadas en el caso SISO. Así, para el sensor 1, según indica la tabla 4.7, deben compaginarse las estructuras (1 2 1) para el gas tolueno y (1 4 2) para el gas octano. De tal manera que la ecuación a estimar, descompuesta en sus parámetros es:

$$a_0y(n) + a_1y(n-1) = b_{11}x_1(n-1) + b_{12}x_1(n-2) + b_{21}x_2(n-2) + b_{22}x_2(n-3) + b_{23}x_2(n-4) + b_{24}x_2(n-5) \quad (4.58)$$

Deben determinarse por tanto, 7 parámetros diferentes.

Tabla 4.8 Estructura ARX MISO a partir de estimaciones ARX SISO para el sensor 1	
Parámetros estimados	$a_0 = 1$ $a_1 = -0.0898$ $b_{11} = -0.0207$ $b_{12} = -0.0154$ $b_{21} = -0.0440$ $b_{22} = -0.0108$ $b_{23} = -0.0044$ $b_{24} = -0.0034$
PSNE (%) zona de estimación	14.7 ± 0.6
PSNE (%) zona de validación	17.2 ± 1.4

Como las mejores estructuras en el caso del sensor 1 contenían ambas un valor $n_a = 1$, no ha habido problemas a la hora de generar la estructura MISO. Sin embargo, para el resto de sensores los valores de n_a de las estructuras SISO son diferentes para cada gas. Pero para el modelo MISO, debe cumplirse que el polinomio $A(q)$ sea común a ambas entradas, por tanto, alguna estructura debe sacrificar parte de su descripción óptima del tipo SISO.

Para el sensor 2, se han probado dos casos, utilizar $n_a=2$, con lo cual el modelo SISO para octano pierde un parámetro en su adecuación al sistema MISO o el caso $n_a=3$, donde el modelo SISO para tolueno, queda sobredimensionado.

Los resultados del PSNE en tanto por ciento se presentan a continuación:

	Estructura 2 [5 3] [1 2] ($n_a=2$)	Estructura 3 [5 3] [1 2] ($n_a=3$)
PSNE(%) zona estimación	10.5 ± 0.4	11.3 ± 0.5
PSNE(%) zona validación	10.9 ± 0.9	12.3 ± 0.9

En el siguiente apartado de aceptación de resultados, se discutirá sobre estos valores y los presentados anteriormente. Puede sin embargo, adelantarse que estos resultados no mejoran los rendimientos hallados con el análisis de correlación aplicado al problema MISO. Cabe preguntarse, no obstante si las estructuras óptimas para el modelo SISO siguen siendo válidas para el MISO. Para investigar sobre la cuestión, lo mejor es probar otras estructuras sin la restricción de utilizar las halladas en el caso SISO. Fruto de estas experiencias se ha extraído finalmente una estructura común para los sensores que presenta un rendimiento que se presenta a continuación.

La tabla 4.10 muestra para cada sensor, el resultado para la estructura 3,3,1 asociado al tolueno y 3,4,1 asociado al octano:

método ARX	Sensor 1	Sensor 2	Sensor 5	Sensor 6
estimación %	5.0 ± 0.2	3.7 ± 0.1	5.3 ± 0.2	3.9 ± 0.1
validación %	7.0 ± 0.5	5.5 ± 0.4	9.2 ± 0.7	7.3 ± 0.6

Por otra parte, es interesante en este momento comprobar la calidad blanca del ruido. En la siguiente figura se muestra la magnitud espectral del ruido, pudiéndose comprobar que la caída en unidades de decibelios es de 1 decibelio aproximadamente.

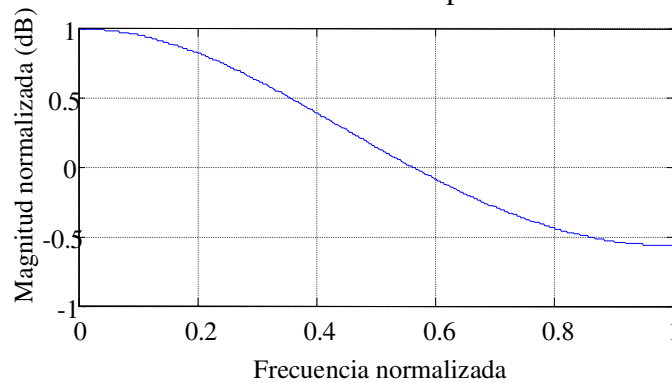
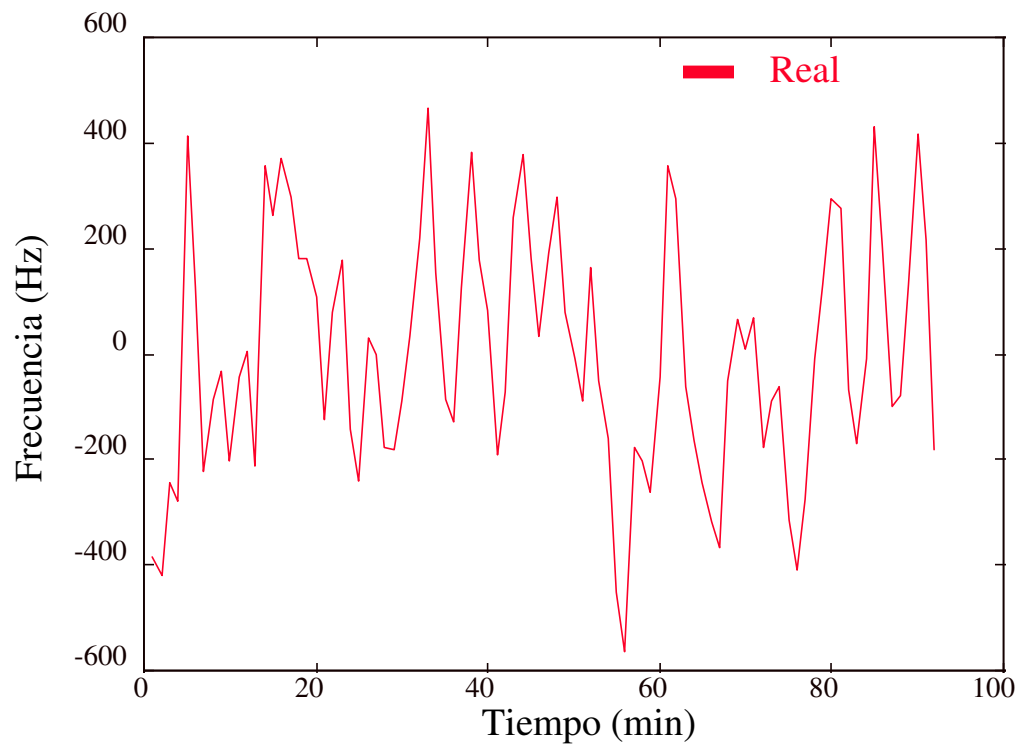


Figura 4.23 Espectro del ruido en el sistema

En la figura 4.24 se muestra una comparación entre los resultados reales y los del modelo de estructura ARX.



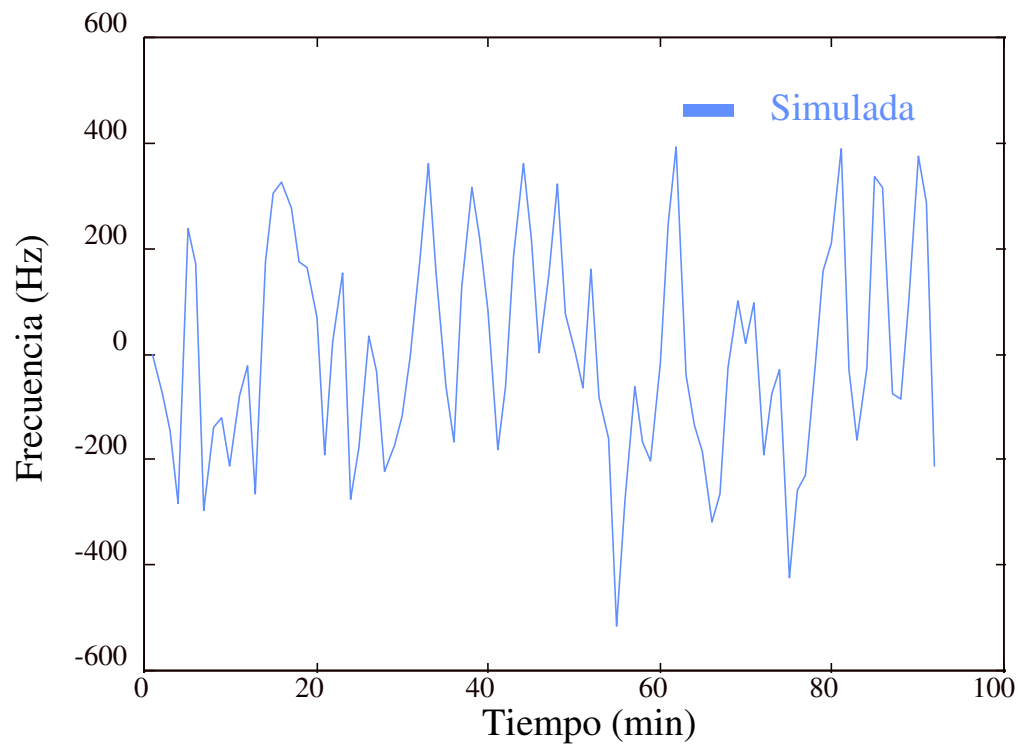


Figura 4. 24 Comparación entre los datos reales y el modelo ARX

Aceptación del modelo

Al realizar los ensayos MISO con las mejores estructuras SISO, cabía esperar una franca mejoría sobre los resultados del análisis de correlación. Ciertamente, la mejora ha sido importante, reduciéndose el valor de la función PSNE de un 30 % a un 10% aproximadamente, para estructuras MISO procedentes de la unión de las mejores estructuras SISO. Respecto al estudio realizado sobre las estructuras cuando en el caso SISO los valores del parámetro n_a eran diferentes, los resultados no son concluyentes y no permiten decidir claramente entre uno u otro valor de n_a . Cabe comentar que el buen rendimiento de los modelos SISO aplicados a tolueno se pierde en presencia de octano cuando se ataca el problema MISO. Esto es lógico, y ya se pudo predecir a la vista de los resultados estáticos presentados en el capítulo anterior.

Existen estructuras MISO que mejoran aun más el rendimiento respecto a los presentados a partir de estructuras SISO óptimas. Con estas nuevas estructuras se consiguen errores del PSNE alrededor del 5%. Curiosamente, el factor determinante en estas estructuras es el correspondiente al retraso del octano. Bajando este valor de 2 a 1, los modelos ARX MISO rinden mejor. En la figura 4.19 se puede comprobar que introducir un retraso 2 o 1 en el octano no genera grandes diferencias en la función de error.

No se presentan en este caso los resultados del análisis de los residuos ya que aportan la misma información que en el caso del modelo FIR.

Queda claro a la vista de todo lo expuesto que las mejores estructuras SISO pueden ser directamente exportables, en este problema, hacia modelos más complejos, pero existen estructuras MISO mejores. Por otra parte, a pesar de la mejora, existe aun un porcentaje de señal que los modelos lineales no son capaces de explicar, y que por tanto, vale la pena intentar atacar el problema desde perspectivas no lineales.

4.14. Elección entre método paramétrico y no paramétrico

A raíz de los resultados de este estudio y antes de continuar con la modelización no lineal, vale la pena discutir brevemente, a modo de conclusión, los beneficios y los inconvenientes que aporta la elección entre un método como el análisis de correlación o la estimación mediante un modelo paramétrico. Ha de quedar claro que el objetivo de los estimadores es hallar una función $\hat{y}(n)$ que prediga los valores de la señal $y(n)$ futuros del sistema a partir de las señales del propio sistema para tiempos anteriores a n . La filosofía, no obstante, de los estimadores no-paramétricos como el basado en análisis de correlación y los paramétricos es radicalmente diferente.

Sea $\hat{y}(n)$ la predicción realizada de un sistema SISO LTI mediante un método de análisis de correlación y sea $\varepsilon(n)$ el error de la predicción, definido como:

$$\varepsilon(n) = y(n) - \hat{y}(n) \quad (4.59)$$

El estimador será tanto mejor cuanto menor sea $\varepsilon(n)$ para cualquier n . Sin embargo a efectos prácticos, en un análisis de correlación y para un número finito N de datos, el error de predicción viene marcado, no tan solo por la señal de perturbación $v(n)$ sino por otro término que tiene en cuenta el error cometido en la estimación por usar un número finito de datos.

$$\begin{aligned}\varepsilon(n) &= y(n) - \hat{y}(n) = \sum_{k=0}^{\infty} g(k)x(n-k) + v(n) - \sum_{k=0}^{\infty} \frac{R_{yx}(k)}{R_{xx}(0)} x(n-k) \\ &= v(n) + \sum_{k=0}^{\infty} \left[g(k) - \frac{R_{yx}(k)}{R_{xx}(0)} \right] x(n-k)\end{aligned}\quad (4.60)$$

Puede demostrarse ⁸ que la varianza del segundo sumando de la expresión final del error de predicción decae como $1/N$, y que por tanto el error de predicción para el análisis de correlación tiende a la señal de perturbación $v(n)$ cuando N tiende a infinito.

Para el mismo sistema, un modelo paramétrico realiza una predicción $\hat{y}(n, \theta)$ en la cual se tiene en cuenta una cierta modelización de la perturbación $v(n)$ aleatoria en términos de una función de transferencia $H(q, \theta)$ y de una señal $e(n)$ aleatoria que representa ruido blanco gaussiano. El hecho de poder realizar buenas conjeturas sobre el valor de la perturbación del sistema, permite mejorar la estimación de la señal de salida y de hecho este conocimiento se introduce en el cálculo de $\hat{y}(n, \theta)$ a través $H(q, \theta)$.

Además, los modelos paramétricos contienen, valga la redundancia, una serie de parámetros que permiten realizar un ajuste de los mismos a los datos reales según un cierto criterio. En efecto, si el error de predicción se define ahora como:

$$\varepsilon(n, \theta) = y(n) - \hat{y}(n, \theta) \quad (4.61)$$

puede definirse una función $V_N(\theta)$ como:

$$V_N(\theta) = \frac{1}{N} \sum_{n=1}^N \varepsilon^2(n, \theta) \quad (4.62)$$

Esta función mide cuan buenos son los valores de los parámetros del modelo en referencia al error global que se comete en la estimación. Naturalmente, la elección de los valores del vector θ ha de ser tal que minimice $V_N(\theta)$. Siendo esto así, el error de predicción para los modelos paramétricos se reduce a la señal $e(n)$.

Efectivamente:

$$\begin{aligned}y(n) - \hat{y}(n, \theta) &= G(q)x(n) + H(q)e(n) - H^{-1}(q)G(q)x(n) - (1-H^{-1}(q))y(n) \\ &= (1-H^{-1}(q)) \cdot (G(q)x(n) - y(n)) + H(q)e(n) = \\ &= (1-H^{-1}(q)) \cdot (H(q)e(n)) + H(q)e(n) = e(n)\end{aligned}\quad (4.63)$$

si el modelo es adecuado, el ajuste de $H(q, \theta)$ y $G(q, \theta)$ a las funciones $H(q)$ y $G(q)$ es tan bueno como lo pueda ser un polinomio de orden 2 a los datos de una parábola.

Tanto los métodos paramétricos como los no paramétricos, son capaces de extraer información del sistema que ayuda a conocer el comportamiento del mismo. La cuestión a plantear a priori ante un problema identificativo es qué método escoger. El criterio de elección debe estar guiado por conceptos como sencillez de aplicación y rendimiento del método. Por supuesto, el concepto de sencillez y rendimiento son relativos en el sentido de que no es posible conocer cuán sencillo o qué rendimiento (cualquiera que sea la definición concreta de ambos términos) ofrece un método antes de utilizarlo en el

problema concreto. Cada problema de identificación es diferente, precisa de un estudio particular y no es posible establecer a priori que método es mejor. Si además el método elegido es el paramétrico, esto supone que se deben realizar una nuevas elecciones para determinar el tipo de modelo paramétrico (ARX, ARMAX, BJ, etc.) y la estructura de dicho modelo.

Si es posible, sin embargo, establecer ciertos criterios generales, y en ese sentido, se puede decir que los modelos no paramétricos son más intuitivos e indican clara y directamente como responde el sistema. Una vez que se elige este método, no es necesario realizar más elecciones. Los algoritmos del proceso se encargan de, a partir de los datos, ofrecer una estimación del modelo. Sin embargo, el método no-paramétrico presenta un peor rendimiento que el paramétrico, y aun en ausencia de señales ruidosas, genera un cierto error ineludible causado por el hecho de tratar con N finita. Además son incapaces, en la práctica, de representar a sistemas con realimentación. Por otro lado, los métodos paramétricos, a pesar de una mayor complejidad para el usuario de un proceso identificativo, y tal vez una mayor dificultad interpretativa, ofrecen (para una estructura y orden adecuados) un error despreciable para sistemas sin ruido y para un mismo número N finito de datos que el método no paramétrico. En este sentido, los métodos paramétricos ofrecen un mayor rendimiento, a costa de un mayor esfuerzo.

En realidad ambos métodos pueden trabajar combinados. Debido a la sencillez de aplicación del método no paramétrico, es usual realizar una primera identificación del sistema a través de él, ya que proporciona, una visión global de la complejidad del sistema y ofrece estimaciones útiles que ahorran trabajo en el desarrollo de estructuras paramétricas.

4.15. Resumen de los resultados y conclusiones

En este capítulo se han desarrollado diferentes tipos de identificación y modelización lineal aplicados al sistema experimental.

Previamente a la aplicación de los modelos al sistema, se ha estudiado el rendimiento de los algoritmos a través de un sistema generado artificialmente. Con ello se ha pretendido presentar de una manera clara las ventajas y desventajas de cada tipo de modelo lineal.

Se ha demostrado la sustancial mejora en el rendimiento de los modelos dinámicos respecto a los modelos estáticos.

Para el caso de sistemas SISO se ha comprobado el rendimiento de los algoritmos lineales, que en el caso del sistema sometido a experiencias con gas tolueno es prácticamente inmejorable. Sin embargo, se ha constatado un cierto grado de no linealidad frente a las experiencias con gas octano.

Para el caso de sistemas MISO, aparece también un cierto grado de no linealidad justificable por la no linealidad de la respuesta al octano antes comentada. La discrepancia entre los datos reales y la aproximación lineal, no siendo desmesurada, si que justifica el intento de realizar una aproximación no lineal.

En cuanto a los algoritmos de identificación, se han comprobado que la identificación lineal paramétrica presenta un mejor rendimiento que la no paramétrica a costa de un cierto mayor grado de complejidad.

Bibliografía

- ¹T. Glad L. Ljung, *Modeling of Dynamic Systems* (Prentice Hall, New Jersey, 1994).
- ²L. Ljung, *System identification. Theory for the user* (Prentice Hall, New Jersey, 1987).
- ³Gwilym M. Jenkins George E.P. Box, *Time series analysis forecasting and control* (Holden-Day, Oakland, 1976).
- ⁴Jhon M. Parr Charles L. Phillips, *Signals, Systems and Transforms* (Prentice Hall, New Jersey, 1995).
- ⁵Ronald W. Schafer Alan V. Oppenheim, *Digital Signal Processing* (Prentice Hall, New Jersey, 1975).
- ⁶L. Ljung, *System identification Toolbox User's Guide (For use with MATLAB)* (The Mathworks Inc., 1995).
- ⁷K.J. Astrom, "Maximum Likelihood and Prediction Error Methods," *Automatica* **16**, 551-574 (1980).
- ⁸P. Stoica T. Soderstrom, *System Identification* (Prentice Hall, New Jersey,).
- ⁹S.A. Billings,... "Structure detection and model validity test in the identification of nonlinear systems," *IEE Proceedings* **130**, 193-199 (1983).
- ¹⁰H. Akaike "Information theory and an extension of the maximum likelihood principle" *Proc. 2nd Int. Symp. Information theory* 267-281 (1973)
- ¹¹H. Akaike, "A new look at the statistical model identification". *IEEE Trans. AC-19*, 716-722
- ¹²J. Rissanen. "Shortest data description and consistency of orders estimates of an ASMA process". *Int. Symp. on Systems Optimization and Analysis*. (1979)
- ¹³R. Isermann. "Practical Aspects of Process Identification". *Automatica*, **Vol 16**, 575-586 (1980)
- ¹⁴V. Strejc. "Trends in Identification", *Automatica* **Vol 17**, 7-21 (1981)
- ¹⁵. P.E. Wellstead. "Non-Parametric Methods of System Identification", *Automatica*, **Vol 17**, 55-59 (1981)
- ¹⁶K.R. Godfrey. "Correlation Methods" *Automática*, **Vol 16**, 527-534, (1980)
- ¹⁷I.J. Leontaritis, S. A. Billings, "Model selection and validation methods for non-lineal systems"

Capítulo 5

Métodos de identificación dinámica: Aproximación no lineal

- *Siempre se ha dado por supuesto que algo tan complicado como la sociedad humana se convertiría rápidamente en un sistema caótico y, por tanto, impredecible; pero lo que yo he hecho es demostrar que al estudiar la sociedad humana resulta posible escoger un punto de partida y hacer suposiciones adecuadas que suprimirán el caos. Eso hará posible predecir el futuro...*

Isaac Asimov: Preludio a la fundación

- *Pero existe otra clase de comportamiento, que la física maneja mal. Por ejemplo, todo lo que tenga que ver con la turbulencia: el agua que sale de un surtidor; el aire que se desplaza sobre el ala de un avión; el clima; la sangre que fluye a través del corazón. Los sucesos turbulentos se describen mediante ecuaciones no lineales. Son difíciles de resolver .. De hecho, habitualmente es imposible resolverlos.*

Michael Crichton: Parque jurásico

Prólogo del capítulo

Aunque los modelos lineales estudiados alcanzan buenas prestaciones en la estimación de la respuesta del sistema, la parte de señal que estos modelos no alcanzan a explicar tiene la suficiente importancia como para justificar el ataque de la identificación desde el punto de vista no lineal. El esfuerzo invertido en este ataque no lineal se verá recompensado con modelizaciones capaces de abarcar un tanto por ciento de información del sistema cercano al 100%. Conseguir este objetivo justifica ciertamente el uso de la teoría no lineal que sigue.

Si bien la identificación de sistemas dinámicos lineales es un tema no tan solo bien estudiado sino que ampliamente aplicado en diferentes campos como por ejemplo en ingeniería, control, biomedicina o econometría, no es menos cierto que el estudio de los sistemas no lineales ha recibido una menor atención. Esta descompensación se debe básicamente al hecho de que el estudio de sistemas no lineales supone un grado de complejidad elevado que restringe su uso a sistemas muy específicos. Existen, sin embargo, un relativamente amplio abanico de aproximaciones no lineales diferentes, algunas de las cuales están siendo objeto de un mayor interés últimamente^{2 4 5}, en este trabajo se han ensayado cuatro tipos diferentes: expansión en series de funcionales, sistemas de bloques o cajas, estimación paramétrica no lineal y, por último, redes neuronales. El desarrollo matemático formal de todo ello puede encontrarse en la literatura específica^{6 7 1 8}, aunque en los siguientes apartados se apuntan algunos aspectos fundamentales de los mismos, vinculados sobretodo a la aplicación para la cual van a ser utilizados.

5.1. Desarrollo de un sistema en series funcionales

Ha quedado establecido que un sistema lineal e invariante en el tiempo queda completamente caracterizado por la función de respuesta impulsiva $g(n)$.

$$y(n) = \sum_{k=0}^{\infty} g(k)x(n-k) \quad (5.1)$$

Sin embargo, si para los sistemas LTI, la respuesta impulsiva constituye una descripción completa del mismo, para los sistemas no lineales es tan solo una aproximación lineal de un desarrollo más general.

Si un sistema invariante en el tiempo es no lineal, bajo ciertas restricciones la relación entre la entrada y la salida del sistema puede formularse según la siguiente expresión^{6 36 37}:

$$\begin{aligned} y(n) = & h_0 + \sum_{j_1=0}^{\infty} h_1(j_1)x(n-j_1) + \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} h_2(j_1, j_2)x(n-j_1)x(n-j_2) \\ & + \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \sum_{j_3=0}^{\infty} h_3(j_1, j_2, j_3)x(n-j_1)x(n-j_2)x(n-j_3) + \dots \\ & + \sum_{j_1=0}^{\infty} \dots \sum_{j_n=0}^{\infty} h_n(j_1, \dots, j_n)x(n-j_1)\dots x(n-j_n) + \dots \end{aligned} \quad (5.2)$$

Si el sistema es además causal, se cumple que $h_n(j_1, \dots, j_n) = 0$ para cualquier $j_k < 0$ siendo $n=1, 2, \dots$ y $k=1, 2, \dots, n$.

Este impresionante desarrollo se denomina serie funcional y fue desarrollado por el célebre matemático Vito Volterra a finales del siglo pasado como generalización de la expansión en serie de Taylor, aunque, de hecho, la aplicación de estas series a la teoría de sistemas no lineales no se debe a Volterra sino a Norbert Wiener.

En reconocimiento a Volterra la serie 5.2 se denomina serie de Volterra y cada término $h_i(j_1, \dots, j_n)$ con $i=1, 2, \dots, n$ es una función de la serie denominado núcleo de Volterra o kernel de Volterra.

En cualquier caso, la serie puede describirse de la siguiente manera:

$$y(n) = H_0 + H_1[x(n)] + H_2[x(n)] + H_3[x(n)] + \dots + H_n[x(n)] \quad (5.3)$$

donde

$$H_n[x(n)] = \sum_{j_1=0}^{\infty} \dots \sum_{j_n=0}^{\infty} h_n(j_1, \dots, j_n) x(n-j_1) \dots x(n-j_n) \quad (5.4)$$

En esta representación, H_n , que representa el sumatorio aplicado sobre los kernels, se denomina operador de Volterra de orden n .

Un desarrollo como el descrito en la ecuación anterior representa al sistema a estudiar como una serie de cajas negras con salida $y_n(n) = H_n[x(n)]$ unidas para producir la salida $y(n)$ tal y como queda esquematizado en la figura 5.1. Más adelante se retornará a este concepto de modelización en forma de cajas.

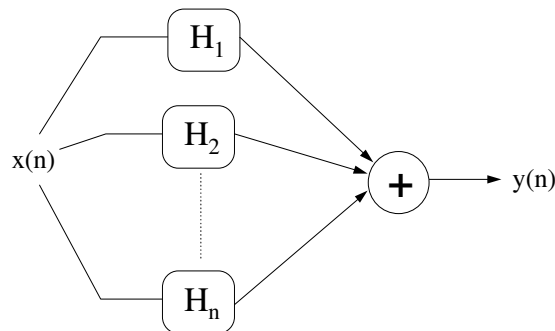


Figura 5.1 Modelo no lineal mediante operadores H_i

5.2. Descripción de la serie: Kernels de Volterra

Cada uno de los kernels de la expansión en serie de Volterra representa un cierto orden de aproximación de la serie a la señal $y(n)$. Así, el kernel de orden cero (h_0), que es de hecho una constante, representa una aproximación al modelo del sistema de orden cero. Un modelo de este estilo representaría al sistema con un valor constante. Es decir, al igual que las series de Taylor pueden desarrollarse alrededor de un punto, las series de

Volterra pueden extenderse alrededor de un nivel constante. De todas formas no es este primer kernel el más interesante.

Continuando con el análisis de la expresión 5.2, si $h_0 = h_2 = h_3 = \dots = 0$, la serie de Volterra se convierte en el producto de convolución que describe a los sistemas lineales e invariantes en el tiempo, siendo h_1 la respuesta impulsiva y conformando, el conjunto, una aproximación de orden 1.

$$y(n) = H_1[x(n)] = \sum_{j=0}^{\infty} h(j)x(n-j) \quad (5.5)$$

Si se restringe ahora la discusión al segundo orden para un sistema invariante en el tiempo, la señal de salida del sistema se define en términos del kernel de segundo orden $h_2(j_1, j_2)$ como:

$$y(n) = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} h_2(j_1, j_2)x(n-j_1)x(n-j_2) \quad (5.6)$$

El kernel de segundo orden puede ser interpretado como una extensión de la respuesta impulsiva a dos dimensiones, es decir una superficie de respuesta impulsiva que relaciona la salida del sistema con diferentes productos retrasados de la señal de entrada consigo misma. Ya que se supone que el sistema es causal, entonces $h_2(j_1, j_2) = 0$ para todo $j_1 < 0$ y para todo $j_2 < 0$ y la superficie h_2 sólo ocupara el cuadrante positivo del sistema de coordenadas tridimensional formado por los índices j_1 , j_2 y el kernel h_2 .

Realmente, el kernel $h_2(j_1, j_2)$ no responde exactamente a la definición de respuesta impulsiva. De hecho, como respuesta exacta a un impulso unidad para un sistema exclusivamente de segundo orden sólo se obtiene el kernel $h_2(j, j)$:

$$y(n) = H_2[\delta(n)] = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} h_2(j_1, j_2)\delta(n-j_1)\delta(n-j_2) = h_2(n, n) \quad (5.7)$$

Es decir, la respuesta impulsiva, como tal respuesta a un impulso de la entrada, sólo se corresponde con la diagonal del kernel h_2 .

Se debe remarcar que en esta discusión se ha considerado un sistema sólo dependiente del segundo orden en la forma: $y(n) = H_2[x(n)]$, pero la forma general de un sistema de segundo orden no es tan simple, y se describe así:

$$y(n) = N_2[x(n)] = H_0 + H_1[x(n)] + H_2[x(n)] \quad (5.8)$$

Es fácil demostrar ⁶ que la presencia de operadores de ordenes inferiores no altera la estimación de h_2 , pero si se desea determinar el kernel h_1 del sistema N_2 de la ecuación 5.8, resulta necesario conocer antes el kernel de segundo orden ya que la respuesta a un impulso del sistema N_2 viene dada por la expresión:

$$N_2[\delta(n)] = h_1(n) + h_2(n, n) \quad (5.9)$$

Por tanto, la respuesta impulsiva de primer orden $h_1(n)$ sólo puede ser conocida cuando la respuesta impulsiva de segundo orden $h_2(n, n)$ ya ha sido estimada.

Este desarrollo puede extrapolarse a órdenes superiores, y así pueden determinarse respuestas impulsivas de orden 3, 4... en general, para un kernel de orden p , la "respuesta impulsiva" relativa a ese orden y para un sistema MISO viene descrita por la expresión:

$$g_p(n) = p! H_p(n-T_1, \dots, n-T_p) \quad (5.10)$$

donde T_1, \dots, T_p , son los retrasos de cada uno de los impulsos en la entrada del sistema MISO⁶. De nuevo la determinación de kernels de orden inferior está limitada al conocimiento sucesivo y en cascada de los kernels de orden inmediatamente superior.

5.3. De los kernels de Volterra a los kernels de Wiener

La teoría básica de los kernels de Volterra expuesta hasta ahora ya ha dejado claro un problema práctico importante: su estructura impide que sean determinados de forma sencilla. Por un lado, para sistemas de orden finito se puede intentar, tal y como se ha expuesto, aislar cada uno de los operadores $H_i[x(n)]$ y ir derramando las sucesivas respuestas impulsivas comenzando desde el orden más alto hasta el más bajo. Sin embargo, no existe un método exacto para el caso en que el sistema no sea de orden finito, y deben aplicarse entonces métodos aproximados. En cualquier caso, el trabajo para calcular los kernels es largo y laborioso. Por otro lado, se puede⁶ demostrar que las series de Volterra presentan problemas de convergencia: sólo convergen para un rango limitado de amplitudes de la señal de entrada.

Para solventar estos problemas, Wiener reformuló las series de Volterra formando un nuevo conjunto de funcionales denominados funcionales G . Estos funcionales tienen la propiedad de ser ortogonales cuando la señal de entrada se corresponde con un proceso aleatorio de ruido blanco gaussiano. Las propiedades de ortogonalidad de estos funcionales hacen desaparecer los problemas de convergencia y de estimación de los funcionales de Volterra. De hecho, las series de Wiener son la mejor representación de un sistema, en el sentido de mínimos cuadrados, usando solamente n -kernels^{26 27 28 29 30}. La serie de Wiener se expresa de la siguiente forma

$$y(n) = \sum_{i=0}^{\infty} G_i(k_i, x(n)) \quad (5.11)$$

donde k_i es el kernel de Wiener de orden i , y G_i es el funcional G de orden i . Los funcionales de Wiener se definen a partir de la siguiente propiedad:

$$E[H_m[x(n)]g_n[k_n, k_{n-1(n)}, \dots, k_{0(n)}, x(n)]] = 0 \quad \text{para } m < n \quad (5.12)$$

donde $x(n)$ es una señal de ruido blanco gaussiano.

Es decir, para $x(n)$, el funcional $g_n[k_n, \dots, k_{0(n)}, x(n)]$ es ortogonal a cualquier funcional de Volterra de orden menor que n . Siendo entonces $x(n)$ una señal de entrada en forma de ruido blanco gaussiano, de media 0 y varianza P , y siendo además el sistema a identificar un sistema no lineal e invariante en el tiempo, los primeros cuatro funcionales se definen así^{6 10 3}:

$$\begin{aligned}
G_0[k_0, x(n)] &= k_0 \\
G_1[k_1, x(n)] &= \sum_{m=0}^{\infty} k_1(m)x(n-m) \\
G_2[k_2, x(n)] &= \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} k_2(m_1, m_2)x(n-m_1)x(n-m_2) - P \sum_{m=0}^{\infty} k_2(m, m) \\
G_3[k_3, x(n)] &= \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \sum_{m_3=0}^{\infty} k_3(m_1, m_2, m_3)x(n-m_1)x(n-m_2)x(n-m_3) \\
&\quad - 3P \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} k_3(m_1, m_2, m_2)x(n-m_1)
\end{aligned} \tag{5.13}$$

donde P es la densidad espectral de potencia del ruido blanco de la entrada.

Tanto las series de Volterra como las de Wiener son descripciones de las relaciones entrada y salida de un sistema no lineal invariante en el tiempo. La principal diferencia entre ellas es que las series de Wiener son ortogonales para entrada de ruido blanco gaussiano, aunque, de hecho, si el sistema no presenta no linealidades mayores que las de segundo orden, ambas descripciones son equivalentes, aunque no su estimación¹³.

5.4. Estimación de kernels mediante correlación (Lee-Schetzen)

Para efectuar el cálculo de los kernels de Wiener existen diversas alternativas. Una de ellas, quizás la más utilizada es el método de correlación cruzada con señales de ruido blanco, propuesto por Lee-Schetzen^{9 11 12 35}

Su desarrollo se basa en aplicar la propiedad de ortogonalidad de los kernels de Wiener. A partir de esta propiedad es fácil demostrar que^{6 9} si $y(n)$ se describe según 5.11, entonces $E[G_n[k_n, x(n)]] = 0$ para todo $n \geq 1$, siendo $x(n)$ ruido blanco. Es decir, para ruido blanco gaussiano, y para $n \geq 1$, la esperanza de un funcional G es cero. La consecuencia inmediata es que:

$$E[y(n)] = \sum_{i=0}^{\infty} E[G_i(k_i, x(n))] = G_0[k_0, x(n)] = k_0 \tag{5.14}$$

Por tanto:

$$k_0 = E[y(n)] \tag{5.15}$$

Aplicando el mismo método para kernels de orden superior, aparecen correlaciones cruzadas de las entradas de ruido blanco gaussiano y como resultado general, los kernels de Wiener se pueden calcular usando la siguiente expresión:

$$k_n(j_1, \dots, j_n) = \frac{1}{n!P^n} E \left\{ \left[y(n) - \sum_{m=0}^{n-1} G_m[x, k_m] \right] x(n-j_1) \dots x(n-j_n) \right\} \tag{5.16}$$

Obsérvese de esta ecuación la simetría existente en el producto de $x(n-j_i) \dots x(n-j_k)$, de tal modo que, particularizando por ejemplo en el orden 2, $k_2(j_1, j_2)$ es equivalente a $k_2(j_2, j_1)$.

Existen numerosos trabajos donde se estudian sistemas no lineales con modelos basados en estimaciones de kernels de Wiener a través de este método^{3 10}. Se deben matizar algunos aspectos relativos al cálculo práctico de los kernels. Uno de estos matices que deben exponerse es que la cantidad de cálculos necesarios para estimar los kernel crece exponencialmente a medida que aumenta el grado del kernel. Con ser esto una desventaja, no es tan importante como la que se deriva del hecho de utilizar un número finito de muestras, ya que un estimador basado en la operación de correlación, presenta una varianza que decae como $1/N$ y por lo tanto, la estimación de la correlación sólo es eficiente cuando en número de datos usados N tiende a infinito. Este problema es similar al que ya apareció en el caso de la estimación de la respuesta impulsiva de orden uno para sistemas LTI.

5.5. Aplicación de la estimación Lee-Schetzen a un problema sintético

Para ilustrar la teoría expuesta antes de adentrarse en la aplicación práctica que ocupa a este trabajo, se plantea la resolución con el problema sintético ya empleado en los análisis lineales. Sin embargo y aun no siendo estrictamente correcto, para demostrar ciertas propiedades de interés que no se han puesto de manifiesto en el desarrollo lineal, conviene general el sistema en ausencia de la señal $e(n)$ de ruido. Algunos de los procesos que se van a presentar a continuación son una reiteración de ciertos aspectos desarrollados durante el estudio lineal. No obstante, se tratará ahora de enfocar estos resultados hacia la consecución de la modelización no lineal y por tanto se profundizará en algunos puntos que anteriormente no recibieron atención. El sistema queda entonces definido según la siguiente ecuación:

$$y(n)=2x(n-1)+x(n-2)+0.5x(n-2)x(n-3)+0.9x(n-4)x(n-7)+0.5x(n-5)x(n-5) \quad (5.17)$$

Se observa que el sistema es no lineal a través de los tres últimos sumandos. Se utiliza como entrada $x(n)$ una señal aleatoria formada por 1000 datos cuya media presenta un valor cero y su varianza es de valor uno. La estimación de la parte lineal del sistema se realiza utilizando el análisis de correlación, que, de alguna forma, es la simplificación del método de Lee-Schetzen al orden 1. La primera aproximación es el kernel de orden cero, es decir el valor medio de la salida. En el ejemplo planteado el valor medio, que obviamente depende de la realización concreta de la señal aleatoria de la entrada, está alrededor de 0.4; en el caso concreto presentado, el valor medio, y por tanto el funcional de Wiener de orden cero y el kernel de orden cero es $G_0=0.42$.

Ya quedó claro en el estudio de las aproximaciones lineales que la media de la salida debía eliminarse para que la siguiente aproximación se desarrollase correctamente. El siguiente paso es calcular el kernel de orden uno o respuesta impulsiva, y así configurar el funcional de orden uno, que conforma junto con el kernel anterior, la estimación lineal del sistema. Aplicando pues las herramientas lineales a este sistema se obtiene, para 1000 muestras, una respuesta impulsiva de primer orden como la mostrada en la figura 5.2:

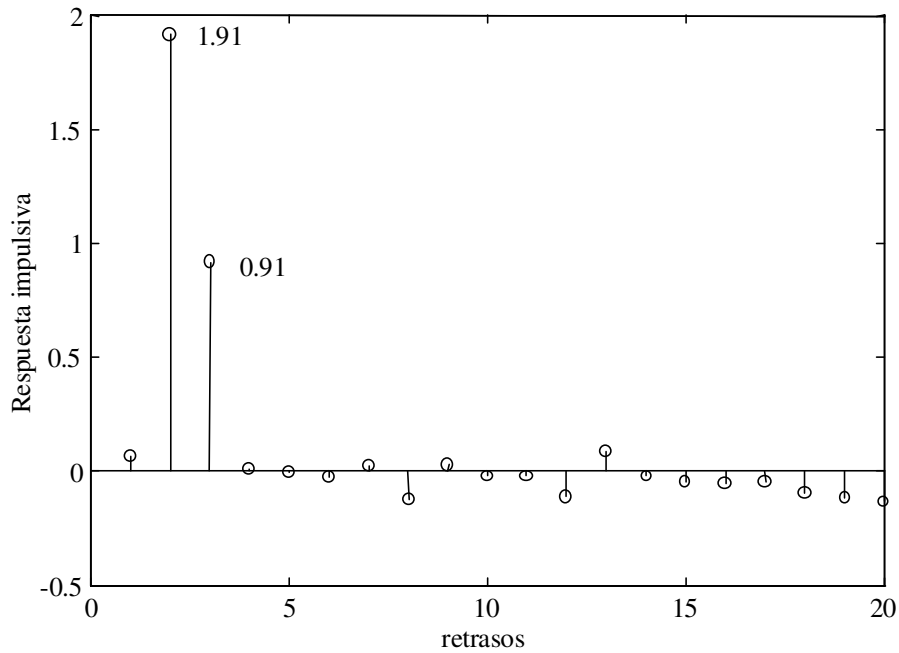


Figura 5.2 Estimación de la respuesta impulsiva de primer orden para el sistema sintético

Esencialmente, la forma es idéntica a la presentada en los apartados de aproximación lineal. Las pequeñas diferencias provienen de disponer de una estadística finita, es decir, son fruto de la realización concreta de las señales aleatorias (realizadas varios ensayos, se ha estimado que la desviación estándar de los valores significativos de la respuesta impulsiva, es inferior a 0.05).

Más allá de la respuesta impulsiva aparecen los términos no lineales. Para la estimación de estos términos se deben implementar algoritmos en base a ecuaciones como la presentada en 5.16. Así, para el primer término no lineal, hay que estimar el kernel de segundo orden que se representa en esta figura:

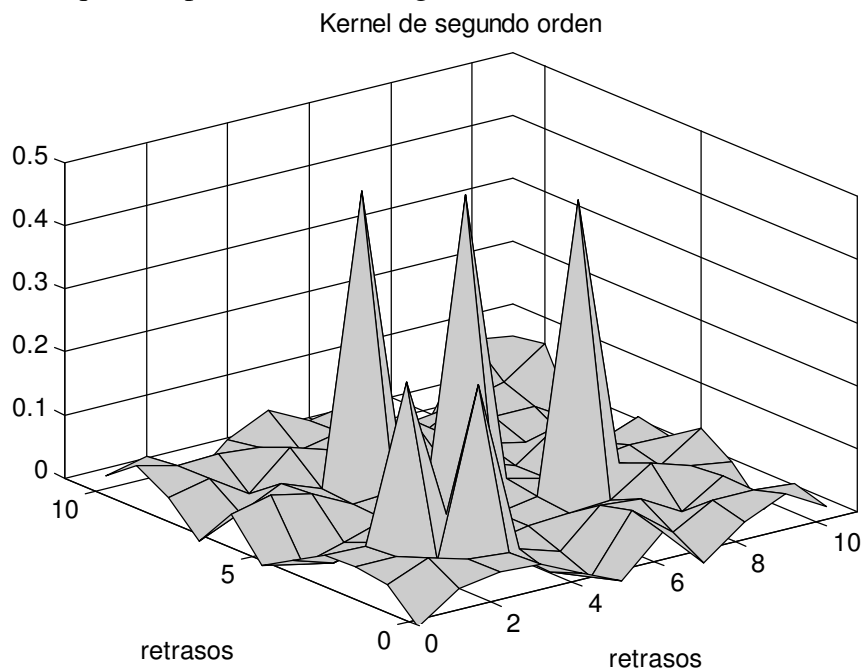


Figura 5.3 kernel del segundo orden para el sistema de la ecuación 5.17

Puesto que el kernel es simétrico, la estimación no es capaz de distinguir entre la posición (i,j) o la (j,i). La figura no puede mostrar entonces, otro resultado que la presencia dos picos asociados a las posiciones (2,3)-(3,2) y (4,7)-(7,4) que se corresponden, con los términos no lineales $x(n-2)x(n-3)$ y $x(n-4)x(n-7)$. Además la amplitud de los picos ha de ser de valor, dentro del error de cálculo debido a la estadística finita, la mitad del valor de los coeficientes no lineales de la ecuación. En la figura, las amplitudes de los picos en (2,3) o (3,2) son de 0.25, mientras que las amplitudes en (4,7) o (7,4) son de 0.43 aunque, de nuevo, dependiendo de la muestra de datos aleatorios, estos valores oscilan.

El tanto por ciento de señal no estimada (PSNE) por los diferentes grados de aproximación y para diferente cantidad de puntos, se presenta en la tabla 5.1:

Tabla 5.1 PSNE en función del número de datos usados en la estimación			
	Aproximación 0	aproximación 1	Aproximación 2
PSNE(%) 100 puntos	107.2 ± 7.5	21.2 ± 1.5	17.5 ± 1.2
PSNE(%) 300 puntos	102.2 ± 5.9	24.3 ± 1.4	15.5 ± 0.9
PSNE(%) 1000 puntos	100.7 ± 2.2	24.5 ± 0.5	3.9 ± 0.1
PSNE(%) 3000 puntos	100.3 ± 1.3	23.7 ± 0.3	1.3 ± 0.02
PSNE(%) 10000 puntos	100.0 ± 1	23.9 ± 0.2	0.3 ± 0.00

Como se ve en la tabla 5.1, si en lugar de utilizar muestras de 1000 datos, se amplía el número hasta 10000, los resultados mejoran ostensiblemente.

La figura muestra este kernel de segundo orden, observándose los picos mucho mejor definidos sobre el fondo general.

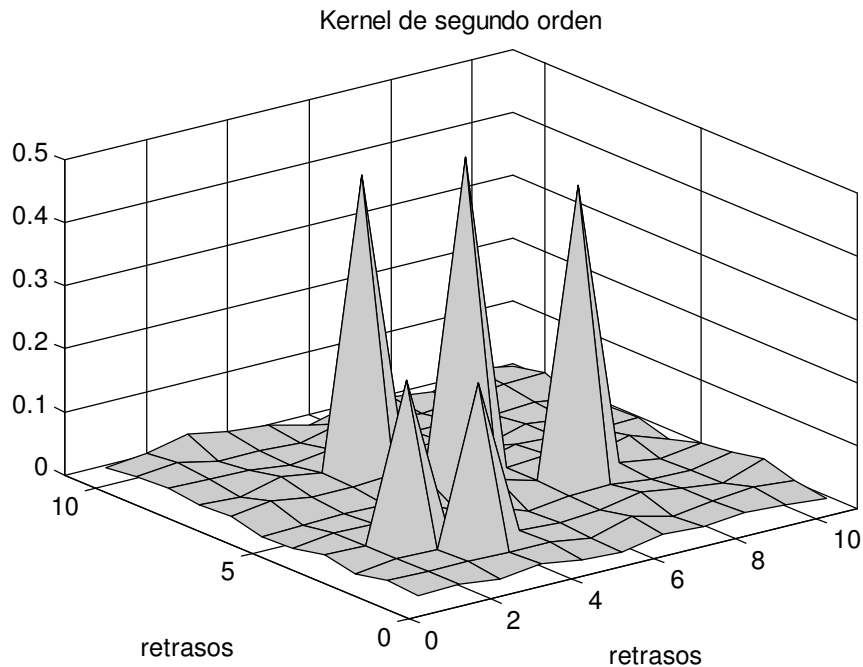


Figura 5.4 Kernel de 2º orden con resolución correspondiente a N=1000 puntos

Por contra, si se utilizan solo 300 datos, se puede comprobar que el error de estimación crece considerablemente.

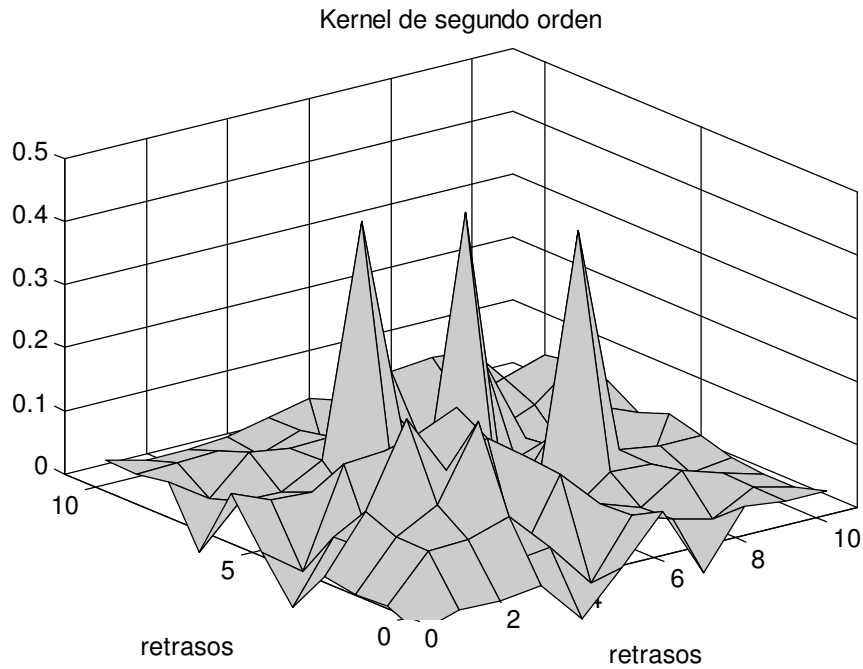


Figura 5.5 Kernel de 2º orden con resolución correspondiente a N=300 puntos

Con este ejemplo se ha observado que el número de datos influye gravemente en el resultado, obteniéndose mejores resultados cuanto más datos se utilicen, aunque por contra, el proceso de cálculo se hace más lento. En el caso de la aproximación lineal, la diferencia no es, en la práctica, apreciable, pero el número de cálculos a ejecutar cuando se realiza una aproximación no lineal crece de tal forma que las diferencias según el número de puntos se hacen notables. Sin embargo, este ejemplo no representa una situación real, ya que todo sistema contiene una cierta señal de ruido que en este caso no se ha contemplado. El ruido que aparece en los resultados presentados en el ejemplo es únicamente debido a la estadística de los datos. De hecho, si se practica el análisis de residuos a estos resultados, se comprueba que existe cierta correlación de los residuos con la señal de entrada, no cumpliéndose las condiciones de los tests de análisis de residuos lineal sin que esto indique que el modelo sea inadecuado. Simplemente, la calidad del estimador no es suficientemente buena. (El ejemplo presentado en el apartado lineal si que cumplía con las condiciones de los test de residuos, pero también el sistema contenía un cierto nivel de ruido).

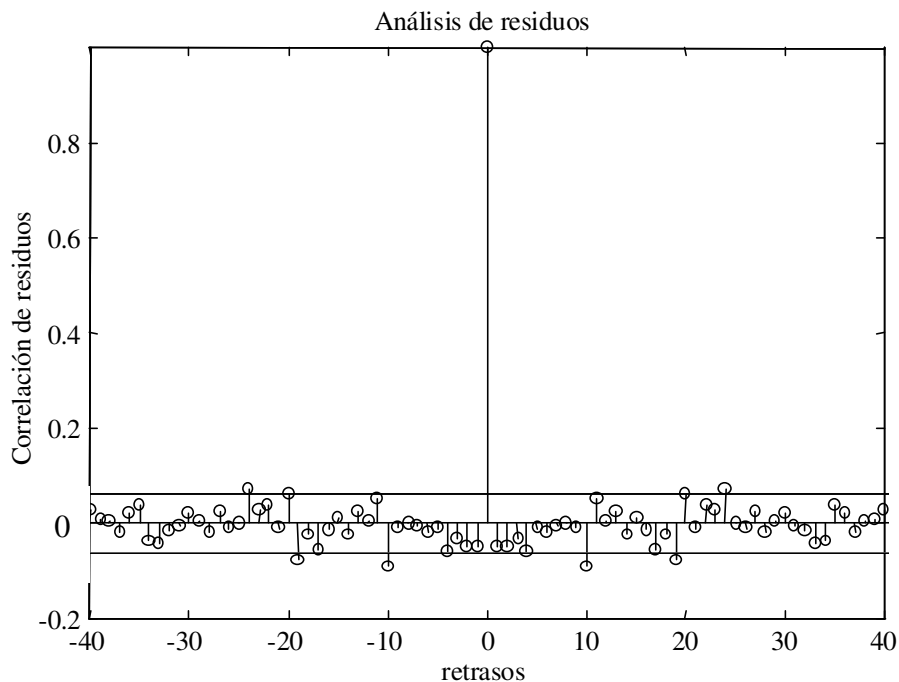


Figura 5.6 Test 1 del análisis de residuos : $R_{\epsilon\epsilon}$

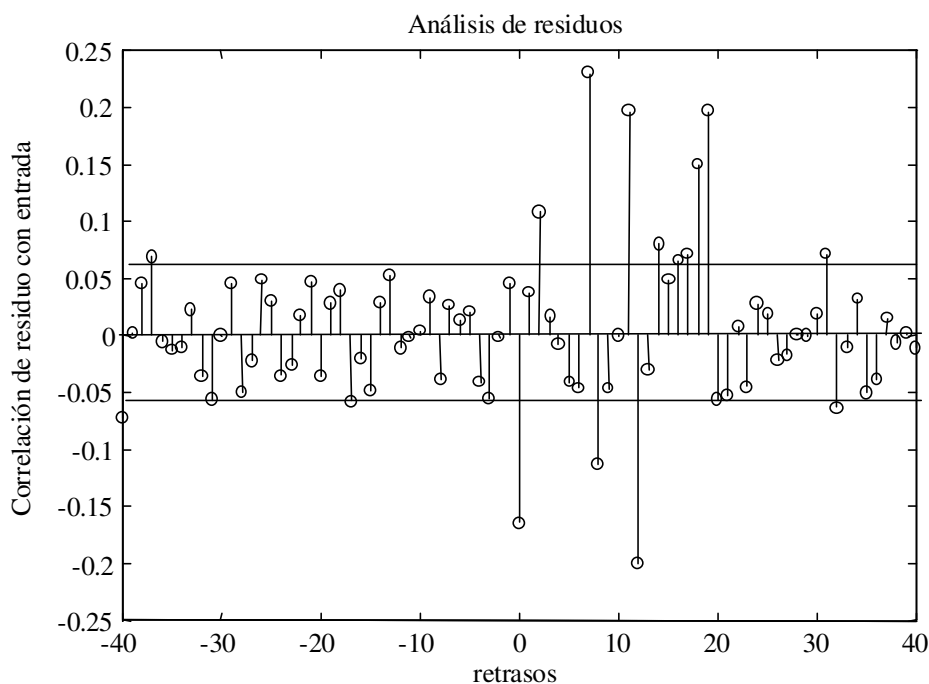


Figura 5.7 Test 2 del análisis de residuos: $R_{\epsilon x}$

Se puede observar en esta ampliación del test de correlación residuo-entrada como, siendo el modelo estimado adecuado al sistema, y aún cumpliéndose el hecho de que el residuo es blanco, existe una correlación con la entrada. Esta correlación es ineludible y procede únicamente de la diferencia entre los coeficientes de la función expresada en la ecuación 5.17 y los kernels estimados, por el hecho de usar una estadística finita.

Ciertamente, esta correlación no se puede eludir, pero sólo queda patente en ausencia de ruido adicional.

Para el caso de la estimación no lineal, la figura 5.8 presenta un resumen de los tests de análisis de residuos sobre el sistema 4.29 estimado en series de kernels, demostrándose que el residuo no contiene información del sistema.

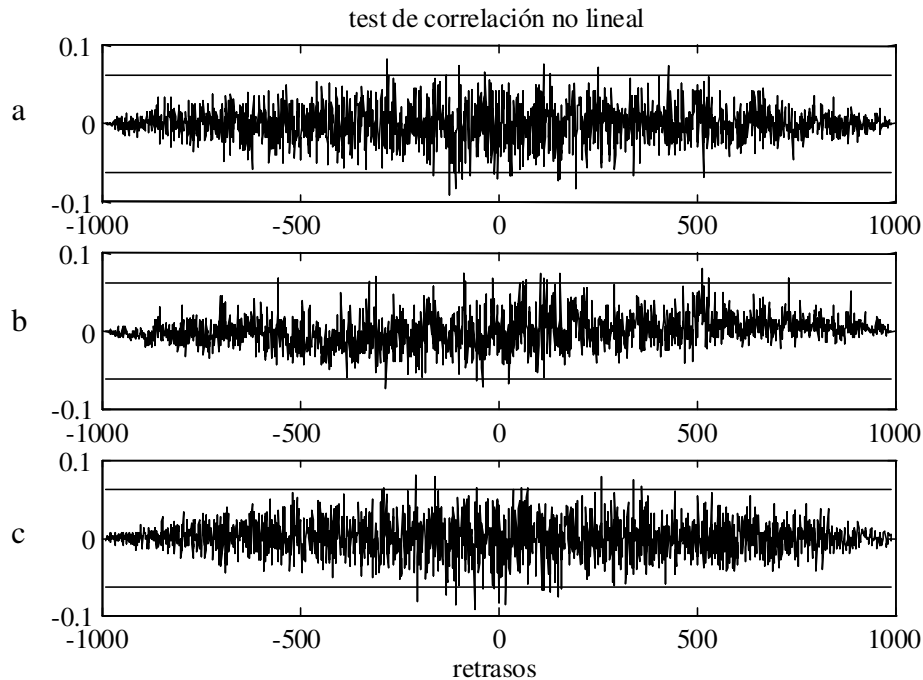


Figura 5. 8 Test de residual adicionales: a) correlación $R_{xx\epsilon}$ b) correlación $R_{xx\epsilon\epsilon}$, c) correlación $R_{\epsilon\epsilon x}$

Para acabar de completar estos análisis, falta analizar un punto que quedo sin cerrar en el análisis lineal. Quedó entonces claro que si se realizase el análisis de una aproximación lineal sobre este problema, el test de correlación $R_{\epsilon_{xx}}(n)$ debía fallar, ofreciendo un pico evidente puesto que en el residuo queda latente toda la información del sistema no lineal. La posición de dicho pico debía localizarse en el retraso 5 que se corresponde con la información aportada por el binomio $0.5x(n-5)x(n-5)$:

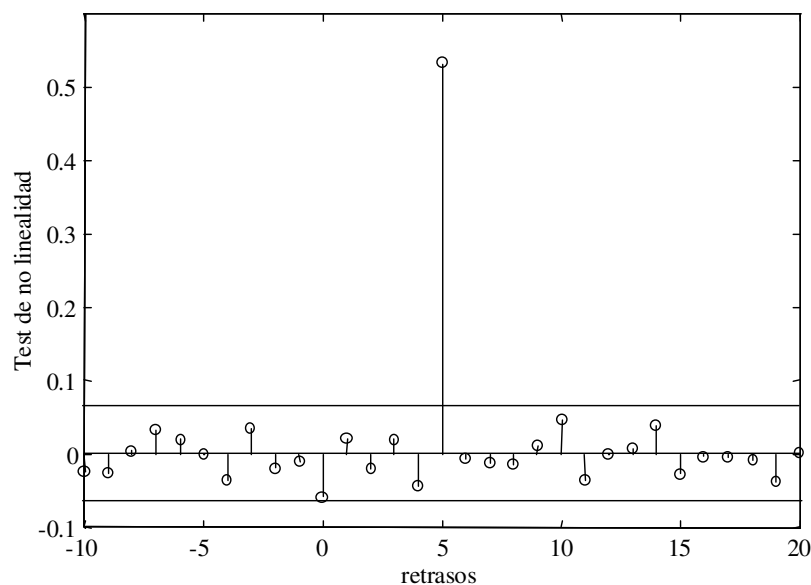


Figura 5. 9 Test de correlación $R_{\epsilon x(n-5)x(n-5)}$

Sin embargo, la estimación no lineal, como se ve en la figura anterior ha conseguido desentrañar esta información no lineal y, como se ve en la figura de test no lineales, no ha presentado ningún pico que revele la existencia de esta información en el residuo. Debe quedar claro, no obstante que este test sólo es útil para esta correlación concreta, y que no es capaz de dar información acerca de otras correlaciones. En este sentido, si se realizan los tests $R_{e(x(n))x(n-1)}$ y $R_{e(x(n))x(n-3)}$ para la estimación lineal y la no lineal, se puede observar que en la estimación lineal todavía queda información por determinar proveniente de los términos $0.5x(n-2)x(n-3)$ y $0.9x(n-4)x(n-7)$, mientras que la no lineal parece haber recogido toda la información del sistema.

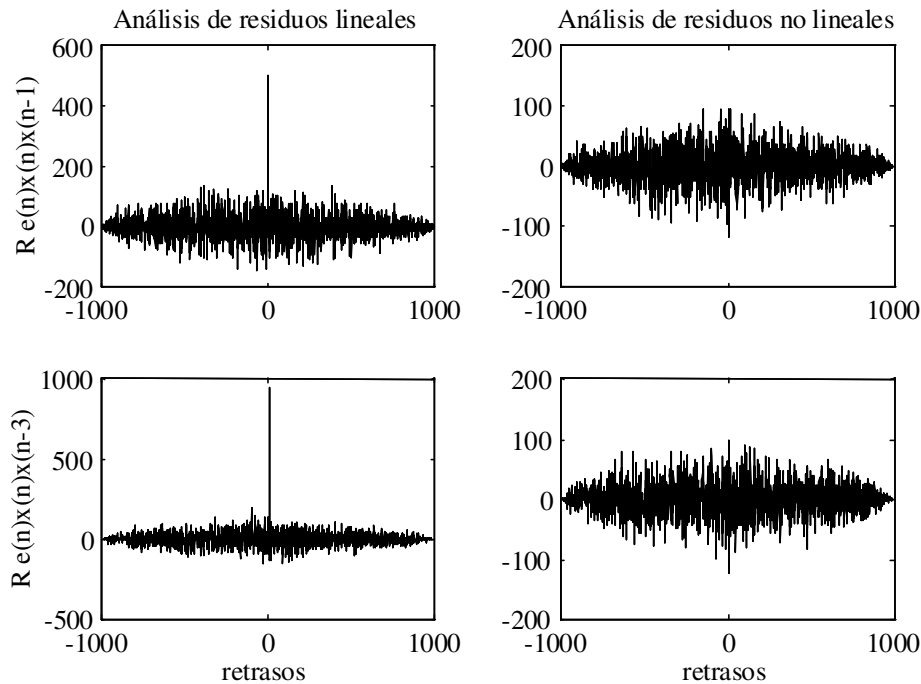


Figura 5.10 Análisis de residuos adicionales

Queda claro que los tests de análisis de residuos no abarcan, ni pretenden abarcar todo el espectro de posibles correlaciones entre entradas y salidas (de hecho eso es tarea de los kernels) sino que intentar dar unas pautas para marcar claramente la adecuación o no de los resultados de una estimación a través de una serie de correlaciones consideradas como habituales en sistemas no lineales

Vistos, a través de este ejemplo, los aspectos fundamentales de la estimación no lineal y del análisis de residuos, se va a entrar a analizar los datos del sistema sensor que ya se estudiaron desde el punto de vista lineal.

5.6. Aplicación de la estimación Lee-Schetzen al sistema

5.6.1 Identificación de un sistema SISO

Antes de plantear el problema desde el punto de vista multientrada se va a realizar, igual que en el caso lineal, el paso previo de analizar el sistema desde el punto de vista SISO. El sentido de hacerlo así viene dado por el hecho de que, aunque la estimación lineal del caso SISO era bastante precisa, al menos en el caso de gas octano no lo era suficientemente como para no permitir un cierto margen de maniobra a la estrategia no

lineal. Siendo los experimentos los mismos que en los apartados lineales, se entra a continuación a considerar directamente los resultados del método de Lee-Schetzen en la identificación del sistema.

Ajuste y validación

Los datos utilizados para esta estimación proceden del experimento 2 cuyos resultados ya se han expuesto anteriormente. Para la estimación de los diferentes modelos de los canales de salida del sistema se han utilizado 300 pares de datos entrada salida, mientras que para la validación se han usado 80 pares, como siempre, diferentes de los utilizados en la estimación.

La figura muestra el resultado del kernel de 2° orden para el gas octano sensor 2

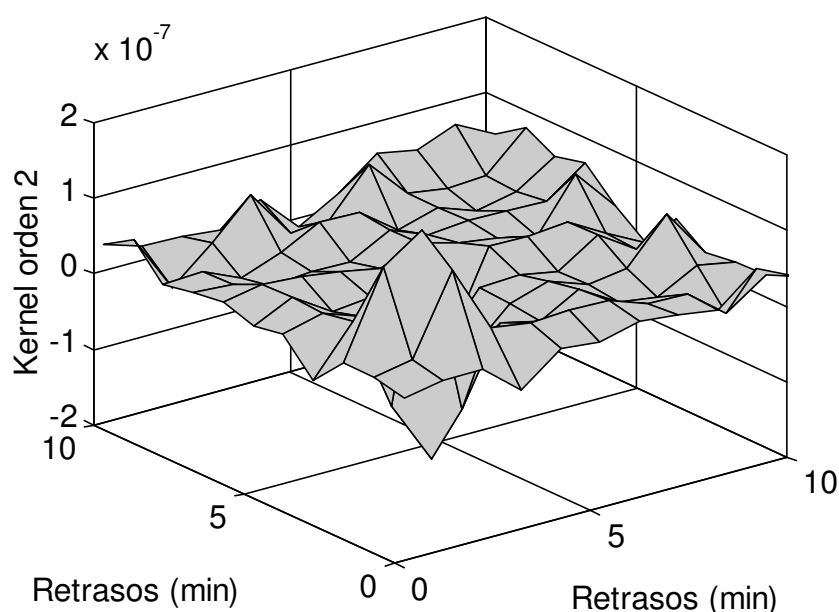


Figura 5.11 Estimación del kernel de orden 2 mediante correlación

La tabla 5.2 muestra los resultados para todos los sensores y los gases:

Tabla 5.2				
Sensores/ Gases	Octano		Tolueno	
	Estimación(%)	Validación(%)	Estimación(%)	Validación(%)
S1	3.2 ± 0.1	6.0 ± 0.5	3.2 ± 0.1	4.7 ± 0.4
S2	3.6 ± 0.1	3.9 ± 0.3	3.2 ± 0.1	4.1 ± 0.3
S5	4.4 ± 0.2	11.8 ± 0.9	3.0 ± 0.1	4.1 ± 0.3
S6	3.7 ± 0.1	5.0 ± 0.4	2.9 ± 0.1	4.0 ± 0.3

La figura 5.12 muestra una estimación del kernel de orden 2 con el doble de datos entrada salida. Se observa una mayor precisión en la resolución de la gráfica y los resultados sobre el valor de PSNE también mejoran. Hay que recordar, sin embargo, la limitación impuesta de 300 puntos para realizar todas las estimaciones. Son, por tanto, la figura 5.11 y la tabla 5.2 los resultados a considerar.

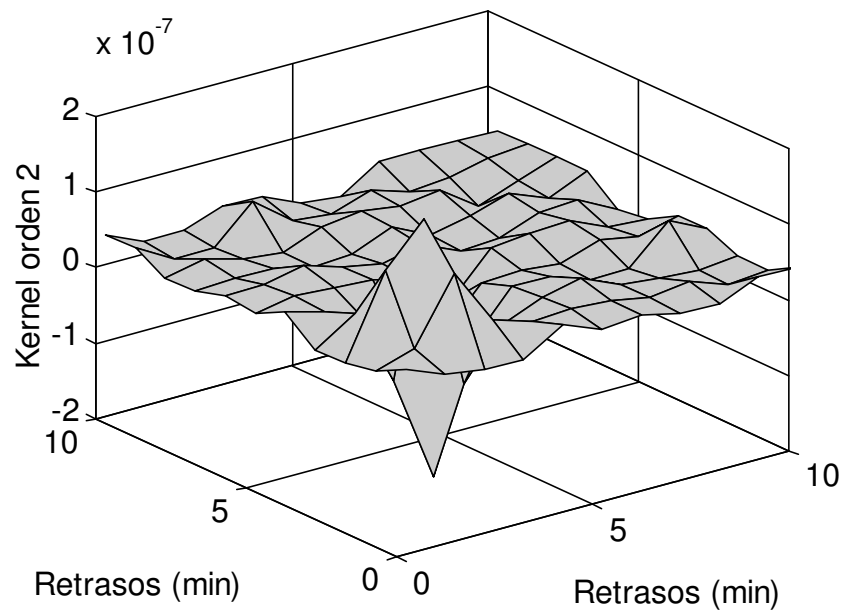


Figura 5. 12 Estimación del kernel de 2º orden con más resolución

Aceptación del modelo

Aunque los resultados hallados presentan ciertas mejoras respecto los resultados SISO lineales, estas no son significativas. En realidad, el tratamiento del problema a través de correlaciones ha llegado al límite de la información que puede extraerse usando sólo 300 puntos. Se demuestra por tanto que el hecho de que los desarrollos sean de tipo lineal o no lineal no es tan importante como el error que impone la limitación estadística.

5.6.2. Estimación de un modelo MISO mediante el método Lee-Schetzen

En este apartado se aborda directamente el problema no lineal para un sistema MISO generado a través del experimento 3, utilizando la teoría de kernels de Wiener estimado a través del cálculo de correlación ideado por Lee-Schetzen.

Los sistemas MISO no lineales plantean la aparición de nuevos kernels que hasta este punto no se habían tomado en consideración. Efectivamente, ya que los sistemas MISO contienen varias entradas, en la teoría de los kernels de Wiener aplicados a un sistema MISO no lineal aparecen, a partir del segundo orden, una serie de kernels nuevos que no aparecen en los sistemas SISO. El caso MISO debe, como generalización del caso SISO, disponer de kernels que, a partir del segundo orden de aproximación, contengan las relaciones no lineales de cada entrada consigo misma. Sin embargo, es necesario tener en consideración las posibles relaciones del sistema con productos cruzados de las entradas. En el caso de aproximación de orden dos, las combinaciones de las entradas serán productos cruzados dos a dos de las diferentes entradas. Estos kernels de orden 2, que referencian las relaciones entre productos de entradas cruzadas, se denominan kernels cruzados o crosskernels. La notación para el caso de sistema MISO de dos entradas (que, es el caso que ocupa a este trabajo) es la siguiente:

$$\begin{aligned}
G_0(x_1, x_2) &= k_{0,0} \\
G_1(x_1, x_2) &= \sum_{i=0}^N k_{1,0}(i)x_1(n-i) + \sum_{i=0}^N k_{0,1}(i)x_2(n-i) \\
G_2(x_1, x_2) &= \sum_{i_1=0}^N \sum_{i_2=0}^N k_{2,0}(i_1, i_2)x_1(n-i_1)x_1(n-i_2) - P_1 \sum_{i=0}^N k_{2,0}(i, i) + \\
&\quad \sum_{i_1=0}^N \sum_{i_2=0}^N k_{0,2}(i_1, i_2)x_2(n-i_1)x_2(n-i_2) - P_2 \sum_{i=0}^N k_{0,2}(i, i) + \\
&\quad \sum_{i_1=0}^N \sum_{i_2=0}^N k_{1,1}(i_1, i_2)x_1(n-i_1)x_2(n-i_2) \\
&\dots
\end{aligned} \tag{5.18}$$

Efectivamente, en la definición de las ecuaciones aparece, en el segundo orden, un término que hace referencia a productos cruzados entre valores retrasados de las dos entradas $x_1(n)$ y $x_2(n)$.

La estimación del crosskernel mediante el método de Lee-Schetzen sigue la siguiente fórmula:

$$k_{1,1}(i_1, i_2) = \frac{1}{P_1 P_2} E[y(n)u_1(n-i_1)u_2(n-i_2)] \tag{5.19}$$

Introducida la formulación MISO se ha de aplicar a los datos del experimento 3, pero hay que hacer notar el significativo aumento de variables a calcular por el hecho de trabajar con un sistema MISO.

La figura 5.13 muestra el autokernel de segundo orden para el sensor 2 y gas octano.

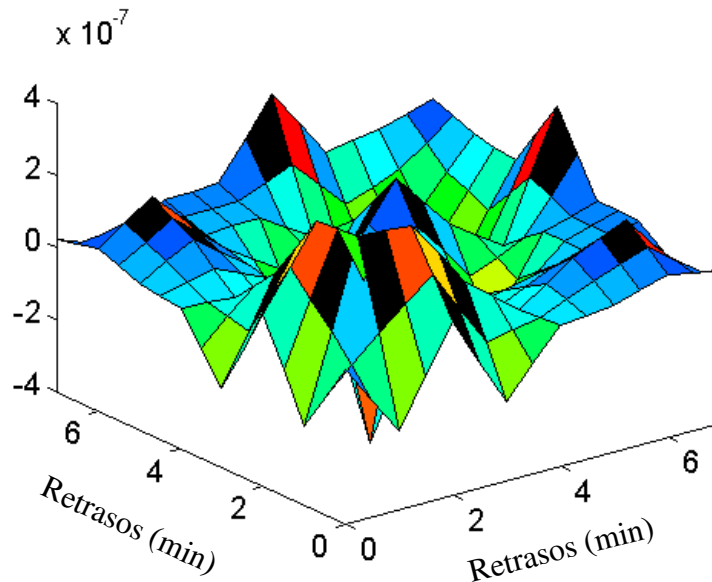


Figura 5.13 Estimación del autokernel de 2º orden para el sensor 2 y gas octano

La tabla 5.3 resume los resultados para esta aproximación de segundo orden.

	Sensor 1	Sensor 2	Sensor 5	Sensor 6
PSNE estimación (%)	79.3 ± 3.2	86.9 ± 3.5	76.7 ± 3.1	84.7 ± 3.4
PSNE validación (%)	53.6 ± 4.2	55.8 ± 4.4	54.7 ± 4.3	56.8 ± 4.5

Aceptación del modelo

Como puede observarse y como cabía esperar a partir de los ajustes con el sistema SISO, los resultados son decepcionantes. Es evidente que esta aproximación no lineal basada en el método de correlación no sólo no mejora las prestaciones de la estimación lineal MISO sino que las empeora. Si el análisis de correlación ya era poco satisfactorio para el caso de dos gases y aproximación lineal, en el caso no lineal, el ruido inherente en las señales y, sobretudo, la limitación en la cantidad de datos utilizados frente al aumento de variables a calcular por el hecho de trabajar con un sistema MISO no lineal, hacen inviable un buen ajuste a través de este método. Ante la evidencia de los resultados no es necesario plantearse el análisis de residuos: es evidente que no cumplirán los requisitos exigidos.

5.7. Estimación mediante modelado por cajas

El método de desarrollo en series funcionales de Wiener o Volterra proporcionan una caracterización de un sistema no lineal, pero, sin embargo, el cálculo de ordenes altos de los kernels es farragoso y tiene una difícil interpretación. Hasta ahora sólo se ha explicado el segundo orden no lineal y sin embargo, ya se ha comprobado la complejidad del cálculo y su bajo rendimiento en relación al número de datos. Si el sistema presentase no linealidades de orden alto, el problema de la estimación no sería factible a través de la correlación.

Existe una alternativa para desarrollar modelos de sistemas no lineales; son las denominadas redes de bloques o cajas^{15 16 17 31 32}. Con estos modelos los operadores lineales dinámicos quedan representados por un bloque L mientras que los operadores no lineales estáticos están representados por bloques N. El modelo del sistema completo quedará constituido por una serie de bloques L y/o N interconectados en serie y paralelo. De hecho, cada bloque L o N no es más que una cierta representación matemática, y no tienen porque relacionarse con componentes reales del sistema. Trabajando bajo esta concepción, se pierde el sentido de sistema como una única caja negra, transformándose en una serie interconectada de subsistemas representados por las cajas L-N. De alguna manera, los modelos de cajas están relacionados con la interpretación dada en la figura 5.1 a través de los operadores $H_i(x(n))$ de Volterra. De hecho, tal y como demostró Korenberg^{14 8} un sistema discreto no lineal con memoria finita y con una aproximación de grado M en series de Volterra con $M < \infty$ puede representarse por un número finito de bloques en paralelo LN o en cascada LNL.

Un sistema LNL puede esquematizarse según el siguiente diagrama de bloques:

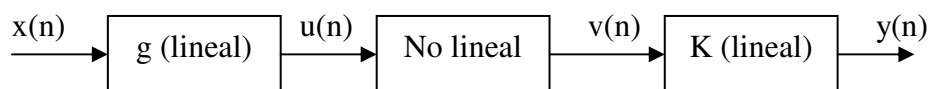


Figura 5.14 Esquema de un sistema LNL

El primer bloque del diagrama, el bloque g , representa un sistema lineal dinámico. Le sigue a continuación un sistema estático no lineal, para acabar de nuevo con un bloque lineal dinámico k . El proceso a seguir para estimar cada uno de los bloques puede encontrarse en la literatura^{16 17 31 32} pero es un proceso largo y complejo. Si además el sistema no se adapta a la forma LNL, la convergencia de la estimación hacia el mejor modelo en sentido de mínimos cuadrados no está garantizada. El complicado proceso de estimación de un modelo LNL puede simplificarse si se utilizan unas ciertas subclases de los propios modelos: las estructuras de cajas de Wiener (no confundir con las series de Wiener) y las estructuras de cajas de Hammerstein^{17,22,23,24}. La figura 5.16 presenta diagramas para describir ambas estructuras. La estructura general LNL puede entonces considerarse como una combinación de un modelo lineal seguido de una estructura Hammerstein o como una estructura Wiener seguida de una caja lineal. Diferentes artículos^{15,16,17} hacen referencia a como estimar estas subclases de estructuras, aunque los pasos básicos a seguir para estimar una estructura Wiener son los siguientes

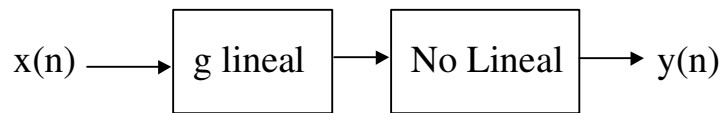


Figura 5.15 Esquema de la estructura de Wiener

- 1) Estimar g a partir de una estimación lineal entre $x(n)$ e $y(n)$.
- 2) Estimar $u(n)$ a partir del operador lineal g y $x(n)$
- 3) Ajustar un polinomio no lineal entre $u(n)$ e $y(n)$
- 4) Reestimar $u(n)$ a través del inverso de dicho polinomio
- 5) Reestimar el operador lineal g esta vez entre $x(n)$ y $u(n)$
- 6) Volver a estimar $u(n)$ a partir del nuevo operador g y $x(n)$
- 7) Ir al punto 3 y ejecutar una nueva iteración o pasar al punto 8 y finalizar
- 8) Ajustar un polinomio entre $u(n)$ e $y(n)$ y determinar con el la estimación de $y(n)$,

Sin embargo, si bien la filosofía de las cajas puede permitir una mejor interpretación física de la realidad del sistema y son una alternativa eficaz al tedioso problema de cálculo de los kernels, no existe un método sencillo para obtener fácilmente la topología de cajas que mejor se ajuste al sistema.

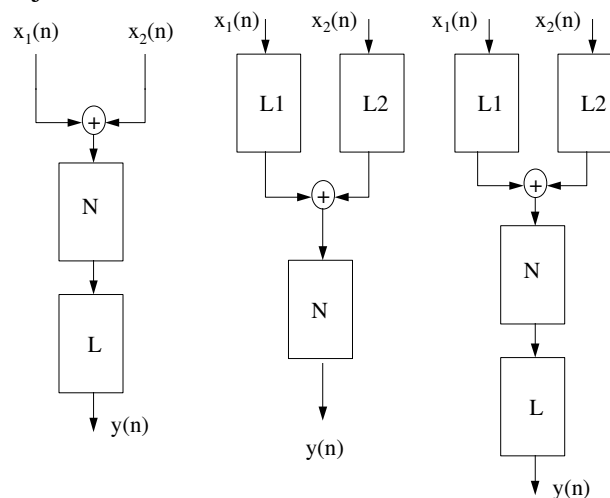


Figura 5.16 Diferentes topología de modelos por cajas

La figura 5.16 muestra algunas posibles estructuras de cajas a usar en el caso de dos entradas y una salida. El esquema representado a la derecha es una estructura general LNL para un sistema MISO de dos entradas, en la que cada entrada ha de pasar en cascada por una serie de bloques lineal-nolineal-lineal. La estructura central es la subclase LN de Wiener para un sistema MISO de dos entradas. En ella, la salida de los bloques lineales se suman y la señal unidimensional resultante se introduce en un bloque no lineal. La estructura Wiener es relativamente sencilla y muy empleada en los problemas de identificación. Debido a esta simplicidad y a sus buenos resultados es la primera estructura que suele probarse cuando el problema se sale de los márgenes lineales. El esquema de la izquierda es una aplicación para dos entradas de una estructura de bloques de Hammerstein. De alguna forma, la estructura Hammerstein es más sencilla que la de Wiener, pero no es tan útil; de forma que algunos sistemas no lineales no pueden ajustarse a estructuras Hammerstein. De hecho, la demostración de Korenberg acerca de la relación entre una aproximación a un sistema en serie de Volterra finita y los modelos de cajas LN o LNL no puede ajustarse a los modelos de cajas de Hammerstein, que son de hecho un modelo de cajas NL. Esta diferencia en el orden de las cajas, es por tanto crucial, y así, siendo una estructura más sencilla, no engloba la generalidad de los modelos LN.

Como conclusión y resumen de esta exposición, puede considerarse que la verdadera utilidad de los modelos de bloques reside en modelar sistemas con no linealidades más allá del segundo orden. En el caso del sistema de medida de gases objeto del presente trabajo, no será necesario, como se demostrará más adelante llevar la modelización no lineal más allá del segundo orden y por tanto, en este caso, los modelos de bloques no aportan mejoras apreciables a los resultados ya conseguidos.

5.8. Aplicación del modelado de cajas a un problema sintético

Siguiendo el mismo desarrollo llevado a cabo hasta ahora, se van a aplicar las técnicas de modelado por cajas a un problema sintético antes de utilizarlas con el sistema de sensores real. Sin embargo, a diferencia de lo acostumbrado hasta ahora, no se va a entrar a estudiar directamente el sistema ejemplo 4.29, sino que se van a generar y estudiar una serie de sistemas crecientes en complejidad. De esta manera se pretende dar una visión más completa de las particularidades de este tipo de modelización.

Dentro de las posibles estructuras, se va a centrar la atención en el desarrollo de cajas con estructura de Wiener por su relativa sencillez y su equivalencia con los desarrollos en serie de Volterra.

Se inician los ensayos con un sistema muy simple cuya relación entre la entrada y la salida se muestra a continuación:

$$y(n) = x(n-1) + x(n-1)*x(n-1) + 0.1*e(n) \quad (5.20)$$

donde $x(n)$ es una señal de ruido blanco con distribución gaussiana, de media cero y desviación estándar uno y $e(n)$ también es una señal de ruido blanco con distribución gaussiana, de media cero y desviación estándar uno.

La estructura a probar sobre este sistema consiste en una estructura de Wiener de una entrada y una salida:

En el primer paso, la estimación lineal y la no lineal para este sistema ofrece los resultados que se muestran en las siguientes figuras:

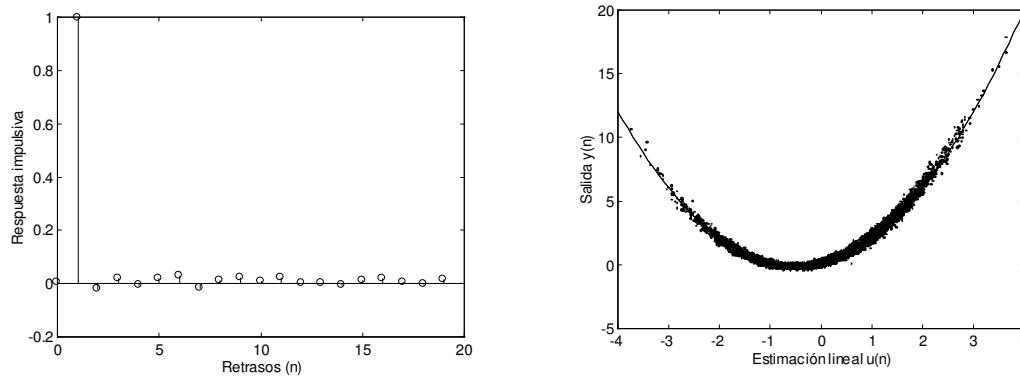


Figura 5. 17 Resultados de la estructura Wiener

La primera figura muestra la respuesta impulsiva que, efectivamente muestra un valor significativo en el retraso 1. El cálculo de esta relación lineal, correspondiente al paso 1 expuesto anteriormente, se realiza a través de una estimación paramétrica según un modelo FIR. En la segunda figura se muestra en línea continua el polinomio no lineal de orden 2 estimado como relación entre la señal $u(n)$ e $y(n)$. Los puntos representan la relación efectiva entre $u(n)$ y $v(n)$.

En el ajuste lineal, el valor de la función PSNE es de un 66.5 %, mientras que con la aproximación no lineal se llega a valores de 1% de la PSNE en el primer paso. Aunque la estimación lineal presente un valor tan alto de la función PSNE, esto no es indicación de que no sea una buena estimación. El valor PSNE alto viene motivado por el hecho de que la parte no lineal tiene un peso muy importante en este sistema. El algoritmo de estimación debe ahora seguir iterando, y sin embargo no parece posible que, con el nivel de ruido presente, los siguientes pasos puedan mejorar más este primer resultado, (de hecho eso es lo que ocurre). Sólo en el caso concreto en el que la realización particular de las señales $x(n)$, $y(n)$ y del ruido $e(n)$ produce un primer resultado desviado en el primer paso, las siguientes iteraciones pueden corregirlo. Sin embargo, lo habitual es que ya la primera iteración converja a un valor mínimo. Los valores ofrecidos de la función PSNE son orientativos puesto que, por supuesto, dependiendo de la realización concreta de las señales $x(n)$ e $y(n)$, dichos valores pueden estar sujetos a pequeñas variaciones. Hay que remarcar, como siempre, la dependencia de los cálculos con el número de puntos utilizado. En el ejemplo presentado se han generado 1000 datos x - y , y con ellos se ha estimado una respuesta impulsiva de hasta 20 términos (a la vista del sistema, de hecho era innecesario conocer tantos datos de la respuesta impulsiva, pero realmente nunca se dispone de un conocimiento a priori de la relación x - y). Si se le exige al mismo algoritmo de cálculo realizar la misma estimación pero contando con un número de puntos igual a 100, los resultados son de un rendimiento pésimo como muestra la figura 5.18

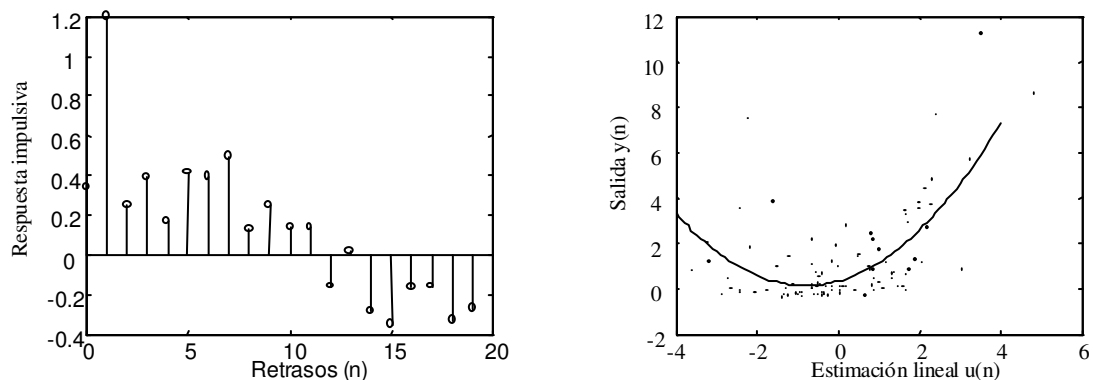


Figura 5.18 Resultados de la estructura Wiener

Se observa como la representación de la respuesta impulsiva puede llevar a engaño al presentar unos retrasos con valores que no parecen en absoluto despreciables aunque, esos si, el retraso 1 sigue apareciendo como una dependencia destacada. Por otro lado, la dispersión de los datos respecto al polinomio no lineal estático de orden 2 estimado es mucho mayor que en el caso anterior. De hecho, en este primer paso de la estimación los valores de la función PSNE son de un 89% para la estimación lineal y de un 50% para la estimación no lineal. Tampoco en este caso las siguientes iteraciones logran mejorar estos pobres resultados. Ahora bien, para acabar de confirmar este aspecto, y puesto que son conocidas a priori las relaciones del sistema, se va a ejecutar la estimación pero sin pretender hallar tantos puntos de la respuesta impulsiva. Se debe enfatizar que el conocimiento de las relaciones del sistema es un conocimiento no disponible habitualmente y que por tanto se está haciendo uso de una información privilegiada para el desarrollo de estas estimaciones. Sin embargo, ya se expuso anteriormente, que el objetivo de estos problemas sintéticos se centra en comprobar el rendimiento de los algoritmos de identificación, la estricta aplicación del método se reserva para el problema real. Hecha esta salvedad, queda claro de la ecuación 5.20 que sólo es necesario estimar 2 datos, correspondientes al retraso 0 y al retraso 1, así que esa es la cantidad de datos mínima necesaria a estimar en la aproximación lineal. Para el ajuste no lineal estático basta con un polinomio de orden 2. Desarrollado el algoritmo con estos nuevos parámetros, los resultados son los siguientes:

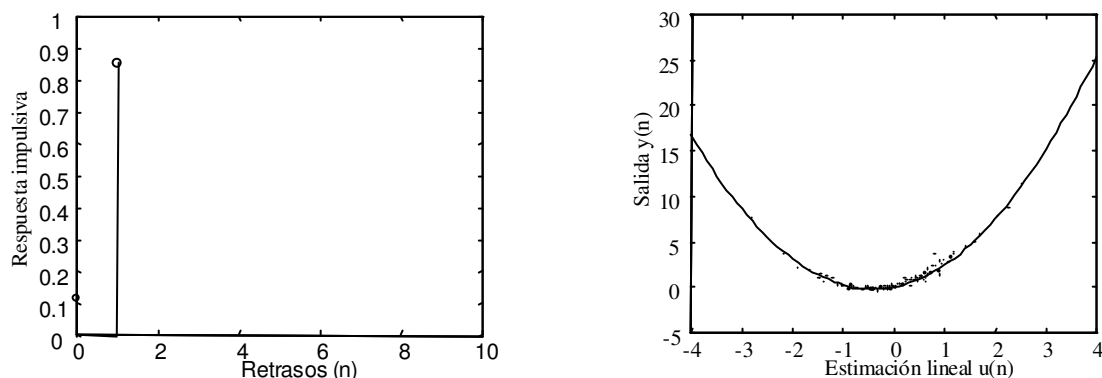


Figura 5.19 Resultados de la estructura de Wiener

Los valores de PSNE en este primer paso son de aproximadamente 80% para el ajuste lineal y de cerca del 1% para el no lineal. Aun pueden obtenerse mejores resultados si se fuerza al término de retraso 0 a tener un valor 0; en cualquier caso queda demostrado

que el número de datos afecta al cálculo pero que con un número suficiente de los mismos y con la estructura adecuada se puede extraer toda la información del sistema. Aunque los resultados obtenidos son buenos hay que tener en cuenta que el sistema probado es muy simple y se debe investigar el rendimiento para sistemas más complejos. Por ejemplo, el simple hecho de introducir un nuevo retraso hace que esta estructura, tal y como se ha desarrollado no sea capaz de extraer toda la información. Sea entonces un nuevo sistema como el que sigue:

$$y(n) = x(n-1) + x(n-2) + x(n-1)*x(n-1) + 0.1*e(n) \quad (5.21)$$

Para trabajar con este sistema se va a adoptar la estructura de Wiener presentada en la figura 5.18. Suponiendo que no se conoce la relación entre la entrada y salida, se van a generar más muestras de la respuesta impulsiva que las estrictamente necesarias (en realidad tres es el número de muestras de la respuesta impulsiva para describir a este sistema desde el punto de vista lineal). El número de datos con el que se trabaja en este caso va a ser 1000, se va a estimar una respuesta impulsiva con 10 muestras. El polinomio no lineal se va a mantener en el orden 2.

Los resultados de la primera iteración son los siguientes:

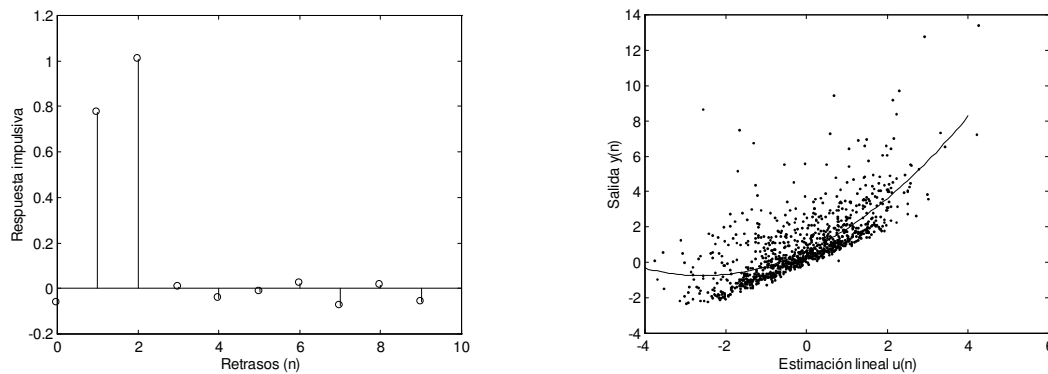


Figura 5.20 Resultados de modelo de Wiener para sistema no de Wiener

De la figura que representa la respuesta impulsiva se puede observar como los únicos datos válidos están, como no podía ser de otro modo, en el retraso 1 y en el retraso 2. Aunque ambos valores deberían ser iguales entre si e iguales a 1, la primera iteración del algoritmo no es capaz de llegar a este resultado. Por otra parte, la relación entre la estimación lineal y la no lineal, presentada en la otra figura, ya no es exactamente cuadrática como en el caso del sistema 5.20. Lógicamente, al ajustar un polinomio cuadrático a la estimación lineal, aparecen términos del tipo $x(n-1)*x(n-2)$ y $x(n-2)*x(n-2)$ como residuo. El sistema no depende de ellos pero al mismo tiempo, con la estructura planteada, son inevitables. Este hecho hace que el rendimiento de la estimación se empobrezca. El hecho es que en este caso, las sucesivas iteraciones consiguen mejorar algo la estimación del primer paso, aunque el rendimiento global no es, ni mucho menos, igual que en el sistema anterior. Este hecho era de esperar puesto que en este caso, la estructura Wiener solo es una aproximación a la estructura del sistema. La función PSNE para la estimación lineal adopta un valor superior al 50%. En el primer paso no lineal este valor se transforma en un 45% mientras que las sucesivas iteraciones pueden conseguir valores del 40%. Pero esta mejora de las sucesivas iteraciones se

concentra sobre todo en estimar mejor los parámetros lineales. Así, después de dos iteraciones, como se puede apreciar en la figura 5.21, los valores re-estimados de la respuesta impulsiva se asemejan más a los valores teóricos

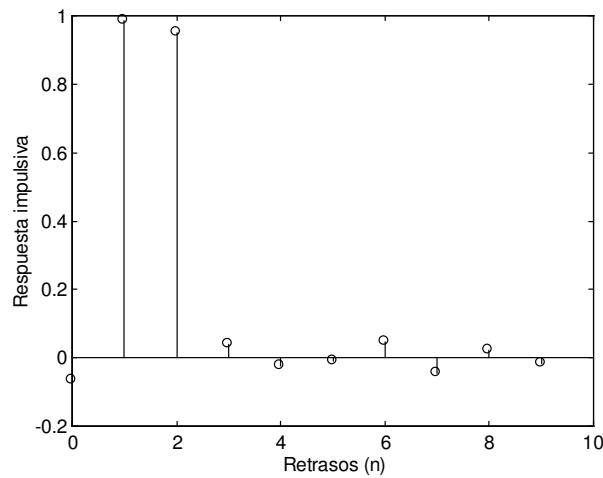


Figura 5.21 Respuesta impulsiva después de las iteraciones

Está claro que con esta estructura no es posible extraer la información del sistema. Es posible, sin embargo pensar en una estructura más elaborada que pudiese extraer dicha información de forma efectiva. La forma general de consiste en una serie de bloques L-N en paralelo, pero en el caso del sistema 5.21 bastaría con una estructura de tres bloques L-N tal como se muestra en la figura 5.22:

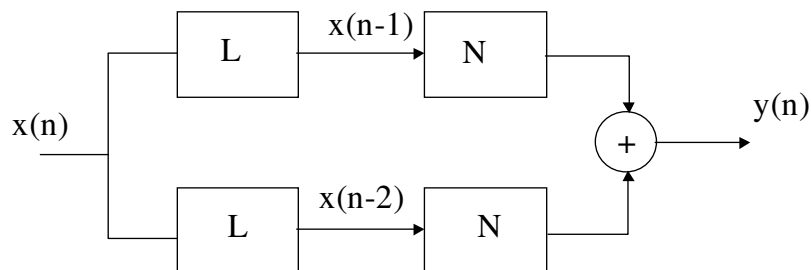


Figura 5.22 Estructura correcta para ajustar el sistema

Esta estructura tiene en cuenta la existencia de dos dinámicas diferentes, correspondientes al retraso uno y al retraso 2. Sin embargo, esta estructura sencilla no sería capaz de resolver una dependencia cruzada $x(n-1)x(n-2)$. Más aun, ¿Qué estructura se debería emplear si, como en el caso del ejemplo 4.29, asiduamente empleado, las dependencias no lineales presentan términos cuya dinámica no aparece en la parte lineal? Es evidente que en este caso, la estructura no puede ser en absoluto sencilla, sino que debe considerarse una arquitectura basada en una innumerable serie de bloques L-N dispuestos en cascada o paralelo. La selección de la estructura es por tanto crucial, pero basándose en el único conocimiento de las señales de entrada y salida no es posible establecer cual es la mejor estructura. El operador se ha de conformar con probar diversas arquitecturas Wiener, Hammerstein o L-N-L hasta encontrar un conglomerado de cajas que funcione bien. De hecho existen estrategias que permiten elucidar la topología, pero precisan del conocimiento de al menos los kernels de Wiener hasta el

orden tres, con lo cual el problema se vuelve prácticamente inabordable para esta caso⁸. Por supuesto esto no es tarea fácil e incluso, como en el caso del ejemplo 4.29, es inabordable. En cualquier caso, a continuación se presentan los resultados para el sistema 4.29 ajustado mediante una estructura como la de la figura 5.18,

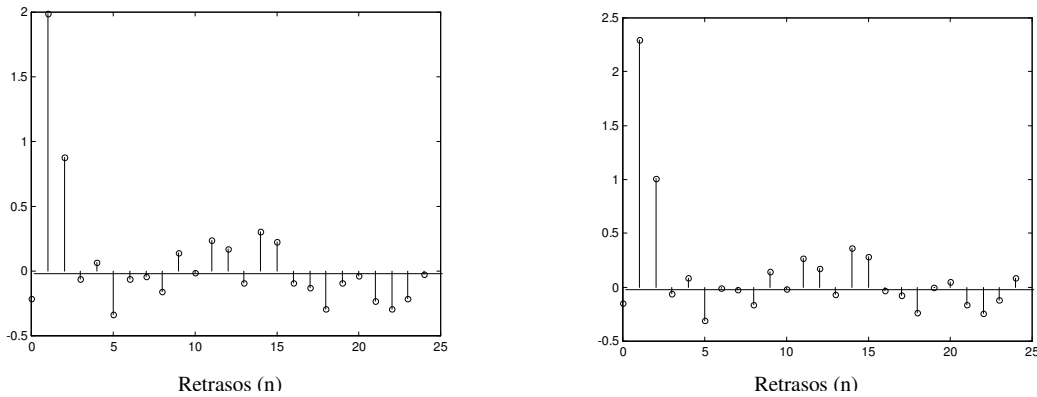


Figura 5. 23 Respuestas impulsivas para diferentes iteraciones

En esta figura se presentan las respuestas impulsivas calculadas en la primera iteración (gráfica izquierda) y en la última (gráfica derecha). Se observa que en ambos casos distingue cual es la dinámica lineal correcta, aunque los valores son mejores en el primer paso que en el último. Sin embargo, a medida que se realizan las iteraciones, el error disminuye hasta estabilizarse alrededor del 20% de PSNE (habiendo partido de valores alrededor de 30% de PSNE). Estos valores, son excesivamente altos teniendo en cuenta que se trata ya de una estimación no lineal. No obstante, el problema radica, como ya se ha demostrado, en la estructura. Una estructura como la de la figura 5.18 ya no puede presentar un rendimiento mejor, pero ¿cuál es la estructura adecuada?. Generar estructuras a ciegas no es una buena estrategia, y por otra parte, estructuras complejas precisan de gran cantidad de datos, por tanto, a falta de mas información, no se puede seguir avanzando por este camino.

5.9. Aplicación del modelado por cajas al sistema

Ajuste de datos y validación

Dejando definitivamente atrás el estudio del sistema SISO, se va a considerar directamente la estimación desde el punto de vista MISO. Se han utilizado para ello, los datos del experimento 3 y una aproximación universal de bloques LN en cascada propuesta por Korenberg. El método usado en la identificación es iterativo, calculando el mejor sistema de Wiener para el trío de datos (x_1, x_2, y) , siendo x_1 y x_2 las concentraciones de gas octano y tolueno y siendo y la salida de los sensores.

La estructura es la presentada en la figura 5.24:

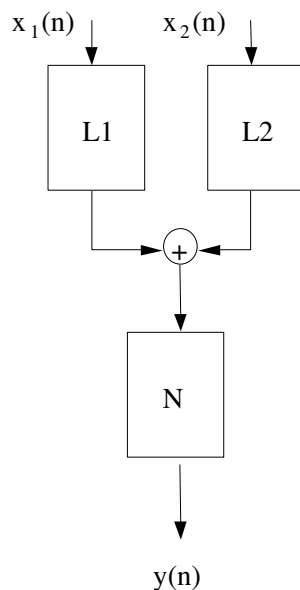


Figura 5. 24 Estructura Wiener de dos entradas, utilizada

A continuación se calcula el error cometido como $e = y - y_{\text{predicción}}$. Se identifica el mejor modelo de Wiener para el trío e, x_1, x_2 adicionándose de esta nueva contribución al modelo y recalculando de nuevo el error. Este proceso se repite hasta alcanzar un cierto límite de tolerancia preestablecido. La experiencia con los datos manejados demuestra que el descenso del error es muy lento, indicando este hecho que se necesitarían una gran cantidad de estructuras de bloques en paralelo. Con este método se alcanzan rendimientos similares a los métodos paramétricos lineales.

Tabla 5. 4 Resumen de resultados				
	Sensor 1	Sensor 2	Sensor 5	Sensor 6
PSNE estimación %	4.8 ± 0.2	3.6 ± 0.1	5.3 ± 0.2	3.8 ± 0.1
PSNE validación %	5.9 ± 0.5	5.3 ± 0.4	8.6 ± 0.7	6.8 ± 0.5

No resulta necesario realizar el análisis de residuos puesto que los resultados de la estimación no mejoran los ya obtenidos anteriormente.

Aceptación de resultados

Aunque evidentemente, el resultado es mucho más satisfactorio que el obtenido a través de los kernels calculados mediante correlación, el resultado no es mejor que las aproximaciones lineales realizadas a través del método paramétrico. Por este motivo, tampoco se realizarán los tests de análisis de residuos, que no aportarán mayor información que la que ya es evidente. De lo visto en los ejemplos sintéticos hay que deducir que la estructura empleada no es la óptima, pero tampoco es posible conocer cual es la mejor estructura posible. Aunque los valores de PSNE, si se comparan con los valores obtenidos en los ejemplos anteriores, son relativamente bajos, esto sólo se debe a que la no linealidad de los experimentos no es tan acusada como la expuesta en los mencionados ejemplos. La realidad es que este método, planteado de esta forma no es capaz de extraer toda la información, pero además tampoco hay una garantía que modificando la estructura se consigan mejores resultados. La única conclusión clara es

que el método de estimación por cajas puede funcionar bien si se acierta con la estructura adecuada, pero en este caso, la falta de información al respecto obliga a seguir investigando otros métodos no lineales.

5.10. Estimación mediante un método paramétrico no lineal: NARX

La estimación de los kernels de Wiener a través del método de Lee-Schetzen puede considerarse como una generalización del análisis de correlación presentado en el capítulo anterior. En el mismo sentido, puede desarrollarse un método paramétrico no lineal como generalización del método paramétrico también presentado en el capítulo anterior. Para ilustrar el procedimiento, supóngase un modelo de regresión lineal ARX

$$\hat{y}(n|\theta) = \varphi^T(x,y)\theta \quad (5.22)$$

donde $\varphi^T(x,y)$ es el vector que contiene las señales que describen al sistema:

$$\varphi(x,y) = [x(n-nk), x(n-nk-1), \dots, x(n-nk-nb+1), y(n-1), \dots, y(n-na)] \quad (5.23)$$

siendo nk el retraso puro entre la entrada y la salida y donde θ es el correspondiente vector de parámetros:

$$\theta = [b_1, \dots, b_{nb}, a_1, \dots, a_{na}] \quad (5.24)$$

Ya se ha descrito en el apartado 4.10 del capítulo anterior, la estimación del vector θ , que se realiza mediante algoritmos de mínimos cuadrados y que en el caso ARX estaba relacionada con correlaciones entre las señales del sistema.

Cuando el sistema resulta ser no lineal, debe modificarse el modelo lineal ARX por un modelo no lineal denominado NARX^{18 19 20}. Entonces, este modelo NARX vendrá descrito por alguna combinación no lineal a partir de las señales que describen al sistema. Esta combinación puede representarse en una expresión como la que sigue a continuación:

$$y(n) = F_r[x(n-d), x(n-nk-1), \dots, x(n-nk-nb+1), y(n-1), \dots, y(n-na)] \quad (5.25)$$

siendo F_r una función no lineal de orden de no linealidad r que afecta a las señales que actúan en el sistema.

Expandiendo la función F_r en forma de polinomio y redefiniendo las señales como:

$$\begin{aligned} V_1 &= x(n-nk), V_2 = x(n-nk-1) \dots V_{nx} = x(n-nk-nb+1), \\ V_{nx+1} &= y(n-1), \dots, V_s = y(n-na) \text{ con } s = nb + na \end{aligned} \quad (5.26)$$

entonces la expansión puede expresarse como:

$$y(n) = \sum_{i=1}^s \theta_i V_i + \sum_{i=1}^s \sum_{j=1}^s \theta_{ij} V_i V_j + \dots + \underbrace{\sum_{i=1}^s \dots \sum_{m=1}^s \sum_{n=1}^s \theta_{ij\dots mn} V_i V_j \dots V_m V_n}_{r \text{ veces}} \quad (5.27)$$

Donde la señal $y(n)$ se presenta como un desarrollo polinómico de grado enésimo de las funciones V_i . Evidentemente, la regresión lineal del modelo ARX es un caso particular de este modelo general y se corresponde con el primer sumatorio de la expresión anterior.

La definición del sistema no lineal en forma paramétrica queda pues bien clara, pero el objetivo ahora consiste en la estimación de cada uno de los parámetros θ de la expresión 5.23. Sería deseable conseguir un método de estimación similar al obtenido para el caso paramétrico lineal. Afortunadamente, es posible alcanzar este objetivo si se redefinen los productos de las señales $V_1 \cdots V_n$ como nuevas señales V_{s+k} : $V_{s+1} = V_1 V_1$, $V_{s+2} = V_1 V_2, \dots$

Entonces es posible definir la señal $y(n)$ según la expresión:

$$y(n) = \sum_{i=1}^S \theta_i V_i \quad (5.28)$$

que es cualitativamente igual a la empleada en los modelos ARX, pero donde V_i es un vector de señales que contiene tanto las señales con dependencia lineal como los productos de las mismas y siendo S la dimensión final del vector de señales. Hay que remarcar que a través de este desarrollo se consigue una expresión lineal en los coeficientes (a pesar de que V_i contiene productos no lineales de las señales del sistema) y que por tanto, es posible utilizar el cálculo de mínimos cuadrados descrito en los modelos paramétricos lineales para ajustar un vector de parámetros θ compuesto de parámetros que multiplican tanto a señales lineales como a no lineales.

5.11. Aplicación del método NARX a un problema sintético

Sea el sistema descrito por la ecuación 4.29 (que incluye una señal de ruido) al que se desea ajustar un modelo ARX no lineal (NARX). Resuelto de forma lineal tal y como se realizó en el capítulo 4, resultan dependencias lineales del tipo $x(n-1)$ y $x(n-2)$. Por tanto, estas dinámicas deben ser las primeras a introducir en el vector de funciones $\varphi(x,y)$. No es necesario buscar cuáles son las dependencias dinámicas no lineales puesto que la relación 4.29 las presenta y define totalmente. Dichas dinámicas no lineales son $x(n-2)*x(n-3)$, $x(n-4)*x(n-7)$ y $x(n-5)*x(n-5)$. De nuevo hay que remarcar que esta información dinámica no está a disposición en los problemas reales, y que esa falta de información dinámica es precisamente uno de los problemas fundamentales a resolver. Entonces, el vector $\varphi(x,y)$ que se debe utilizar para estimar un modelo NARX se define así:

$$\varphi(x,y) = [x(n-1) \ x(n-2) \ x(n-2)*x(n-3) \ x(n-4)*x(n-7) \ x(n-5)*x(n-5)]$$

Tal y como queda definido el vector de funciones, puede considerarse al sistema 4.29 como un sistema MISO de cinco entradas, tal como se representa en la figura 5.25:

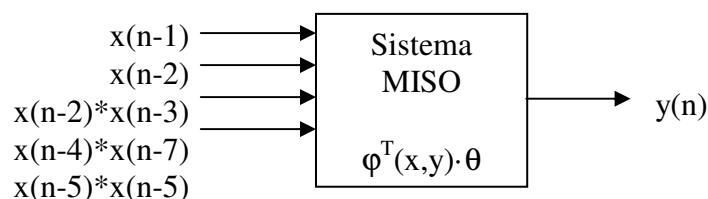


Figura 5. 25 Estructura MISO

Entonces, teniendo en cuenta que no ya no hay más dependencias cruzadas en este sistema la salida es el producto del vector de las entradas con un vector de parámetros θ que se puede estimar a través de los algoritmos ARX. En este caso, el vector de parámetros debería ser exactamente:

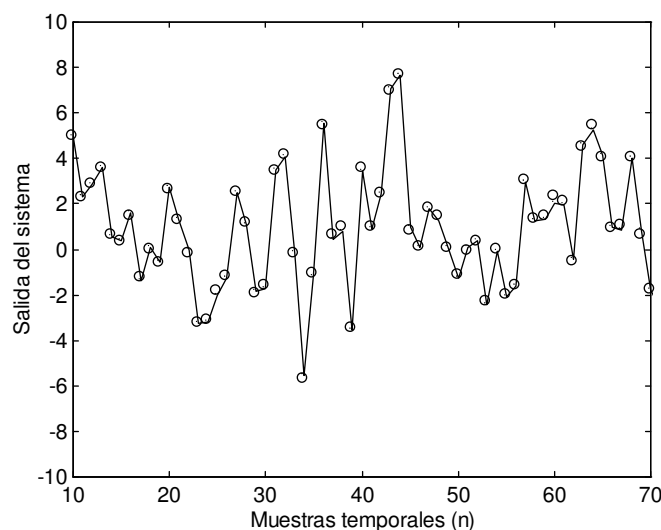
$$\theta = [2 \ 1 \ 0.5 \ 0.9 \ 0.5];$$

y la relación de este vector con las respuestas impulsivas es clara. Los dos primeros términos representan los valores de la respuesta impulsiva del sistema respecto de la señal $x(n)$ para retraso 1 y para retraso 2. El resto de valores del vector θ representan el valor de la respuesta impulsiva para unas supuestas funciones de entrada del sistema cuyas dinámicas son las mismas (según se definen) que la de la salida, y por tanto tienen retraso nulo.

Efectuadas las operaciones de estimación, los valores del vector θ estimado son los siguientes:

$$\theta = [1.9954 \ 0.9998 \ 0.5018 \ 0.9008 \ 0.5001]$$

La figura 5.26 muestra una parte de la estimación de la salida del sistema junto con la salida del sistema real para datos de validación. Los círculos representan la estimación mientras que la línea continua es la señal de salida $y(n)$.

**Figura 5. 26** Comparación entre la estimación y la señal real

Calculado el valor de la función PSNE, éste se sitúa siempre alrededor de 0.2% con datos de validación. Como puede verse, los resultados son excelentes, siempre teniendo en cuenta que la dinámica no lineal ya era conocida. Se va a comprobar a continuación si este buen rendimiento se ve refrendado cuando se aplica al sistema real del cual no se dispone de la información dinámica a priori.

5.12. Aplicación del método NARX al sistema

Ajuste de los datos y validación

Al igual que en el caso del modelado por cajas, se va a estudiar directamente el caso MISO a través del experimento 3. Como siempre, se utilizarán un conjunto de 300 datos para la estimación y otro conjunto de 80 datos para la validación. Aunque, a partir de lo expuesto hasta ahora, el método a seguir es claro, lo que no es tan claro a priori son las dinámicas no lineales que conforman al sistema. El problema es parecido al que ocurre en la estimación lineal, cuando no es conocida la respuesta impulsiva del sistema y se organiza un estudio para estimar mediante análisis de correlación los valores de los n primeros retrasos, donde n ha de ser suficientemente grande como para contener toda la dinámica relevante y suficientemente pequeño como para no realizar cálculos innecesarios. La información extraída de esta estimación se utiliza después en la estimación paramétrica.

Este proceso se ha de intentar repetir en el presente caso. Sin embargo, los resultados del análisis de correlación no lineal, aunque definen una cierta tendencia, no ofrecen información concluyente. Por tanto, dado que no es posible repetir con buena fiabilidad el proceso seguido en el caso lineal pero basándose en la tendencia mostrada por los kernels presentados en las figuras 5.13, 5.14 y 5.15 y en la información lineal se establece la siguiente hipótesis de trabajo:

- Si la estimación lineal no presenta relaciones dinámicas más allá del retraso k , parece lógico suponer que la parte no-lineal del sistema tampoco se verá influenciado por retrasos mayores que k .

Bajo esta hipótesis, el proceso que se ha seguido ha sido el de generar las diferentes combinaciones de retrasos y aceptarlas como válidas siempre que su inclusión en el sistema suponga una disminución significativa del valor de la figura PSNE.

El siguiente mapa muestra en cuadrícula las combinaciones ensayadas. Por otra parte, la inclusión de términos autoregresivos no proporciona mejoras al modelo.

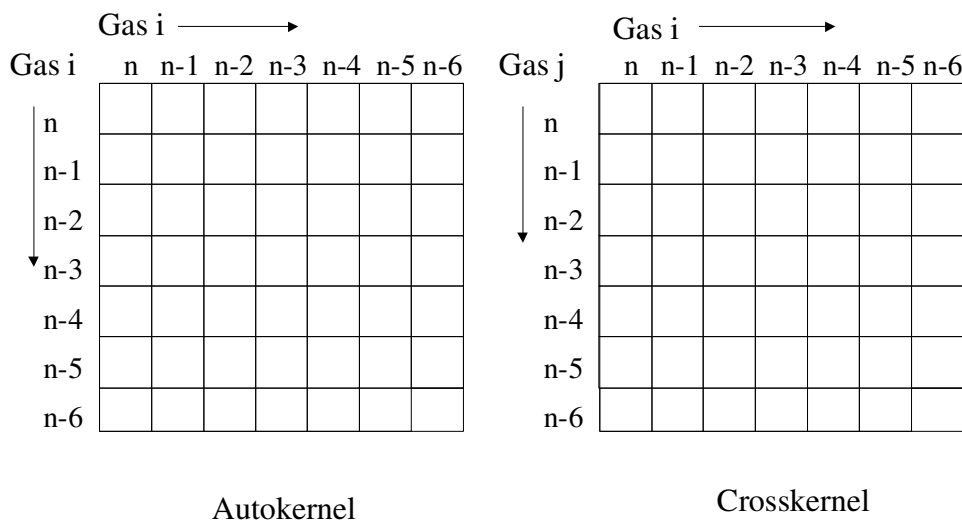


Figura 5. 27 Combinaciones estudiadas

Cada uno de los productos combinados presentes en estas cuadrículas ha sido incluida en la estimación del sistema siguiendo un procedimiento de cálculo por el cual se procede a la adición secuencial de las diferentes diagonales, comenzando por la principal. Siguiendo este proceso, se ha llegado a generar los kernels de segundo orden como el que se muestra en la figura 5.28. En ella, se representa además el kernel de segundo orden calculado a través del método de Lee-Schetzen. Ambas gráficas hacen referencia a resultados sobre el sensor 2.

Los resultados presentados tanto en las gráficas como en las tablas se han obtenido después de modificar los funcionales NARX para ajustarlos exactamente a la estructura de los kernels de Wiener, haciéndolos ortogonales entre sí y mejorando, en consecuencia, la calidad de la estimación. Esta modificación se realiza en base a la ecuación 5.13 (caso SISO) o 5.18 (caso MISO) donde se puede observar que, por ejemplo, el kernel de Volterra de segundo orden para un sistema SISO queda modificado de la siguiente forma:

$$\text{funcional de Volterra } H_2(x(n)) = \sum_i \sum_j h_2(i, j)x(n-i)x(n-j)$$

$$\text{funcional de Wiener } G_2(x(n)) = \sum_i \sum_j h_2(i, j)[x(n-i)x(n-j) - P]$$

Siendo P la densidad de potencia espectral del ruido blanco de la entrada.

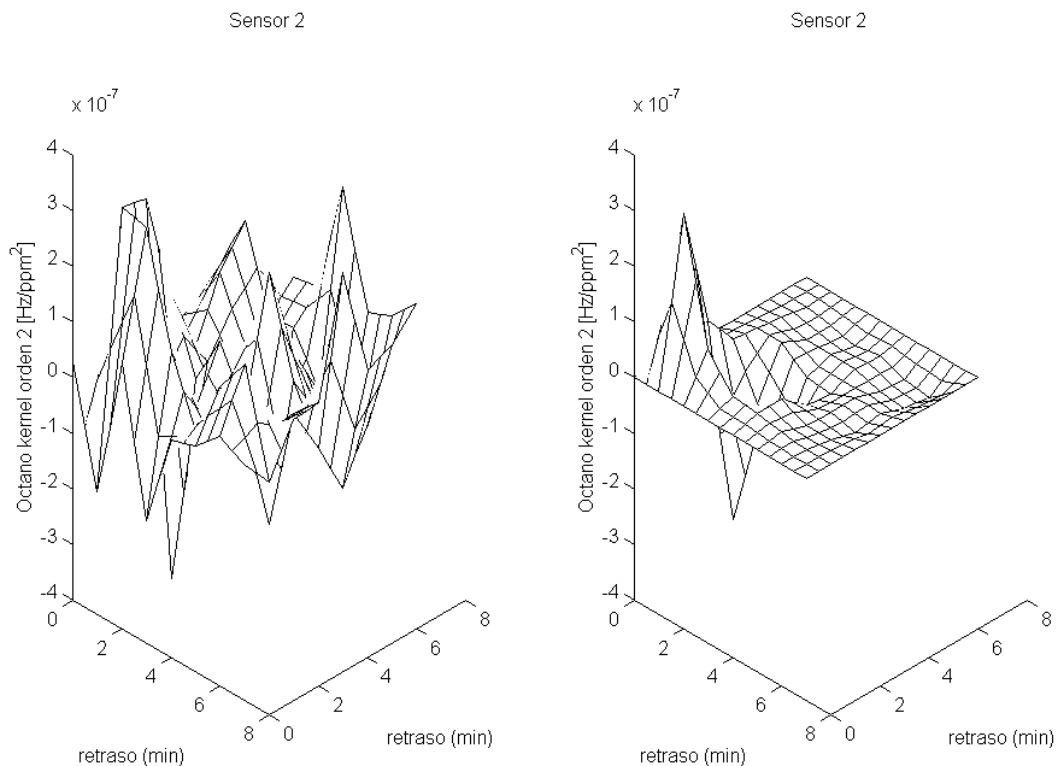


Figura 5.28 Comparación para el cálculo del kernel de 2º orden

Se observa de esta figura que para este caso, de todos los valores presentados en la retícula de autokernel de la figura 5.28, sólo tienen importancia, los correspondientes a retrasos bajos (en la retícula: zona superior izquierda). Similares conclusiones se pueden extraer para el resto de sensores.

El resultado, muestra que el método NARX es mucho más eficiente que el método Lee-Schetzen. Puede apreciarse claramente que los valores del kernel calculado a través del método NARX, presentan una varianza mucho menor que los valores del mismo kernel mediante correlación. De hecho, la estimación NARX, modificada para ajustarse a los funcionales ortogonales de Wiener, es, a semejanza de lo que ocurría en la estimación lineal, una estimación por mínimos cuadrados de los propios kernels de Wiener.

Hay que hacer notar que en referencia a relaciones cruzadas entre las entradas aparece un cierto crosskernel que contiene información valiosa para la estimación del sistema. Si este término cruzado entre las entradas fuese eliminado los valores presentados en la tabla 5.5 serían superiores.

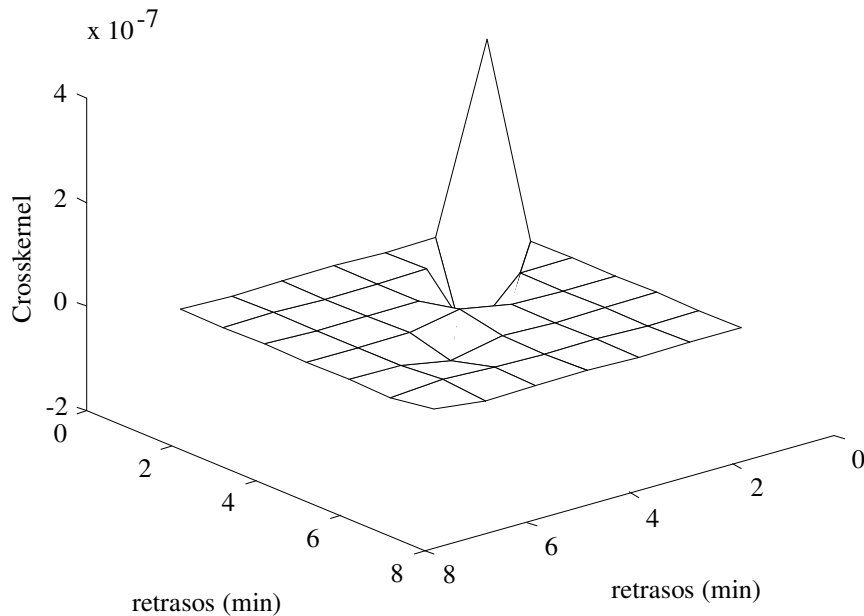


Figura 5.29 Crosskernel estimado

La tabla 5.5 muestra el tanto por ciento de señal no estimada.

Tabla 5.5 Resumen de resultados				
	Sensor 1	Sensor 2	Sensor 5	Sensor 6
Estimación (%)	0.7 ± 0.02	0.3 ± 0.01	0.4 ± 0.02	0.4 ± 0.02
Validación (%)	1.6 ± 0.1	0.6 ± 0.05	0.8 ± 0.06	0.8 ± 0.06

Además de la validación a través de la función PSNE también es necesario realizar un análisis de los residuos. A los tests usuales se han añadido dos más para tener en cuenta el hecho de disponer de dos entradas:

$$\begin{aligned} R_{(x_1 x_2) \epsilon} (m) &= 0 \\ R_{(x_1 x_2) \epsilon^2} (m) &= 0 \end{aligned} \quad (5.29)$$

Como no podía ser de otra forma, el análisis de residuos para esta estimación no lineal cumple con las condiciones preestablecidas. En la siguiente figura se presenta una comparación entre la estimación lineal y la no lineal del test $R_x^2 \varepsilon$ para el sensor 2.

El modelo lineal presenta unos claros picos de correlación con la entrada cuadrada (octano), indicando la no adecuación del modelo. Para el caso del modelo no lineal generado por los kernels de Wiener (estimados por mínimos cuadrados), se observa la desaparición de esta dependencia.

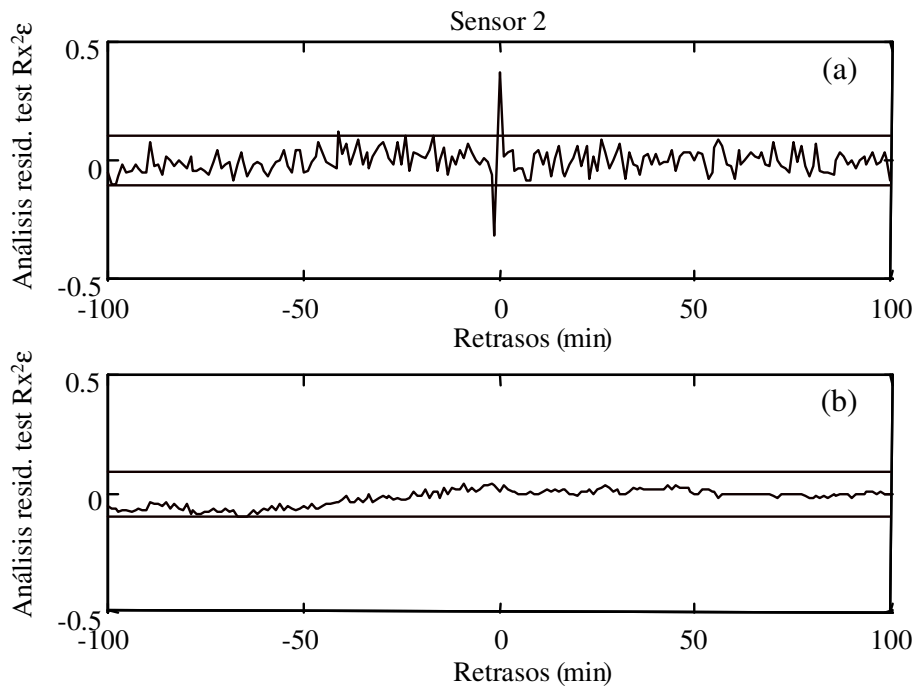


Figura 5.30 test de residuos para un modelo lineal (a) y un modelo no lineal (b)

Aceptación del modelo

Con este modelo no lineal, se llega al límite de resolución del experimento, por tanto, es absurdo ensayar grados de aproximación superiores. Un 1.6 % de error en la estimación de la señal de salida del sensor supone una falta de estimación de una concentración cercana a los 100 ppm que constituye el límite estimado del error que se utiliza en el experimento 3. El modelo puede darse como válido ya que llega al límite de la explotación de la información del sistema: por un lado el valor de PSNE es tan bajo como permiten los datos experimentales y por otro lado los análisis de residuos probados indican que el residuo no contiene información relevante. Obsérvese, sin embargo, que para llegar a este límite, se ha introducido una cierta relación entre las dos señales de entrada. Esto quiere decir que, el sistema no responde de forma igual ante variaciones exclusivas de uno de los gases que ante variaciones de ambos gases.

5.13. Estimación mediante redes neuronales

Aunque los modelos NARX ha resultado ser una vía útil de estimación, se va a investigar también el uso de redes neuronales como opción alternativa^{33 34}. Los modelos NARX ofrecen resultados excelentes y el proceso de cálculo, una vez establecido el

conjunto de funciones, es inmediato. Sin embargo, el proceso para llegar a establecer el mencionado conjunto no es directo a priori. No existe una formulación ni una estrategia bien definida que indique directamente el conjunto de funciones a probar. Establecer dicho conjunto es una tarea que requiere un cierto tiempo y una serie de ensayos para establecer la aceptación o el rechazo. Mediante redes neuronales se trata entonces de comprobar si son capaces de solventar este problema y ofrecer, a la vez resultados similares a los modelos NARX. En el capítulo uno ya se estableció la capacidad de una red de perceptrones multicapa con una capa escondida para aproximar cualquier función continua^{7 21 2}. Puede demostrarse también que la salida de una red neuronal cuyas entradas están formadas por las señales retrasadas de las entradas de un sistema ('Time-delay Neural Network'), puede expresarse como una serie de Volterra infinita⁷. La equivalencia entre el método de Volterra y las redes neuronales queda así bien demostrada. Abundando en este sentido, las redes neuronales no están limitadas a un orden finito de kernels y, por tanto son capaces de representar altos grados de no linealidad, siempre que se dispongan del número suficiente de capas escondidas. Así, la habilidad de las redes neuronales para aproximar funciones no lineales las hace especialmente aptas para su uso en estudios de identificación dinámicos.

En el capítulo uno, se introdujeron las redes neuronales desde el punto de vista de la conexión entre una detección natural y una aproximación matemática a dicha detección. Allí se describió brevemente como funcionaban las neuronas que conforman a una red neuronal así como se presentaron las arquitecturas que se utilizan habitualmente. Ahora, se pretende aplicar esos aspectos a la detección cuantitativa no lineal.

5.14. Aplicación de las redes neuronales a un problema sintético

Para intentar desarrollar una red neuronal ajustada al sistema 4.29 se ha desarrollado una arquitectura basada en una capa escondida de n neuronas no lineales y una capa de salida de una neurona lineal. El número de neuronas de la capa escondida es variable y en función de él, los resultados son mejores o peores, en el sentido de la función PSNE. Puesto que el sistema está perfectamente definido por la ecuación 4.29 y teniendo en cuenta, como siempre, que esta información no es directamente accesible a priori, las entradas que se han de utilizar en esta red son evidentes: la entrada de la red ha de estar constituida por las señales $(n-1)$, $x(n-2)$, $x(n-3)$, $x(n-4)$, $x(n-5)$ y $x(n-7)$. Al ser la red una estructura intrínsecamente no lineal, no es necesario utilizar como entrada las señales producto definidas en la ecuación 4.29 ($x(n-2)*x(n-3)$, $x(n-5)*x(n-5)$ o $x(n-4)*x(n-7)$) ya que la propia arquitectura y las propias neuronas no lineales se encargarán de aproximar estas no linealidades sin necesidad de introducirlas explícitamente. Para ello sólo será necesario disponer del número suficiente de neuronas. En este sentido se han realizado diferentes ensayos con diferente número de neuronas en la capa escondida. Pero antes de mostrar estos resultados se deben discutir algunos aspectos interesantes del proceso de entrenamiento de la red. En primer lugar hay que, de nuevo, volver a remarcar el papel de los datos de validación en estos procesos. Una arquitectura neuronal creciente con el suficiente número de datos y que trabaje sólo con datos de estimación siempre mejorará su rendimiento a medida que su estructura aumente (de la misma forma que los modelos paramétricos también mejoran en la zona de estimación si aumenta el número de parámetros). Nótese que se ha hecho la precisión en el *número de datos suficiente*. El motivo de dicha precisión se debe al hecho de que el número de parámetros que se deben estimar crece rápidamente a medida que la estructura de la red

neuronal aumenta y, si no se dispone de una cantidad de datos de entrenamiento suficiente, no es posible determinar con buena exactitud los valores de los pesos correspondientes a las neuronas de la estructura. Los siguientes resultados tratan de demostrar todos estos efectos. La red que se utiliza es una red de perceptrones multicapa y el algoritmo de entrenamiento es la versión de Levenberg-Mardquart, del algoritmo back-propagation²⁵.

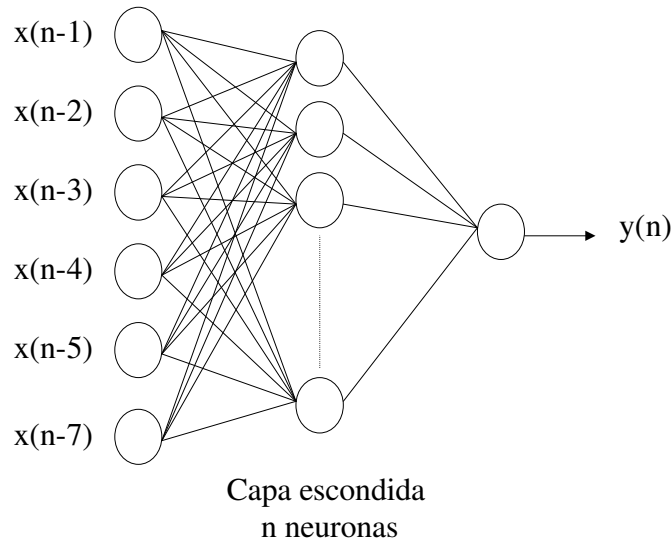


Figura 5.31 Estructura neuronal típica

Tabla 5.6 PSNE en función del número de neuronas de la capa escondida

Neuronas	PSNE Estimación (%)	PSNE Validación (%)
1	22.0 ± 0.9	25.1 ± 2.0
2	15.1 ± 0.6	21.1 ± 1.7
4	5.9 ± 0.2	8.6 ± 0.7
5	1.8 ± 0.07	3.0 ± 0.2
8	0.2 ± 0.00	0.4 ± 0.03
10	0.2 ± 0.00	0.5 ± 0.04
15	0.1 ± 0.00	0.6 ± 0.05
20	0.1 ± 0.00	2.0 ± 0.2
30	0.07 ± 0.00	17.5 ± 1.4

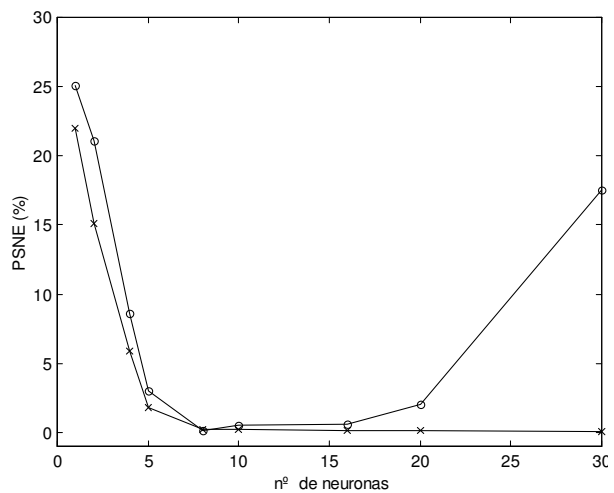


Figura 5.32 valor de PSNE en función del nº de neuronas
x - Estimación
O - Validación

En la tabla 5.7 y el la figura 5.31 se demuestra que la estimación siempre mejora a medida que aumenta la estructura de la red. Sin embargo, para la misma cantidad de puntos, la validación comienza a empeorar a partir de un cierto nivel de complejidad de

la arquitectura neuronal. En el ejemplo presentado se han utilizado 300 puntos. Si se aumenta el número de puntos, los resultados de las estructuras grandes mejoran.

Las estructuras de perceptrones multicapa presentan la peculiaridad de generar una hypersuperficie de error con diversos mínimos locales. Así, los diferentes entrenamientos no tienen porqué converger hacia el mismo mínimo global. El mínimo local al cual converge el algoritmo depende de la inicialización de los parámetros de la red. Por lo tanto, un entrenamiento llevado a cabo con los mismos datos pero dejando la inicialización de los parámetros al azar proporciona diferentes resultados. Este fenómeno es engorroso ya que deben realizarse varios entrenamientos si se desea comprobar si el resultado es bueno. Existen, no obstante, diferentes técnicas para minimizar este efecto. Por ejemplo, un aumento de la estructura de la red hace disminuir la diferencia entre los diferentes mínimos locales, aunque este aumento de estructura está condicionado por la cantidad de datos disponibles, como ya se ha mencionado.

5.15. Aplicación de las redes neuronales a la estimación del sistema

Se pretende analizar el problema a través de una red neuronal multicapa de perceptrones no lineales. El modelo que se pretende implementar está dado por la siguiente expresión:

$$y_p(k+1) = f(y_p(k), \dots, y_p(k-r), x_t(k), \dots, x_t(k-s), x_o(k), \dots, x_o(k-m)) \quad (5.30)$$

Es decir, la salida $y(k+1)$ es una función de salidas retrasada en el tiempo hasta un valor r , y función de las entradas de tolueno y octano (x_t , x_o) con retrasos hasta s y m muestras respectivamente. La arquitectura presenta una capa de entrada con las entradas del sistema, una capa escondida, directamente asociada a la capa de entrada, con un número variable de neuronas y una última capa de salida, de una sola neurona. La función de activación de cada neurona de la capa escondida es, por tratarse de un modelo no lineal, una función no lineal del tipo sigmoide, mientras que la neurona de salida presenta una función de activación lineal.

La figura 5.33 presenta un esquema de la arquitectura de la red a implementar.

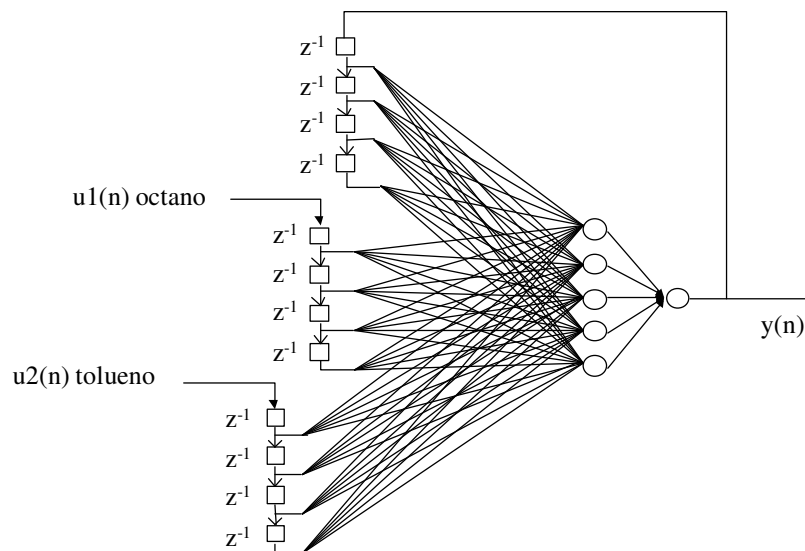


Figura 5.33 Arquitectura de la red multicapa utilizada

El proceso de entreno de la red, por ser una red multicapa de perceptrones no lineales sigue un algoritmo basado en la regla backpropagation.

Ajuste de los datos y validación

Tal y como se viene realizando en los últimos estudios no lineales presentados, se va a utilizar la red para analizar directamente el problema de la estimación MISO. Para ello se utilizan, como siempre, los datos del experimento 3.

Realizadas una serie de ensayos para determinar definitivamente la arquitectura, con los datos disponibles, los mejores resultados aparecen utilizando cinco neuronas en la capa escondida y seis retrasos en las entradas. La tabla 5.7 muestra los valores obtenidos con una red neuronal multicapa.

Tabla 5.7 Resultados de la estimación con redes neuronales				
	Sensor 1	Sensor 2	Sensor 5	Sensor 6
PSNE Estimación (%)	0.8 ± 0.03	0.5 ± 0.02	0.4 ± 0.01	0.4 ± 0.01
PNSE Validación (%)	1.3 ± 0.1	0.8 ± 0.06	0.8 ± 0.06	0.9 ± 0.07

La figura presenta el ajuste realizado para el sensor 2 comparado con los datos reales y el error.

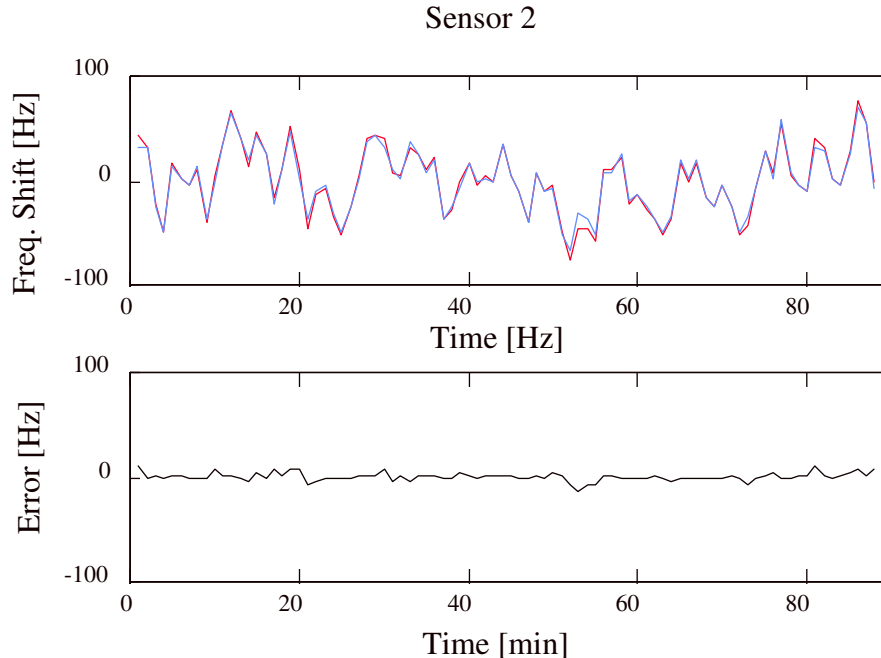


Figura 5.34 Resultados para el sensor 3 con red neuronal no lineal

No es necesario presentar los resultados de los tests de análisis de residuos ya que son similares a los obtenidos con modelos NARX. Los análisis de residuos realizados dan a entender que no queda información no lineal por desvelar o tal información es inferior, en potencia, a la potencia del ruido aleatorio general.

Aceptación de los datos

De la comparación entre las diferentes tablas de resultados, se puede comprobar que el rendimiento de la estimación a través de las redes neuronales es similar al obtenido con los kernels de Wiener estimados a través de mínimos cuadrados (modelos NARX). Sin embargo, a lo largo del proceso de estimación con las redes neuronales se ha podido constatar que cada entrenamiento de la red neuronal no lineal ofrece resultados ligeramente diferentes. Este hecho es indicio de la existencia de mínimos locales diferentes y hace que el uso de redes neuronales para realizar estimaciones sobre estos datos sea tedioso ya que el entrenamiento se debe ejecutar varias veces para asegurar que el resultado obtenido es una buena representación del sistema sensor. Además, si los modelos NARX presentaban un problema en la elección de las funciones, en el caso de las redes, aparece el mismo problema asociado a la elección de la arquitectura. Así, aunque las prestaciones de las redes sean excelentes, todo el proceso de trabajo hace que, para el caso estudiado, sea preferible utilizar el modelo NARX. Por otra parte, no es necesario realizar el análisis de residuos puesto que los resultados son similares a los ya expuestos en el caso NARX.

5.16. Resumen de los resultados y conclusiones

La siguiente tabla resume los resultados para el sistema MISO de los diferentes métodos que se han ido probando a lo largo de esta exposición:

Tabla 5.8 Resumen de resultados				
Método	Sensor 1	Sensor 2	Sensor 5	Sensor 6
Método paramétrico (ARX)	5.0% est 7.0% val	3.7% est 5.5% val	5.3% est 9.2% val	3.9% est 7.3% val
Análisis lineal de correlación	16.9% est 31.9% val	20.2% est 34.8% val	15.8% est 28.2% val	18.4% est 35.6% val
Kernel no lineal (Lee-Schetzen)	79.3% est 53.6% val	86.9% est 55.8% val	76.7% est 54.7% val	84.7% est 56.8% val
Kernel de Volterra no lineal (NARX)	1.1% est 5.6% val	0.4% est 3.3% val	0.6% est 2.2% val	0.6% est 2.3% val
Kernel de Wiener no lineal (NARX)	0.7% est 1.6% val	0.3% est 0.6% val	0.4% est 0.8% val	0.4% est 0.8% val
Modelado por cajas LN	4.8% est 5.9% val	3.6% est 5.3% val	5.3% est 8.6% val	3.8% est 6.8% val
Redes neuronales	0.8% est 1.3% val	0.5% est 0.8% val	0.4% est 0.8% val	0.4% est 0.9% est

De estos resultados y de todo lo expuesto tanto en el capítulo tres, donde se realizaba una estimación estática, como en el capítulo anterior donde se realizaban estimaciones

lineales dinámicas, como en el presente capítulo dedicado a estimaciones no lineales dinámicas, pueden extraerse varias conclusiones.

Ante el problema de estimar una función que relacione la entrada y la salida del sistema de sensores de gas se han aplicado diferentes técnicas de estimación estática, lineal y no lineal, haciendo énfasis en el hecho de utilizar un reducido número de datos para la confección de dicha estimación. Esta condición resulta esencial ya que implica la búsqueda de métodos eficientes y robustos ante una cierta falta de información. Se ha comprobado que los métodos de correlación, a pesar de su simplicidad y fácil aplicación, no son útiles frente al problema de la escasez de datos, ofreciendo rendimientos pobres sobre todo a medida que la relación entre el número de datos y el número de cálculos a realizar disminuye. Esto se hace muy evidente para el planteamiento de sistemas MISO, donde el número de cálculos a realizar es elevado y donde se aprecia que el rendimiento de la estimación por correlación es pésimo.

De entre todos los métodos probados la estimación NARX mediante kernels modificados y las redes neuronales son los que han mostrado un mejor rendimiento. Los modelos NARX ofrecen excelentes resultados aunque el proceso para encontrar las funciones que configuran al sistema puede resultar lento y difícil. Por otra parte, las redes neuronales ofrecen resultados comparables a los modelos NARX, pero configurar su arquitectura es tan lento como determinar el conjunto de funciones a usar en el modelo paramétrico no lineal. Además, mientras que el cálculo no lineal paramétrico es inmediato y en un solo paso, obtener resultados con las redes es más lento y tedioso debido al necesario paso del entrenamiento. A esto hay que añadir que las redes presentan ciertos problemas de convergencia asociado a la no linealidad. Se ha podido comprobar que diferentes entrenamientos de la misma red no producen el mismo resultado, ya que la red suele quedar atrapada en un mínimo local.

Por otra parte, el resto de métodos, como los modelos de cajas no han mostrado su teórica eficiencia, probablemente debido a la difícil selección de topologías adecuadas.

En cuanto a las dependencias y relaciones entre las señales del sistema, se ha comprobado la relación no lineal existente en el sistema. Y se ha llegado a explotar al máximo la información contenida en el sistema gracias al modelo NARX modificado y a las redes neuronales. Por otra parte, se ha evidenciado la existencia de cierta relación cruzada entre las entradas de gas para el caso de experimento MISO. El crosskernel determinado mediante la estimación NARX así lo certifica. Este hecho es uno de los motivos que explica porque en el caso lineal las respuestas impulsivas de orden uno eran cuantitativamente diferentes en el caso MISO y en el SISO, explicación ésta que había sido postpuesta hasta el presente capítulo .

Como conclusión de carácter general cabe apuntar la posibilidad real de obtener buenos modelos dinámicos no lineales, capaces de explotar hasta el límite la información del sistema y útiles para su aplicación a tareas de simulación y/o control, por ejemplo.

Bibliografía

- ¹L. Ljung, *System identification. Theory for the user* (Prentice Hall, New Jersey, 1987).
- ²S. Marco, A. Pardo, F. Davide, C. DiNatale, A. D'Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, U. Weimar, W. Göpel, "Different strategies for identification of gas sensing systems," *Sensors and Actuators B* **34**, 213-223 (1996).
- ³F. Davide, C. DiNatale, A. D'Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, W. Göpel, S. Marco, A. Pardo. "Dynamic calibration of QMB polymer-coated sensors by Wiener kernel estimation," *Sensors and Actuators B* **26-27**, 275-285 (1995).
- ⁴X. Liu D.T. Pham, *Neural Networks for identification, prediction and control* (Springer-Verlag, London, 1995).
- ⁵C.S. George Lee Chin-Teng Lin, *Neural Fuzzy systems* (Prentice Hall, New Jersey, 1996).
- ⁶Martin Schetzen, *The Volterra and Wiener theories of nonlinear systems* (Jhon Wiley & Sons, 1980).
- ⁷S. Haykin, *Neural Networks, a comprehensive foundation* (IEEE Press, 1994).
- ⁸H.W. Chen, "Modeling and Identification of Parallel Nonlinear Systems: Structural Classification and Parameter Estimation Method," *Proceedings of the IEEE* **83**, 39-83 (1995).
- ⁹M. Schetzen Y.W. Lee, "Measurement of Wiener kernels of a non-linear system by using cross-correlation," *Int. J. Control* **2**, 237-254 (1965).
- ¹⁰P.Z. Marmarelis, Ken-Ichi Naka, "Identification of Multi-Input Biological Systems", *IEEE Transactions on biomedical engineering* 88-101 (1974)
- ¹¹W.C. Krenz Y. Goussard, L. Stark, "An Improvement of the Lee and Schetzen Cross-Correlation Method," *IEEE Transactions on automatic control* **AC-30 n° 9**, 895-898 (1985).
- ¹²C.D. Geisler R.E. Wickesberg, "Artifacts in Wiener Kernels Estimates Using Gaussian White Noise," *IEEE Transactions on biomedical Engineering* **BME-31, n° 6**, 454-461 (1984).
- ¹³W. Kranz L. Stark, "Interpretation of Functional Series Expansion", *Annals of Biomedical Engineering*, **19**, 485-508 (1991)
- ¹⁴M.J. Korenberg, "Parallel cascade identification and kernel estimation for nonlinear systems," *Annals Biomedical Engineering* **19**, 429-455 (1991).
- ¹⁵L. D. Jacobson H-W. Chen, J.P. Gaska, "Structural Classification of multi-input nonlinear systems," *Biological Cybernetics* **63**, 341-357 (1990).
- ¹⁶M.J. Korenberg I.W. Hunter. "The Identification of Nonlinear Biological Systems: LNL Cascade Models" *Biological Cybernetics*, **55**, 125-134 (1986)
- ¹⁷I.W. Hunter M.J. Korenberg. "The Identification of Nonlinear Biological Systems: Wiener and Hammerstein Cascade Models" *Biological Cybernetics*, **55**, 135-144 (1986)
- ¹⁸V. Piuri C. Alippi, "Experimental Neural Networks for Prediction and Identification," *IEEE Transactions on instrumentation and measurement* **45 n° 2**, 670-676 (1996).
- ¹⁹R. Haber, "Parametric Identification of nonlinear dynamic systems based on nonlinear crosscorrelation functions," *IEE proceedings* **135 n° 6**, 405-420 (1988).
- ²⁰S.A. Billings S. Chen, "Representations of non-linear systems: The NARMAX model," *Int. J. Control* **49 n° 3**, 1013-1032 (1989).
- ²¹M. Beale H. Demuth, *Neural Network toolbox user's Guide* (The Mathworks Inc, 1994).

- ²² E. Eskinat, S.H. Johnson, W.L. Luyben, "Use of Hammerstein Models in Identification of Nonlinear Systems" *AICHe Journal*, **Vol 37, No 2**, 255-268 (1991)
- ²³ P. Gallman, "A comparison of two Hammerstein Model Identification Algorithms", *IEEE Trans Aut. Control*, 20, 771 (1976)
- ²⁴ D.T. Westwick, R.E. Kearney, "A new algorithm for the identification of multiple input Wiener system" *Biological Cybernetics* 68, 75-85 (1992)
- ²⁵ D. Mardquardt, *J. Soc. Int. Appl. Math.*, 2, 431, (1963)
- ²⁶ D. H. Johnson, "Applicability of white noise non linear system analysis to the peripheral auditory system" *J. Acoust. Soc. Am.* 68, 876-884 (1980)
- ²⁷ Y. Goussard, W. C. Krenz, L. Stark, "An improvement of the Lee Schtzen corss correlation method", *IEEE Trans. On Automatic Control*, 30, 895-898, (1985)
- ²⁸ Y. Shi, K. E. Hecox, "Non linear system identification by M-pulsed secuencias: application to brainstem auditory evoked responses", *IEEE Trans. On biomedical engineering*, 38, 834-845, (1991)
- ²⁹ V. Z. Marmarelis, "Identification of non linear biological systems using Laguerre expansions of kernels", *Ann. Of biomedical engineering*, 21, 573-589, (1993)
- ³⁰ J. D. Victor "Asimptotic approach of generalized orthogonal functional expansions to Wiener kernels" *Ann of biomedical engineering*, 19, 383-399, (1991)
- ³¹ J. Shi, H. H. Sun, "Non linear system identification for cascade block model: an application to electrode polarization impedance" *IEEE Trans. On Biomedical engineering*, 37, 574-587, (1990)
- ³² D. T. Westwick, R. E. Kearney, "a new algorithm for the identification of multiple input Wiener systems", *Biological Cibernetics*, 68, 75-85, (1992)
- ³³ S. Chen, S. A. Billings, P. A. Grant, "Non linear system identification using neural networks", *Int. J. Control*, 51, 1191-1214, (1990)
- ³⁴ K. S. Narendra, K. Parthasarathy, "identification and Control of dynamical system using neural networks", *IEEE Trans. On neural networks*, 1, 4-26, (1990)
- ³⁵ S.A. Billings, "Identification of non linear systems" in "Non linear System Design" Ed. S. A. Billings, J. O. Gray, D. H. Owens, Peter Peregrinus 31-45 (1984)
- ³⁶ D. H. Johnson, "Applicability of white noise non linear system analysis to the peripheral auditory system" *J. Acoust. Soc. Am.* 68, 876-884 (1980)
- ³⁷ Y. Goussard, W. C. Krenz, L. Stark, "An improvement of the Lee Schtzen corss correlation method", *IEEE Trans. On Automatic Control*, 30, 895-898, (1985)

Capítulo 6

Métodos de identificación dinámica: Modelado inverso

- *¡Viviendo marcha atrás! – repitió Alicia con gran asombro -. ¡Nunca he oído una cosa semejante!*
- *Pero tiene una gran ventaja y es que así la memoria funciona en ambos sentidos.*
- *Estoy segura de que la mía no funciona más que en uno – observó Alicia -. No puedo acordarme de nada que no haya sucedido antes.*
- *Mala memoria la que sólo funciona hacia atrás – censuró la Reina*
- *¿De qué clase de cosas se acuerda usted mejor? – se atrevió a preguntarle Alicia.*
- *¡Oh! De las cosas que sucedieron dentro de dos semanas...*
 - Lewis Carroll: Alicia a través del espejo

- *Por el contrario, en el tiempo real, hay una diferencia muy grande entre las direcciones hacia delante y hacia atrás, como todos sabemos. ¿De donde proviene esta diferencia entre el pasado y el futuro? ¿Por qué recordamos el pasado pero no el futuro?*

Stephen W. Hawking: Historia del tiempo

Prólogo del capítulo

Hasta este punto del trabajo se han planteado diferentes estrategias para la identificación de sistemas de sensores de gases desde el punto de vista cuantitativo. Los resultados han sido, tal y como se ha visto, eficientes en mayor o menor medida dependiendo de la estrategia identificativa utilizada. El paso posterior y lógico a la identificación directa, (desde la entrada hacia la salida) es el modelado inverso (desde la salida hacia la entrada). Si el resultado de esta modelización es satisfactorio, el sistema de sensores de gas se comporta de hecho como un analizador cuantitativo de concentraciones de gas en una mezcla multicomponente.

Puede pensarse que, habiéndose obtenido un modelo directo, extraer a partir de él un modelo inverso es una tarea simplemente de rutina. Sin embargo, este es un ejemplo típico de problema mal condicionado, en el cual, pequeños errores provocan grandes desviaciones en la reconstrucción de la señal⁸. Debido a este problema, la cuestión de la modelización inversa se va a realizar con el mismo espíritu que la modelización directa.

Al igual que en la identificación directa, el análisis cuantitativo general inverso de mezclas gaseosas con sensores químicos de bajo costo ha de hacer frente a los problemas originados por la falta de selectividad de los mismos. Estos defectos deben compensarse con la implementación inteligente de sensores con diferentes sensibilidades para cada gas constituyente de la mezcla.

En general, cuando se habla de modelado inverso suele hacerse referencia al modelado estático y, en este sentido, puesto que el sistema suele contener más sensores que gases en la mezcla, el sistema de ecuaciones a plantear está sobredeterminado. Afortunadamente esto no es un problema grave, existen muchas y muy diversas estrategias matemáticas capaces de encontrar soluciones para este problema, como la regresión por mínimos cuadrados o las propias redes neuronales.

Si se desea ir más allá, buscando fórmulas para realizar modelado inverso dinámico que permita contemplar cambios en las mezclas con la misma escala temporal que la dinámica del sistema, deberán generarse aplicaciones que permitan sortear problemas como las no linealidades estáticas y dinámicas o las derivas temporales.

Es obvio que un sistema de modelado inverso cuantitativo tiene múltiples aplicaciones en el campo de medición y en el de control¹³. A pesar de la abundante literatura en estos temas y de las profundas implicaciones que supone para los mismos, no se encuentran demasiadas aplicaciones prácticas. Esta parte del trabajo pretende ser, sin desdeñar los aspectos teóricos, una aplicación práctica concreta del modelado inverso aplicada a la medición gaseosa dinámica.

6.1. El problema inverso

Como ya se ha visto, el sistema de test de sensores descrito en el capítulo dos se puede describir como un sistema multientrada multisalida MIMO, donde las entradas son las concentraciones $\mathbf{x} = (x_1(t), x_2(t), \dots, x_n(t))$ de los gases a estudiar y donde las salidas son las respuestas de los sensores $\mathbf{y} = (y_1(t), y_2(t), \dots, y_n(t))$. En los capítulos anteriores se ha conseguido generar una serie de modelos capaces de transformar las señales de entrada de concentración \mathbf{x} en señales de respuesta de los sensores \mathbf{y} .

El objetivo del problema inverso planteado para este problema es conseguir, a partir de las señales $\mathbf{x}(n)$ e $\mathbf{y}(n)$, un sistema de medida, es decir, un procedimiento de reconstrucción de las señales de entrada de concentración utilizando las señales de

salida de los sensores. En este sentido, debe quedar claro que la concepción de sistema que se va a utilizar es diferente al modelado directo. Desde el punto de vista del modelado directo, el sistema MIMO de test se concibe como la reunión de diferentes sistemas MISO, uno por cada sensor. Si los sensores fuesen lineales y bajo condiciones estáticas, el sistema MIMO puede describirse mediante la ecuación:

$$y=A \cdot x \quad (6.1)$$

donde la matriz A contiene los modelos directos MISO de los sensores. Realizar la inversión de este sistema consiste, teóricamente, en hallar una matriz de transferencia inversa A^{-1} . Si los sensores fuesen perfectamente selectivos, la matriz de transformación sería estrictamente una matriz diagonal. Desgraciadamente, los sensores reales distan mucho de ser perfectamente selectivos. Aun más, normalmente el número de sensores dispuestos en el sistema es superior al número de gases presentes en la mezcla a analizar. Ocurre entonces que aparece un exceso de información, el operador A no es una matriz cuadrada y la solución de la inversión debería buscarse en el sentido de mínimos cuadrados:

$$x=(A^T A)^{-1} A^T y \quad (6.2)$$

Todo este proceso para determinar el sistema inverso no sólo es muy sensible al ruido y a las distorsiones de las señales de los sensores sino que puede perfectamente resultar mal condicionado a causa de las correlaciones entre las respuestas de los diferentes sensores. Por si fuera poco, la falta de selectividad de los sensores, su no linealidad y el hecho de medir en condiciones no estáticas imposibilitan la aplicación de la inversión del sistema a partir de la ecuación 6.1. En general, los trabajos en el campo de la inversión se centran en encontrar modelos no lineales multivariantes de los sensores desde el punto de vista estático⁸. Habitualmente, el exceso de información se resuelve a través del cálculo con algoritmos de minimización, aunque también es corriente usar redes multicapa de perceptrones para obtener las soluciones inversas.

En el presente trabajo, se va a proceder, de forma similar al método usado en sentido directo. Como ya se ha visto, en el sentido directo se generan modelos que se validan según una señal de error procedente de la diferencia entre la salida del modelo y la respuesta real de los sensores. El diagrama 6.1 representa la operación en sentido inverso.

Así, se trata de identificar un modelo inverso y aplicar modificaciones sobre el mismo en base a un criterio de error definido en función de la diferencia entre las señales reales y las definidas a través del modelo. Esto es posible gracias a que la definición de las entradas está bien determinada mediante el sistema de control de los MFC para todos y cada uno de los experimentos diseñados. Los experimentos utilizados usan dos gases: octano y tolueno y se toman las respuestas de cuatro sensores. Por tanto, bajo el punto de vista del modelado inverso, siguiendo la nomenclatura utilizada en el estudio directo, el sistema será MIMO de cuatro entradas (las respuestas de los sensores) y dos salidas (las concentraciones de gas). Siendo esto esencialmente correcto, existen, sin embargo, ciertas particularidades al trabajar con el sistema inverso que deben tenerse en cuenta.

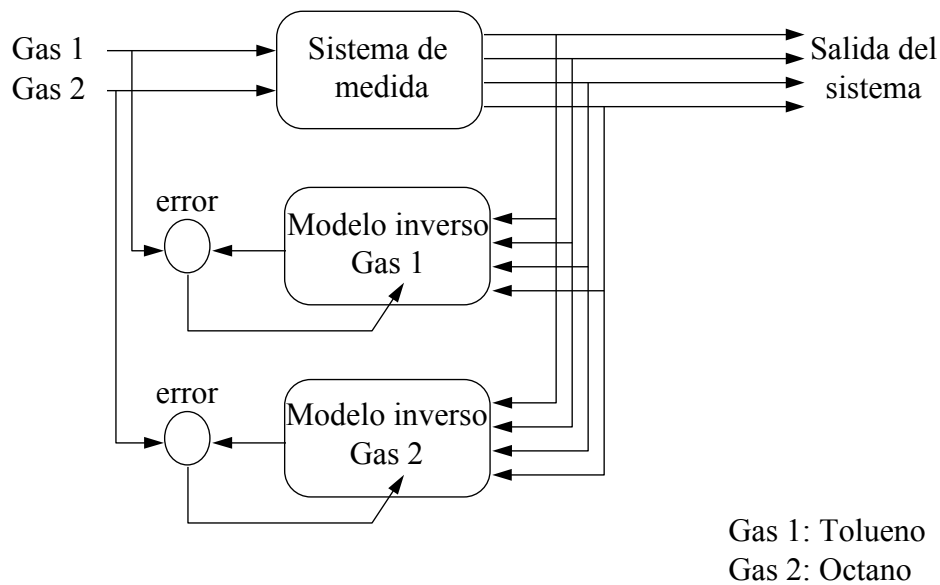


Figura 6. 1 Esquema de la operación inversa

En el estudio directo el sistema es de tipo MIMO, pero si se considera que las salidas de la matriz de sensores son independientes las unas de las otras, el estudio puede reducirse de un sistema MIMO global con n salidas a un conjunto de n sistemas MISO. En el caso inverso, no está claro a priori, si las salidas (concentraciones gaseosas) pueden ser tratadas independientemente, y por tanto se va a trabajar siempre desde el punto de vista MIMO global.

Por otra parte, el hecho de tratar con un sistema inverso de esta manera, supone tratar con un sistema intrínsecamente acausal¹. Efectivamente, supóngase un sistema SISO directo donde la salida depende de entradas anteriores

$$y(n) = f(x(n-1), x(n-2), \dots, x(n-m)) \quad (6. 3)$$

En el siguiente diagrama se representa la relación entre las señales de entrada y salida de un sistema SISO causal. Cada muestra $y(n)$ depende de toda una sucesión de señales de entrada $x(n-k)$ anteriores al momento n . La sucesión puede ser finita o infinita dependiendo de si el sistema es FIR o IIR.

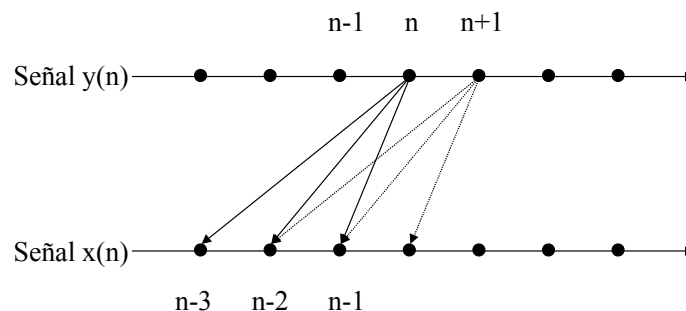


Figura 6. 2 Dependencias dinámicas directas

Centrando la atención por ejemplo en la muestra $x(n-1)$, puede observarse en el diagrama que dicha muestra influye sobre muestras de la señal de salida $y(n)$, $y(n+1)$..., es decir hacia muestras que están en el futuro de la muestra $x(n-1)$.

La siguiente figura esquematiza esta relación acausal de dos formas diferentes. En el primer diagrama de la figura se muestra la relación en la que $x(n-1)$ depende de una sucesión de muestras de la señal de salida, que siempre están en el futuro de $x(n-1)$. La longitud de la secuencia depende de cómo esté definido el sistema directo.

$$x(n-1) = f(y(n), y(n+1), \dots) \quad (6.4)$$

En el segundo diagrama, se representa esta relación desde otro punto de vista: despejando $x(n-1)$ de 6.3 queda una relación de este tipo:

$$x(n-1) = f(y(n), x(n-2), x(n-3), \dots) \quad (6.5)$$

En cualquier caso, la dependencia entre la señal de entrada y la de salida siempre es acausal ya que tanto en 6.4 como en 6.5, $x(n)$ depende de al menos un término que está en su futuro: $y(n)$.

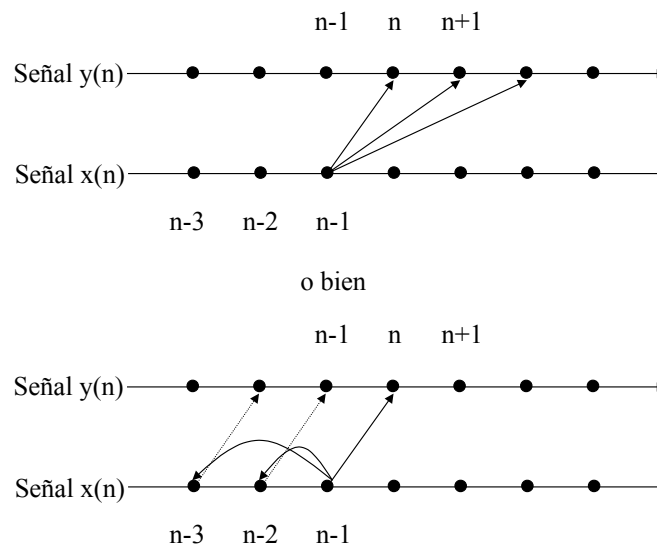


Figura 6.3 Dependencias dinámicas inversas

De estos diagramas se puede observar, y así ocurre en la realidad, que los sistemas que presentan una cierta dinámica en el sentido directo, pueden presentar mejoras en el rendimiento de la estimación inversa si se incluyen dependencias no estrictamente vinculadas en la dinámica. En el ejemplo del diagrama superior de la figura 6.3, para estimar $x(n-1)$ parece que interviene $y(n)$, $y(n+1)$ e $y(n+2)$. Sin embargo, si se incluyen términos del tipo $y(n)$ y/o $y(n-1)$, es muy probable que la estimación mejore, ya que estas muestras contienen información relevante relacionada con la dependencia de $y(n-1)$ y la muestra anterior a $x(n-1)$: $x(n-2)$.

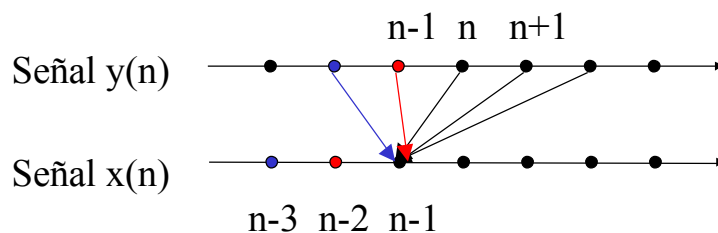


Figura 6. 4 Dependencias dinámicas inversas

En definitiva, los problemas dinámicos del estudio directo también son importantes en el problema inverso. Tal y como ya se a definido, se considera a la matriz de sensores químicos como una unidad de medida dentro del recinto sensorial. En este sentido, y a diferencia de los estudios estáticos, es necesario, al igual que en los estudios directos, tener en cuenta la dinámica de recinto de la matriz de sensores, incluso aunque la dinámica de la cámara sea más lenta que la de los propios sensores. Así, se trata de hallar un modelo dinámico inverso para la matriz de sensores siguiendo el esquema de los modelos directos y teniendo en cuenta los problemas de no linealidad, poca selectividad y posibles derivas temporales. Además se debe aceptar el hecho de que se trata de modelos acausales y que la dinámica del problema contiene, de manera inescrutable, la de los sensores y la de la cámara donde se alojan.

6.2. Aplicación del modelado inverso al sistema real: Experimentos

En este capítulo, se infringe el criterio seguido hasta ahora donde antes de la aplicación de la estimación al problema real se utiliza un problema sintético. En cierta forma, las diferentes estimaciones inversas se basan en procesos ya descritos en los capítulos precedentes, así que no resulta necesario desarrollar ni evidenciar de nuevo el rendimiento de dichos procedimientos, con nuevos problemas sintéticos.

Se han utilizado para la realización concreta de un modelado inverso el experimento 3 y el experimento 4 descritos en el capítulo 3. Recordando ambos experimentos, el número 3 consiste en una secuencia de datos aleatorios, correspondientes a las concentraciones de los gases tolueno y octano a la vez, que presentan distribución gaussiana de valor medio 6500 ppm y desviación estándar 2800 ppm. Gráficas de datos de entrada y salida para este experimento ya se han presentado y los datos fueron usados durante la modelización directa para estudiar los sistemas MISO. El experimento 4 es similar al 3. Se trata también de una secuencia de datos aleatorios que representan las concentraciones de ambos gases. La distribución de estos datos también es gaussiana, y la diferencia esencial se centra en los valores medios y de varianza, siendo el valor medio 1000 ppm y la desviación estándar 360 ppm. Las respuestas de los sensores ante tan poca variación de concentración son muy lineales, con una no linealidad que queda dentro del margen de error que ofrece el ruido experimental. La figura 6.5 presenta la respuesta de los sensores al experimento 4.

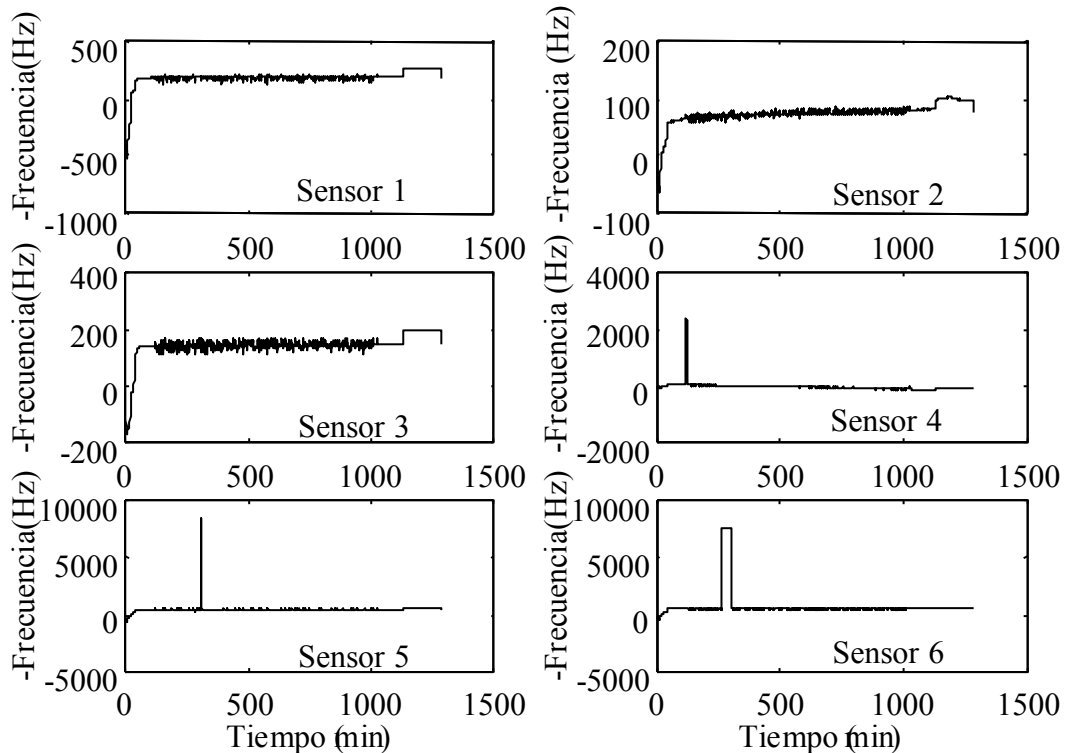


Figura 6.5 -Respuesta de los sensores al experimento 4

6.3. Aplicación del modelado inverso al sistema real: Determinación de los modelos

La dinámica del sistema, como ya se ha dicho, es una combinación del tiempo asociado a la respuesta de los sensores y al tiempo asociado al llenado o vaciado de la cámara¹. Desgraciadamente, ambas dinámicas no pueden separarse de forma efectiva y la contribución a la no linealidad puede ser debida tanto a los propios sensores como a la propia cámara.

Como punto de partida para el ajuste de un modelo inverso se ha escogido un modelo MIMO con respuesta impulsiva finita (FIR). Esta elección se debe principalmente a que es un modelo lineal en los parámetros y permite un ajuste por mínimos cuadrados en un solo paso. Se ha comprobado que generar un modelo IIR a partir de añadir polos al modelo FIR no mejora las prestaciones significativamente.

En definitiva, el modelo FIR que trata de describir al sistema inverso sigue la siguiente ecuación:

$$x_i(n-k) = \sum_{j=1}^r b_{ij}(q)y_j(n) + e_i(n) \quad \text{con } i = 1 \dots m \quad (6.6)$$

donde q es el operador de retraso del polinomio $B_i(q)$, $y(n)$ es la respuesta de los sensores (usadas como datos de entrada), $x(n)$ es la concentración de los diferentes gases (que representan la salida del sistema inverso) y $e(n)$ representa el término de residuo del modelo. Debe notarse la presencia de k retrasos en la salida del sistema inverso (concentraciones) que convierten al modelo en acausal.

A partir de los modelos FIR se han desarrollado algoritmos para encontrar el modelo que mejor se ajusta a los datos del experimento 4. Las siguientes figuras muestran los logaritmos de los valores de la potencia de señal no estimada (PSNE (%)) con datos de

validación tanto para el tolueno como para el octano. Los modelos óptimos para cada experimento y para cada gas se presentan una tabla a continuación de las figuras, donde se precisa también el valor de PSNE calculado en una zona de validación.

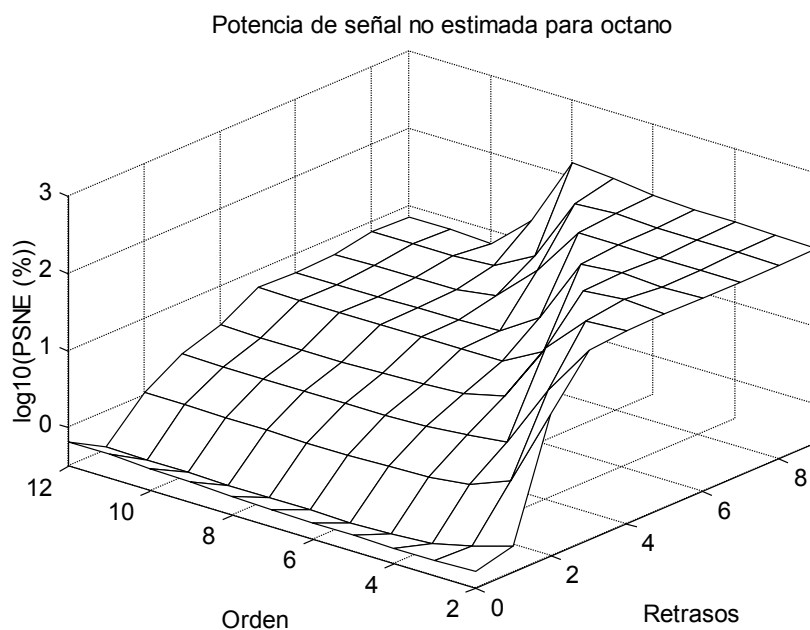


Figura 6. 6 Representación logarítmica del valor de PSNE para gas octano

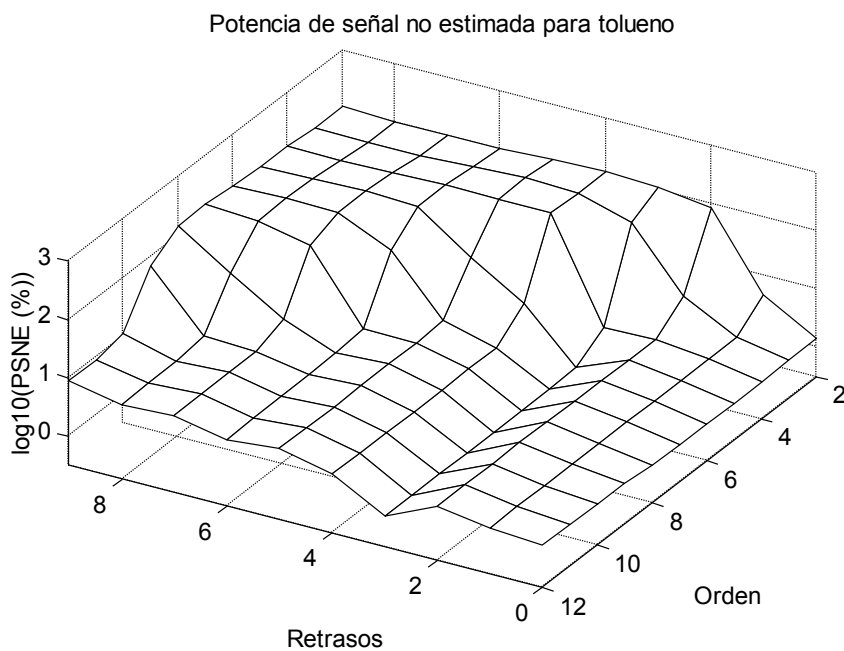


Figura 6. 7 Representación logarítmica del valor de PSNE para gas tolueno

La tabla 6.1 presenta los resultados óptimos hallados con un modelo FIR, tanto para el experimento 3 como para el 4.

Tabla 6. 1 Resultados modelos ARX inversos: zona de validación		
	Tolueno	Octano
Exp. 4	Retraso: 3 Orden: 7 PSNE: $1.0\% \pm 0.08\%$	Retraso: 1 Orden: 5 PSNE: $0.3\% \pm 0.02\%$
Exp. 3	Retraso: 1 Orden: 3 PSNE: $2.3\% \pm 0.2\%$	Retraso: 2 Orden: 4 PSNE: $5.7\% \pm 0.4\%$

El orden de los modelos denota el número de muestras para cada sensor que funcionan como entradas en el modelo, es decir, es el orden del polinomio $B(q)$ más 1.

De los diferentes modelos ensayados para el experimento 3 se puede extraer la figura 6.8 que muestra gráficas con el error en ppm tanto para el tolueno como para el octano en función del orden del modelo. El error que se muestra es el considerado como la diferencia entre el resultado del modelo y el valor real de la concentración.

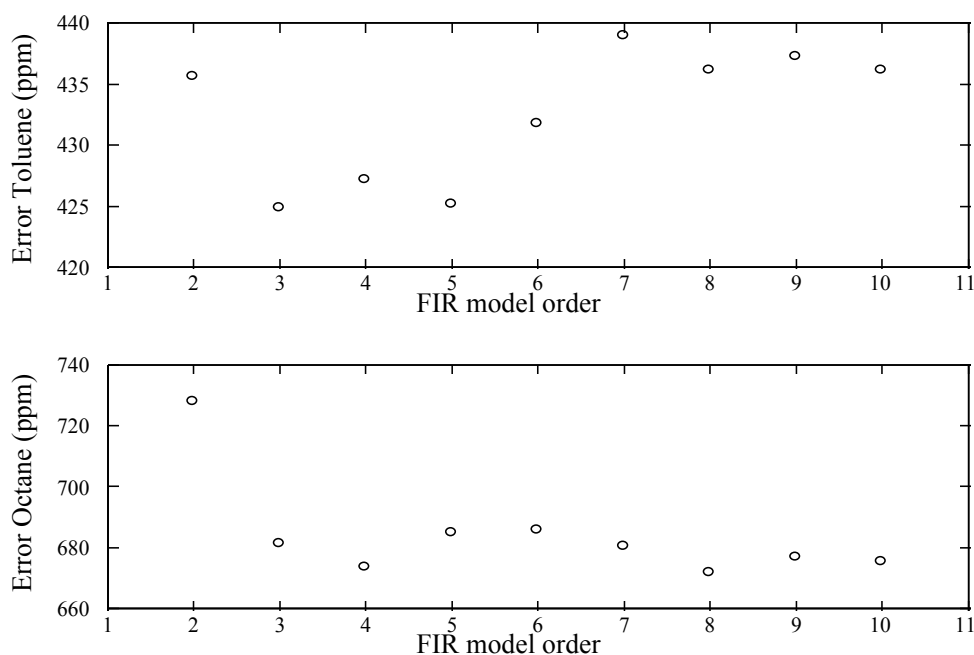


Figura 6. 8 Error rms en función del orden del modelo para el experimento 3

Aceptación de resultados

Los resultados son francamente buenos con un modelo FIR, sobretodo en el caso del experimento 4 donde la desviación estándar es sólo de 360 ppm. Para este caso, las características no lineales son prácticamente despreciables y puede representarse el comportamiento del sistema con un modelo de aproximación lineal. Sin embargo, para excursiones de concentración mayores puede intentarse aproximaciones no lineales que mejoren las características del modelado lineal. Por no reiterar de nuevo con resultados obvios, no se presentan gráficas de análisis de residuos. Si que cabe decir que para el

experimento 3, aquel que hace que el sistema se comporte más no linealmente, el residuo obtenido a partir de la estimación lineal, contiene información evidente, que el modelo no ha sido capaz de extraer.

6.4. Aplicación del modelado inverso al sistema real: modelos no lineales

Para el modelado no lineal, resulta más interesante centrar la atención en el experimento 3, donde existe un grado de no linealidad más acusado que en el experimento 4. De entre todas las estrategias no lineales, y dada la experiencia en identificación directa, se ha centrado la atención en dos: Kernels de Wiener y redes neuronales artificiales.

6.4.1. Estimación inversa mediante el método NARX

Gracias a la analogía matemática entre los kernels de Wiener y los modelos FIR no lineales (NARX) sin polos⁹, se puede, en el proceso de modelado inverso, realizar el cálculo de los propios kernels a través de algoritmos de mínimos cuadrados. Esta metodología ya se utilizó en el caso del modelado directo y se demostró que este proceso era más eficiente que la estimación de correlaciones^{12 13}. La expansión de la serie de Wiener hasta el segundo orden sigue la siguiente expresión:

$$\begin{aligned}
 x(n-k) = & x_0 + \sum_{j=1}^M \sum_{m=0}^{N-1} k_j^{(1)}(m)(y_j(n-m) - y_{j0}) + \\
 & + \sum_{j=1}^M \sum_{m=0}^{N-1} \sum_{r=0}^{N-1} k_{jj}^{(2)}(m,r)(y_j(n-m)y_j(n-r) - P_{jj}) + \\
 & + \sum_{j=1}^M \sum_{l=1}^M \sum_{m=0}^{N-1} \sum_{r=0}^{N-1} k_{jl}^{(2)}(m,r)(y_j(n-m)y_l(n-r) - P_{jl})
 \end{aligned} \tag{6.7}$$

donde M , en el caso que ocupa a esta parte del trabajo, toma el valor 4 ($M=4$) ya que se utilizarán los resultados de 4 sensores. La variable $k_j^{(1)}$ representa al kernel de primer orden, mientras que $k_{jj}^{(2)}$ es el autokernel de segundo orden y $k_{jl}^{(2)}$ es el kernel cruzado de segundo orden. La letra P representa la potencia de la señal siendo entonces P_{jl} la potencia cruzada entre las señales y_j e y_l . Además x_0 e y_{j0} son los valores medios de las señales x e y_j .

Sin embargo, tal y como se ha mencionado, aunque las expansiones en series de Wiener pertenecen al campo de los modelos no paramétricos, para los modelos con memoria finita, es factible realizar una estimación mediante mínimos cuadrados. Este método consigue mejores resultados que los métodos de correlación que, como ya se ha dicho, precisan una mayor cantidad de datos para obtener la misma precisión, especialmente si el sistema con el que se trabaja es MIMO.

Al igual que en el modelado directo, en el proceso inverso se deben realizar algunos estudios previos hasta conseguir encontrar los modelos que proporcionan un error mínimo. En un primer paso se realizaron estimaciones groseras a través de cálculos de correlación para, posteriormente, afinar los resultados a través de algoritmos de mínimos cuadrados. Realizados estos estudios, se pudo comprobar que de los kernels de segundo orden sólo los términos de la diagonal presentan valores significativos, lo que redundaría en modelos más pequeños.

Así, los resultados óptimos se presentan en la tabla 6.2:

Tabla 6.2 Estructura del kernel de Wiener: resultados óptimos	
Kernels de Wiener	
Tolueno	Retraso = 1 Orden = 3 PSNE = 0.5% \pm 0.03% (~100 ppm)
Octano	retraso=2 Orden=4 PSNE = 0.6% \pm 0.04% (~100 ppm)

El aspecto de los kernels de 2° orden se presenta en la figura 6.9:

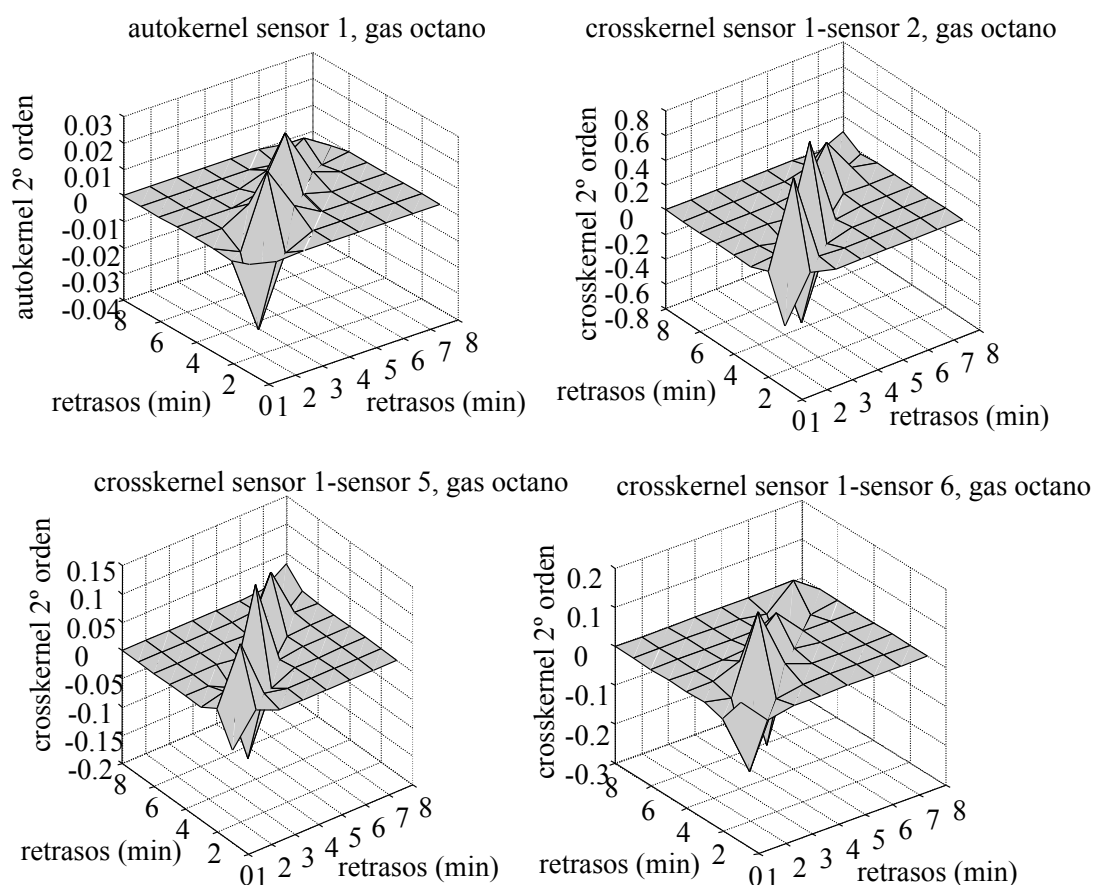
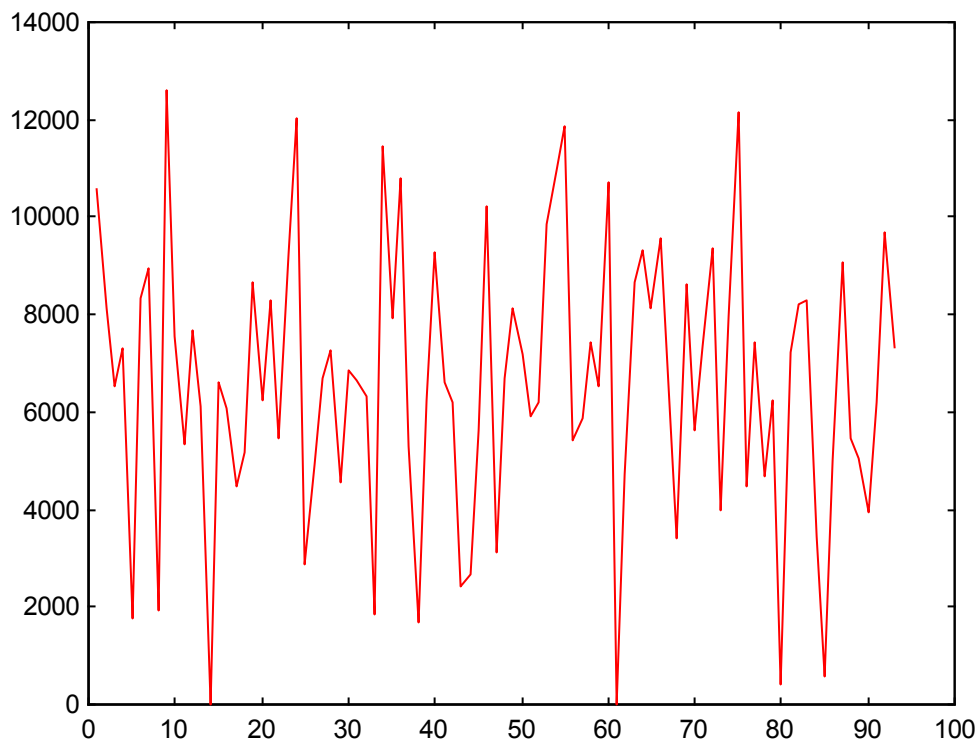
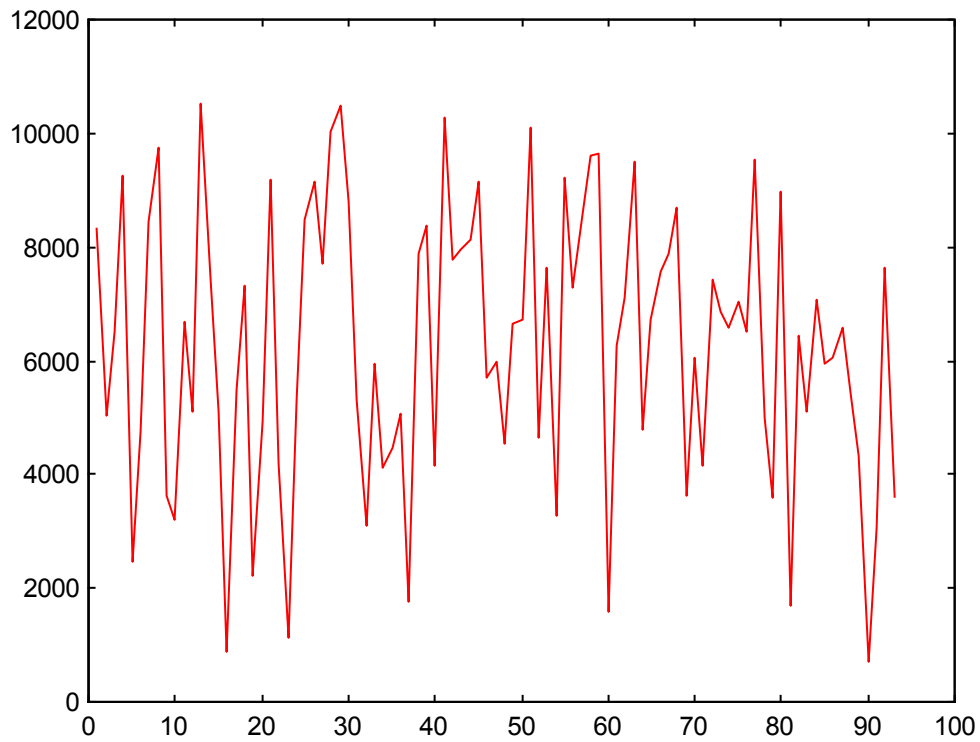


Figura 6.9 kernels de segundo orden para la respuesta del sensor 1

Las siguientes figuras muestran las concentraciones de octano y tolueno a la vez que la simulación de las mismas a través de un modelado inverso en base a kernels de Wiener de primer orden y de segundo orden según el modelo óptimo. Las gráficas en color rojo representan los valores reales de concentración en la zona de validación estudiada mientras que las gráficas en verde representan al resultado ofrecido por el modelo.



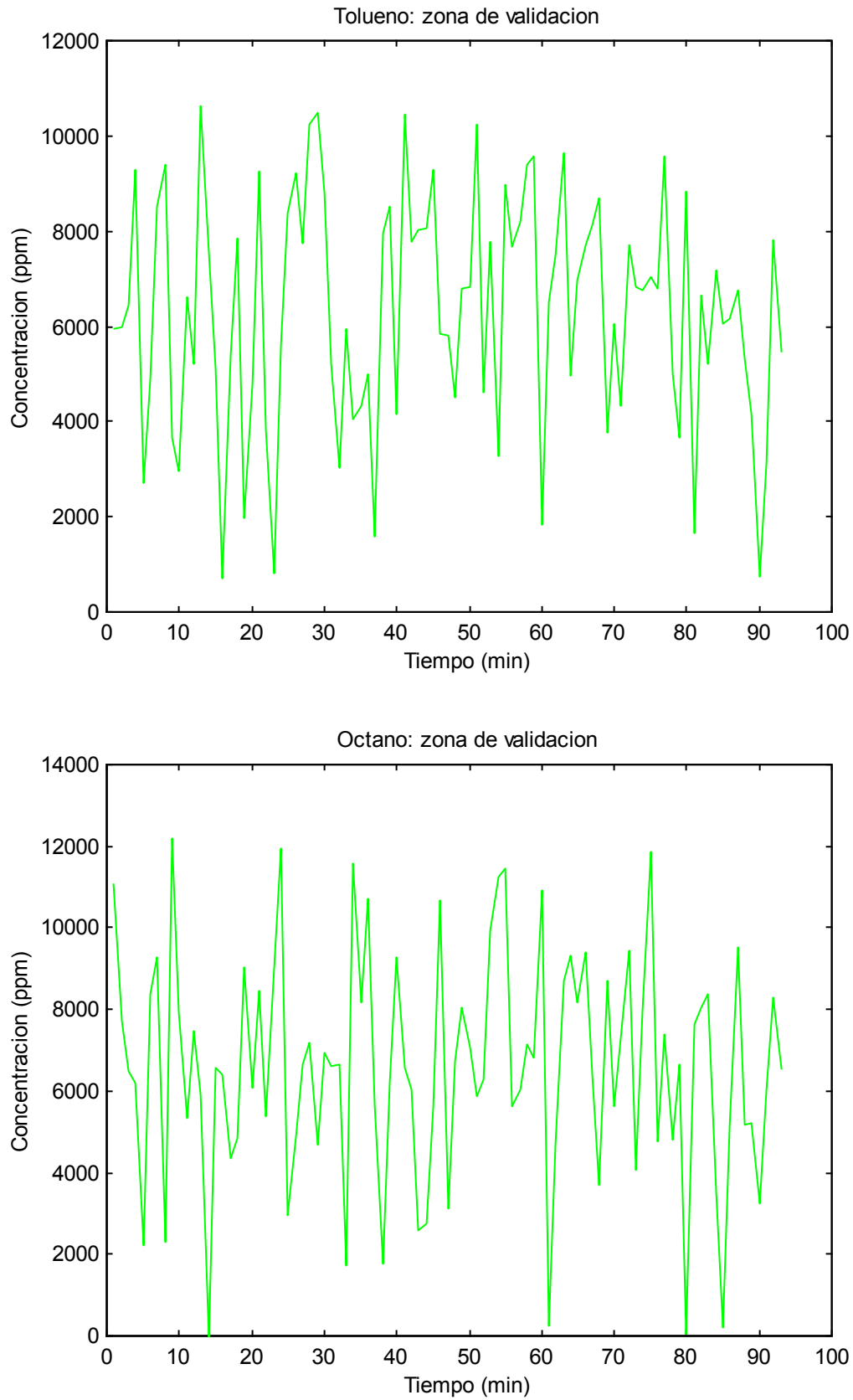


Figura 6. 10 Resultados de la estimación sobre la zona de validación

Aceptación del modelo

Los resultados obtenidos para un modelo inverso a través de los kernels de Wiener calculados mediante mínimos cuadrados han resultado excelentes, llegando a la precisión del experimento real, que se sitúa en el entorno de 100 a 200 ppm de gas. Cabe investigar si las redes neuronales son capaces de llegar a estas prestaciones con la misma o mejor eficacia.

6.4.2. Estimación inversa mediante redes neuronales

Para investigar las redes neuronales artificiales con relación al problema inverso, buscando igualar o mejorar las prestaciones del método basado en kernels de Wiener, se ha experimentado con diferentes estructuras de red neuronal. Algunas de ellas, como las redes recurrentes Elman no aportan buenos resultados, por lo que no se presentan en este trabajo. Las dos estructuras neuronales que consiguen explotar los experimentos con una potencia similar a los algoritmos basados en kernels de Wiener son: redes feedforward y redes Radial Basis ^{2 14}.

Ajuste mediante perceptrones multicapa (MLP).

Un problema inicial a la hora de trabajar con redes multicapa basadas en entrenamiento backpropagation es la elección del tamaño de la red y su conectividad. Este problema ya había surgido en la estimación directa. En aquel caso el problema se atacó desarrollando diversas redes de diferentes dimensiones hasta encontrar la estructura adecuada. Para resolver esta misma situación en relación con la evaluación del problema inverso se han seguido dos estrategias diferentes. El primer camino es esencialmente el mismo que en el modelado directo: examinar diferentes estructuras con un número variable de neuronas en la capa escondida (aunque este número nunca ha sido superior a 10).

La mejor configuración obtenida mediante esta vía de solución se presenta en la tabla 6.3.

Gas	Red Feed-forward (Configuración)	PSNE en zona de validación
Tolueno	Retraso: 1 Orden:3 Neuronas en capa oculta: 5	PSNE = 1.2% ± 0.1% (~300 ppm)
Octano	Retraso: 1 Orden: 4 Neuronas en capa oculta: 3	PSNE = 2.6% ± 0.2% (~450 ppm)

La descripción en cuanto al orden y retraso significa lo mismo que en el caso de estimación NARX, es decir, la dinámica de las señales de la capa de neuronas de entrada se inicia con el retraso uno y se extiende hasta 3 ordenes dinámicos posteriores en el caso del tolueno o 4 para el octano. El número de neuronas en la capa escondida también es diferente para el tolueno (3) que para el octano (4).

La segunda estrategia utilizada en la modelización inversa con redes neuronales se basa en construir una red expresamente grande, teniendo siempre presente que debe prevenirse el problema de sobre entrenamiento de la red a causa de una mala relación

entre las muestras de entrenamiento y el tamaño de la red (*overfitting*)^{10 11}. Lo que se entiende por una red “grande” es una red multicapa con un número de parámetros a calcular del mismo orden de magnitud o superior al número de datos usados en la estimación. Por ejemplo, se han probado redes con 3 capas de 12 neuronas en la capa de entrada que representan las entradas de señal de concentración con retrasos que tengan en cuenta la dinámica, 19 neuronas en la capa escondida para desentrañar las relaciones del sistema, y 1 neurona de salida. Esta arquitectura de red supone, sin contar con el parámetro de bias, una cantidad de pesos a estimar de 228, mientras que los datos de estimación son, en total, 300. Para prevenir el problema de *overfitting* hay que controlar, a la vez que se entrena, el error producido con los datos de validación. Inicialmente, el error en la validación decrecerá, pero si se producen problemas de *overfitting*, este error alcanzará un mínimo y comenzará a incrementar su valor. En este momento debe pararse el proceso de entrenamiento y concluir que la estructura de pesos de las diferentes neuronas que resulte de esta parada brusca por *overfitting* es la que produce el error mínimo en la validación y por tanto es la mejor red que se puede encontrar en estas condiciones.

Para intentar mejorar el rendimiento de las redes MLP dentro del problema inverso y en relación con el tamaño de la red, se han experimentado algoritmos de simplificación de estructura, que permiten eliminar aquellas conexiones neuronales menos significativas del proceso, reduciendo el tamaño de la red sin aumentar apreciablemente el error de predicción⁷. Se trata de los denominados pruning algorithms y son varios los motivos que pueden llevar a utilizarlos: Encontrar una arquitectura mejor ajustada al problema, Reducir el coste de la red (coste tanto en el sentido de tiempo de cálculo y complejidad de las rutinas como en el sentido de coste en la implementación del hardware necesario), la capacidad de la red para reconocer y aprender puede mejorar (aunque no está garantizado), las señales innecesarias o las entradas que no aportan información relevante son eliminadas. En general, y sin entrar en detalles, el proceso seguido para realizar la simplificación es el siguiente:

- 1) Elegir una arquitectura razonable para el problema
- 2) Entrenar la red
- 3) Computar la relevancia de los elementos (conexiones o neuronas) de la red en cuanto a su influencia en el rendimiento de la misma
- 4) Eliminar los elementos poco relevantes
- 5) Entrenar la red resultante de nuevo
- 6) Si el error está dentro de la tolerancia que se especifique, volver al punto tres.
- 7) Regenerar la ultima arquitectura que conseguía estar dentro del margen de error especificado.

La evaluación de la relevancia depende del algoritmo empleado. El más sencillo y el empleado en este trabajo es el denominado “Magnitude Based Pruning”. Consiste en eliminar, después de cada entrenamiento, aquella conexión relacionada con el peso de valor absoluto más pequeño.

Mediante este proceso se consigue mantener buena parte de la potencia y versatilidad de la gran red, pero la reducción de conexiones hace que el número de parámetros a calcular sea mucho menor, permitiendo así el trabajo con cantidades moderadas de datos. La figura 6.11 muestra las estructuras de la red completa y de la red simplificada. Los resultados relativos a estas configuraciones se muestran en la tabla 6.4.

Gas tolueno

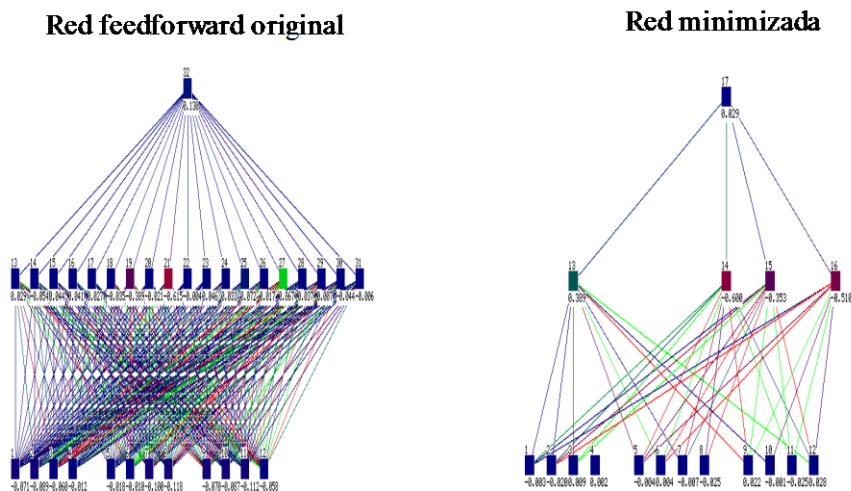


Figura 6. 11 Estructura completa y simplificada de una red neuronal

Tabla 6. 4 Tablas de comparación por tamaño de redes

Gas	Red completa	Red minimizada
Tolueno (Estimación[%])	PSNE = 0.9 ± 0.03 (~350 ppm)	PSNE = 0.9 ± 0.03 (~350 ppm)
Tolueno (Validación[%])	PSNE = 8.6 ± 0.6 (~1060 ppm)	PSNE = 9.2 ± 0.7 (~1090 ppm)

Aceptación de resultados: comparación.

A la vista de esta segunda tabla, se puede concluir que el proceso de generar una red sobredimensionada, para reducirla en un proceso posterior, no ofrece los resultados óptimos para este problema. Si es cierto que el error que comete la red reducida es similar a la red sobredimensionada, pero no es mejor que las redes cuyos resultados se muestran en la tabla 6.4. Parece pues que el algoritmo de redimensión de la red no es capaz de liberarse de la carga de error que introduce el desmesurado tamaño de la red de partida inicial, debido al ‘overfitting’

En referencia a las redes MLP que ofrecen un resultado óptimo, el error que se alcanza a través de ellas, no es menor que el que se consigue con los kernels de Wiener. Además, el tiempo que emplea la red en entrenar y ofrecer el resultado es también muy superior al utilizado con los kernels. No parece interesante por tanto utilizar redes neuronales para modelar un sistema inverso cuando existe un método más eficiente y más rápido. Centrándose en el rendimiento de las redes sobre el problema planteado puede concluirse que son diversas las causas que provocan el bajo rendimiento de las mismas frente a los kernels de Wiener. Por un lado, y en referencia a las redes MLP sobredimensionadas, es evidente que la escasez de datos es un lastre insalvable para ellas; se precisan más datos para entrenar una red grande como la propuesta.

Por otro lado, la no linealidad del problema estudiado, siendo evidente, no es tan acusada como para que al truncar el desarrollo al segundo orden los resultados no sean

ya notablemente buenos. Las redes neuronales planteadas son estructuras de aproximación no lineal, pero para ordenes bajos de no linealidad y frente a la escasez de datos su rendimiento no es mejor (de hecho es incluso peor) que el mostrado por los kernels de Wiener. Con una no linealidad superior, la estimación con kernels hasta el segundo orden sería insuficiente y el rendimiento de las redes respecto de los kernels muy probablemente mejoraría.

Modelización mediante redes neuronales Radial Basis

En el capítulo uno se expusieron una serie de conceptos básicos relativos a las redes neuronales en general y su arquitectura. Hasta este momento, el comportamiento de las redes utilizadas ha seguido el esquema básico desarrollado entonces. Sin embargo, las redes Radial Basis presentan algunas peculiaridades que es conveniente resaltar ahora. La arquitectura de una red Radial Basis (RBF) es similar a las arquitecturas feed-forward; se componen, esencialmente de los nodos de entrada de datos, de una capa de neuronas RBF y de una capa de salida. La diferencia fundamental radica en la neurona RBF, y no en la arquitectura de la red ya que las operaciones que realiza una neurona RBF no son las mismas que las que se ejecutan en una neurona de una red feed-forward²³. El siguiente diagrama pretende esquematizar el conjunto de operaciones que se realizan en una neurona RBF.

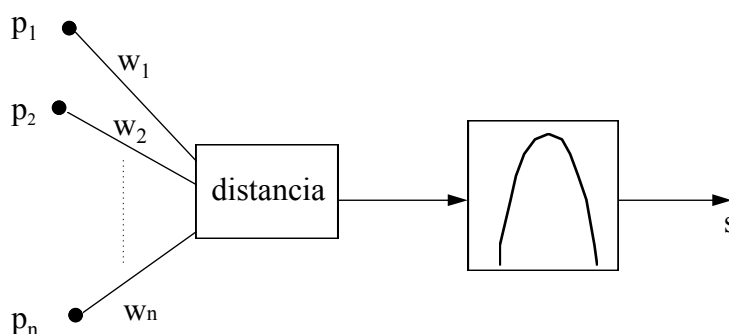


Figura 6. 12 Esquema de la estructura de una neurona Radial Basis

En primer lugar, los datos de la entrada no se combinan linealmente con los pesos de la neurona. La operación que se realiza en su lugar es el cálculo de la distancia euclídea entre el vector de pesos w_i y las entradas p_i . Los pesos definen la posición de la neurona en el espacio n -dimensional y, por tanto, esta operación define la distancia del vector de patrones a la neurona.

La función de activación de las neuronas RBF, por otra parte, no es ninguna de las vistas hasta ahora, sino que se corresponde con una función de tipo gaussiana normalizada su máxima amplitud a 1. Gracias a esta función, si la distancia entre las entradas y los pesos es pequeña, la salida s de la neurona será cercana a 1, mientras que a medida que dicha distancia disminuya, también lo hará el valor de s , siendo prácticamente 0 cuando la distancia sea grande. El resultado en la arquitectura completa es que las entradas alejadas de los pesos de las neuronas RBF ofrecen salidas cercanas a 0 que, en consecuencia, tienen un efecto despreciable en la capa lineal de salida. Por contra, las entradas cercanas a los pesos de las neuronas RBF ofrecen valores altos en su salida y, por tanto, pasan valores apreciables a la capa lineal de salida.

Las redes Radial Basis pueden entrenarse en función de diversos algoritmos^{4 5 6}. El algoritmo que se ha utilizado en el presente estudio es “Orthogonal Least squares” (OLS)⁴.

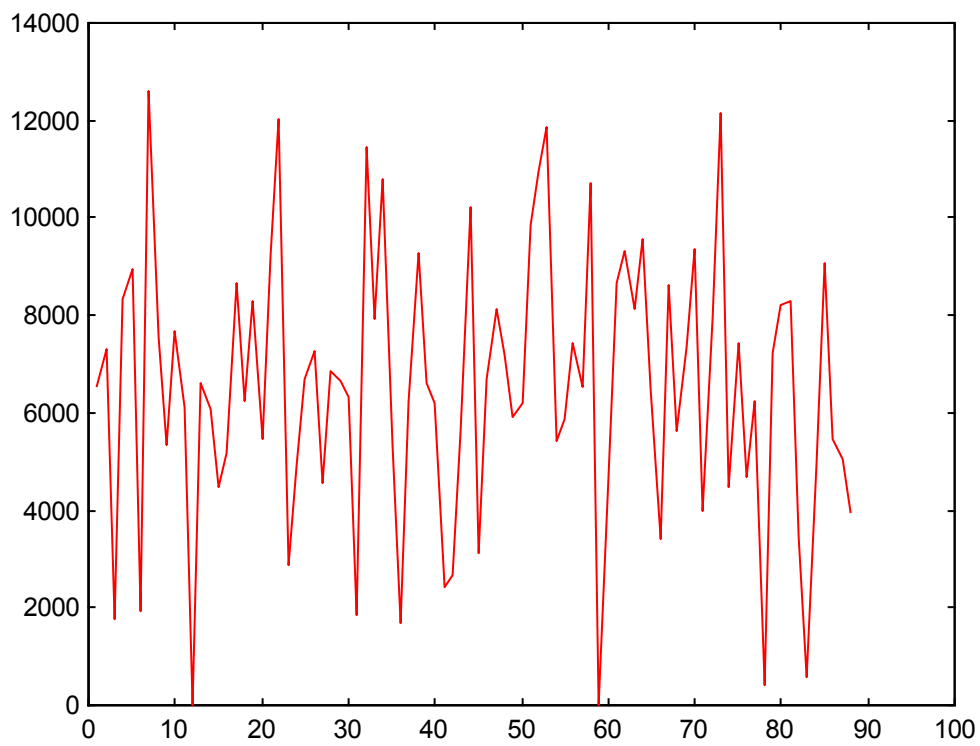
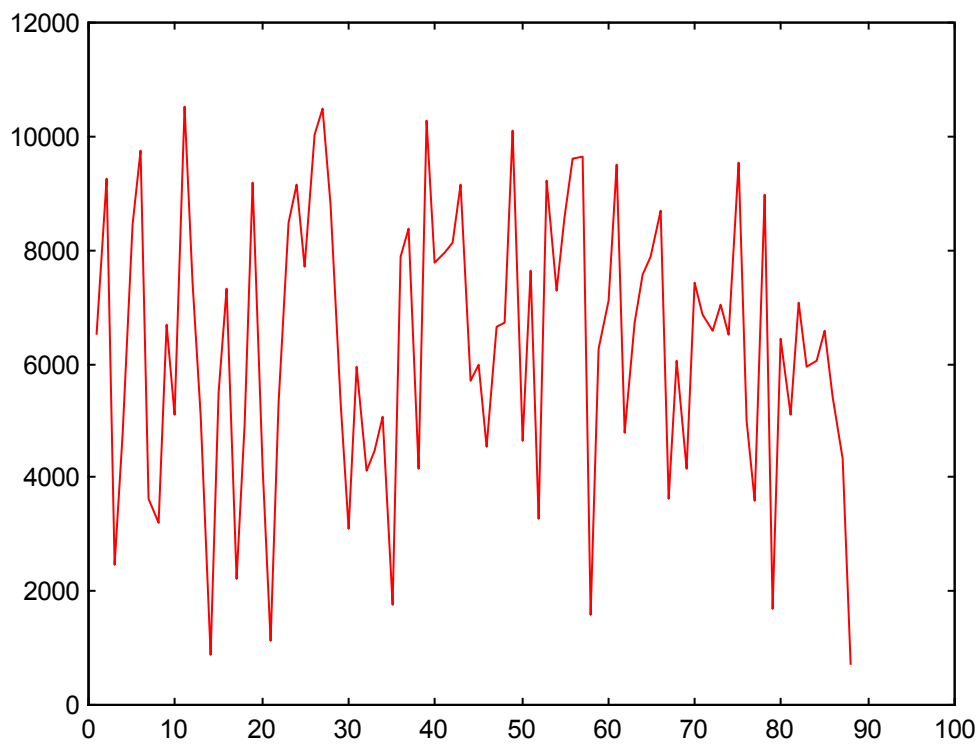
Hablando en términos generales se puede decir que las redes RBF junto con el algoritmo de entrenamiento OLS, tardan menos tiempo en ejecutar los ciclos de entrenamiento que las arquitecturas basadas en feed-forward, aunque por otra parte suelen tener un tamaño mayor. Una ventaja importante respecto a las redes feed-forward no lineales es que el error que ofrecen las redes RBF es único, y no depende de las condiciones iniciales, mientras que las redes feed-forward no lineales, como ya se ha visto, presentan diferentes resultados para diferentes inicializaciones del proceso de entrenamiento.

En la aplicación al trabajo, se han desarrollado dos redes RBF (una para tolueno y otra para octano) con la arquitectura típica de las mismas (capa de entrada, capa escondida y capa de salida). La cantidad de neuronas en la capa escondida se ha ido modificando hasta encontrar en número óptimo. Se utilizan 300 puntos para realizar las estimaciones y 90 para las validaciones. El entrenamiento se realiza a través de algoritmos de mínimos cuadrados y cabe remarcar que los mejores resultados se obtuvieron con un número de neuronas inferior al número de patrones. Para un sistema lineal y sin ruido, lo óptimo es disponer de tantas neuronas RBF como patrones tenga el sistema. De esa manera cada neurona se especializa en un patrón. Sin embargo, en los sistemas reales, con sus problemas de no linealidad y ruido, en general un número demasiado grande de neuronas produce peor rendimiento. Las siguientes figuras presentan los resultados obtenidos con estas redes tanto para el gas octano como para el tolueno. La tabla 6.5 presenta los datos cuantitativos de esta estimación.

Gas	Red RBF	PSNE en zona de validación
Tolueno	Retraso: 1 Orden:3 Neuronas en capa oculta: 80	PSNE = 0.5 % ± 0.03% (~200 ppm)
Octano	Retraso: 2 Orden: 4 Neuronas en capa oculta: 75	PSNE = 0.9% ± 0.07% (~270 ppm)

Para comprobar el efecto de disponer de demasiadas neuronas se presenta la tabla 6.6 aplicada al gas octano. En ella se puede comprobar como, aunque el error en la estimación siempre disminuye, el error de validación comienza a incrementarse a partir de una determinada estructura neuronal. Como siempre, este problema proviene de una inadecuada relación entre el número de datos y el número de parámetros a estimar.

Nº neuronas	20	30	50	80	90	100	120
Estimación	2.9 ± 0.1	2.9 ± 0.1	0.8±0.03	0.4±0.01	0.3±0.01	0.2±0.00	0.1±0.00
Validación	5.3 ± 0.4	3.7 ± 0.3	2.0 ± 0.2	2.1 ± 0.2	2.0 ± 0.2	2.0 ± 0.2	3.9 ± 0.3



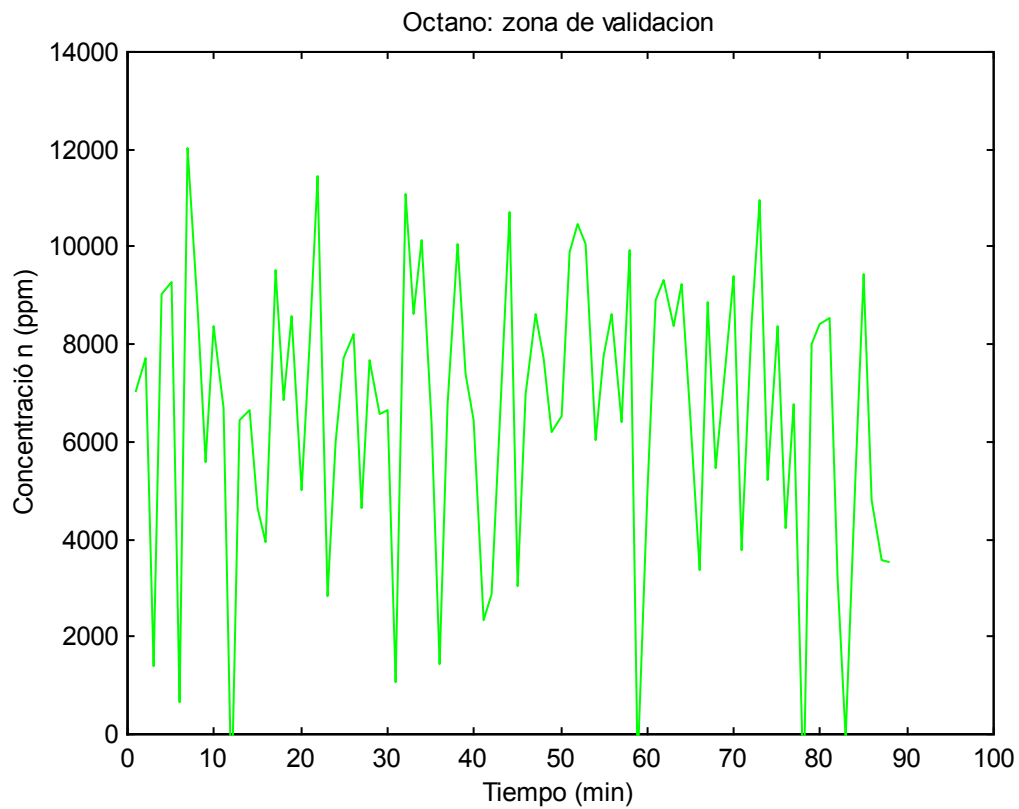
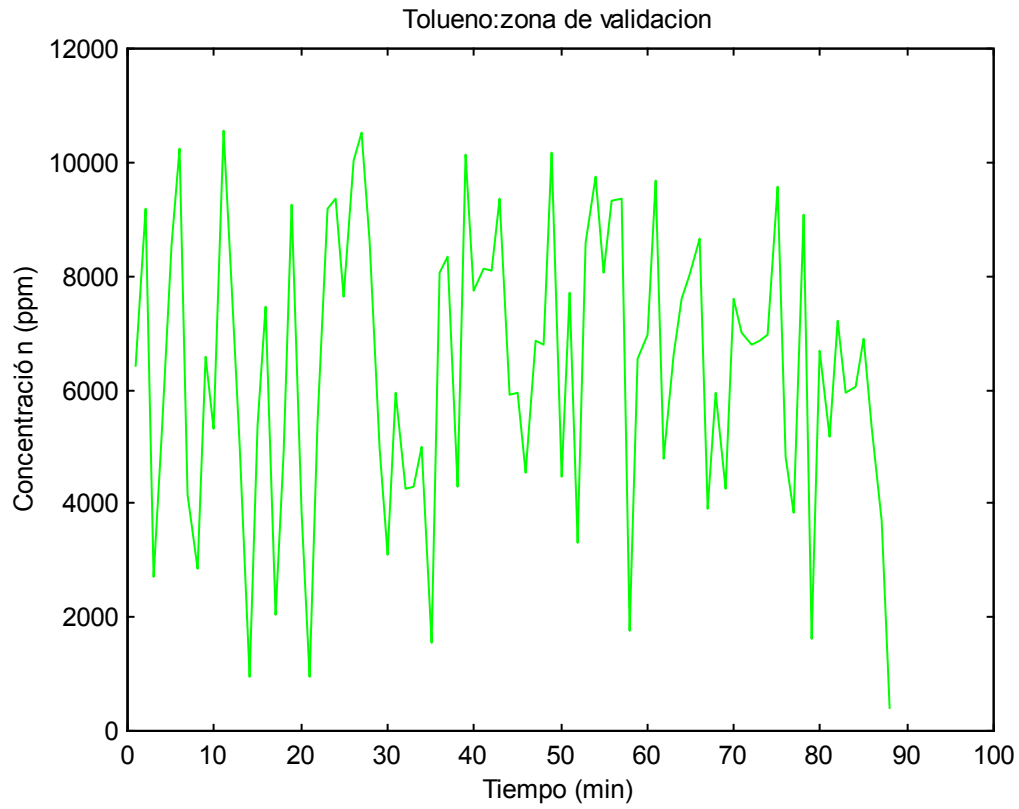


Figura 6.13 Resultados de la estimación en zona de validación

En la figura 6.14 se muestra la evolución del error tanto en la estimación como en la validación para los gases octano y tolueno, en función del número de neuronas.

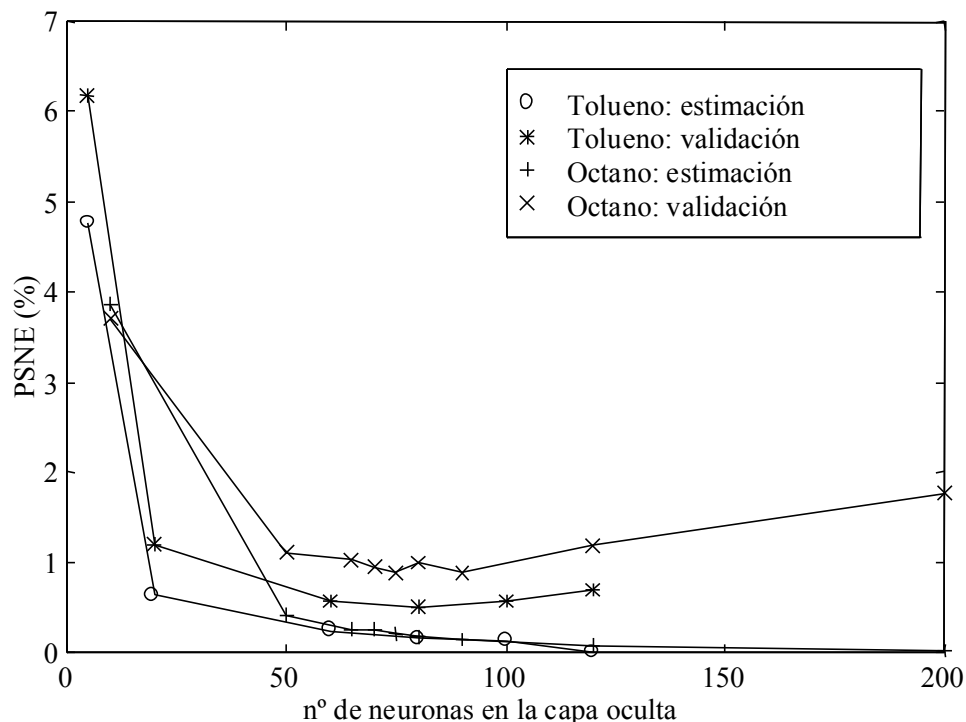


Figura 6.14 Evolución del error en función de la estructura

Las pruebas realizadas con las redes RBF resultan satisfactorias en el sentido que son capaces de llegar, al igual que los kernels de Wiener, al límite experimental. Sin embargo el número de neuronas que se emplean es grande (aunque por debajo del número de datos de entrada). Para el sistema estudiado, el número óptimo se estabiliza alrededor de las 80 neuronas.

Aceptación de los resultados

Los resultados extraídos de una red RBF para la identificación inversa son comparables a los obtenidos con los kernels de Wiener. Es decir, se llega hasta el límite que marca la precisión experimental. Sin embargo, siendo el entrenamiento más rápido que el de las redes feed-forward, y presentando siempre el mismo error, no puede compararse, en cuanto a rapidez al proceso mediante mínimos cuadrados de los kernels de Wiener. Sin embargo, en los kernels de Wiener, el problema fundamental, ya comentado varias veces, consiste en hallar los términos dinámicos significativos. Este problema no existe en las redes Radial Basis entrenadas con el algoritmo OLS, ya que el propio algoritmo se encarga de ir seleccionando las dinámicas significativas.

Cabe destacar la gran diferencia en el número de neuronas necesarias en uno y otro caso. Mientras en las feedforward eran precisas menos de una decena de neuronas, en las RBF un número de neuronas inferior a 20 produce resultados de baja calidad.

6.5. Resumen de resultados y conclusiones

El resumen de los resultados de los diferentes métodos utilizados se presenta en la tabla 6.7.

Modelo	error para Tolueno (%)	error para Octano (%)
No linealidades estáticas	7.8	28.9
FIR	2.3	5.7
Redes de Elman	3.7	4.6
Redes Feedforward	1.2	2.6
Redes RBF	0.5	0.9
Wiener	0.5	0.6

A lo largo de esta exposición, ha quedado demostrado que para pequeñas excursiones en concentración, un modelo sencillo tipo FIR es suficiente para obtener un buen rendimiento como sistema inverso. Sin embargo, para concentraciones mayores, donde los fenómenos no lineales comienzan a jugar un papel importante en los datos, es necesario avanzar hacia algoritmos de procesamiento más complejos.

Se constata el hecho de que los mejores resultados para este caso, se consiguen a través de las redes neuronales del tipo radial basis y de los kernels de Wiener. En ambos casos se consigue llegar hasta los límites del experimento. El método basado en el cálculo por kernels de Wiener es más rápido, pero presenta la dificultad, ya expuesta varias veces, en el conocimiento de la dinámica adecuada. Las redes RBF no presentan este problema ya que el propio algoritmo se encarga de ir escogiendo automáticamente aquellas dinámicas más adecuadas. El problema de las redes RBF es su tamaño (del orden de 80 neuronas para el problema concreto que se trata) y su mayor lentitud.

El propósito de generar un algoritmo de inversión se ve así cumplido a través de los kernels, demostrando que los sensores QMB son susceptibles de ser utilizados como sistema de medida de concentraciones gaseosas de octano y/o tolueno. Los problemas de no linealidad o efectos dinámicos se pueden resolver a través del procesamiento de la señal, llegándose, en la descripción del sistema, hasta el límite del error experimental en los casos óptimos (Wiener y RBF).

Bibliografía

- ¹A.Pardo, S. Marco, J. Samitier “Nonlinear Inverse Dynamic Models of Gas Sensing Systems Based on Chemical Sensor Arrays for Quantitative Measurements”, IEEE Transactions on Instrumentation and Measurement, **47, 3**, 644-651 (1998)
- ²S. Haykin, *Neural Networks, a comprehensive foundation* (IEEE Press, 1994).
- ³M. Beale H. Demuth, *Neural Network toolbox user's Guide* (The Mathworks Inc, 1994).
- ⁴ S. Chen, C.F.N. Cowan and P.M. Grant, “Orthogonal Least Squares Learning for Radial Basis Function Networks” IEEE transactions on Neural Networks, **2,2**, 302-309 (1991)
- ⁵ S. Lee and R.M. Kil “Multilayer feedforward potential function network” Proc. Int. Joint Conf. Neural Networks, **1**, 161-171 (1988)
- ⁶ J.A. Leonard, M. A. Kramer and L.H. Ungar, “Using Radial basis functions to approximate a function and its error bounds” IEEE Trans. Neural Networks **3(4)**, 624-627 (1992)
- ⁷ SNNS (Stuttgart Neural Network Simulator) User Manual, version 4.0
- ⁸ T. Daboczi, I. Kollar. “Multiparameter Optimization of Inverse Filtering Algorithms” IEEE Trans. On Instrumentation and Measurement, Vol 45, n° 2, 417-421 (1996)
- ⁹ D. Kluwer. “Sensor ans Sensory Systems for an Electronic Nose”, Academic Pub., 1992
- ¹⁰ S.C. Hunag, Y.F. Huang. “Bounds on the number of hidden neurons in multilayer perceptrons”. IEEE Trans. Neural Networks 2(1): 47-55 (1991)
- ¹¹ G. Mirchandi, W. Cao. “On hidden nodes in neural nets” IEEE Trans Circuits Syst. 36(5), 661-664 (1989)
- ¹²R. Haber, “Parametric Identification of nonlinear dynamic systems based on nonlinear crosscorrelation functions,” IEE proceedings **135 n° 6**, 405-420 (1988).
- ¹³S.A. Billings S. Chen, “Representations of non-linear systems: The NARMAX model,” Int. J. Control **49 n° 3**, 1013-1032 (1989).
- ¹⁴C.S. George Lee Chin-Teng Lin, *Neural Fuzzy systems* (Prentice Hall, New Jersey, 1996).

Principales aportaciones y conclusiones

- Este trabajo está centrado en el análisis y modelización del comportamiento dinámico de un sistema de medida de gases basado en sensores QMB ante cambios rápidos de dos concentraciones en una mezcla binaria de tolueno y octano en aire sintético en régimen de altas concentraciones.
- Cuando se producen variaciones rápidas en la composición gaseosa la utilización de una calibración estática del sistema para tratar los resultados lleva a errores importantes. Por primera vez, frente a la solución estática, se ha abordado la modelización dinámica de analizadores de gases basados en matrices de sensores no selectivos. Esta modelización contempla el comportamiento conjunto de la matriz de sensores y la cámara donde se alojan los mismos.
- Para abordar el estudio del comportamiento dinámico del sistema se han diseñado un conjunto de experimentos que incluyen, fundamentalmente dos tipos:
Tipo I) señales de concentración en forma de escalones
Tipo II) señales de concentración en forma de ruido blanco con densidad de probabilidad gaussiana.
El diseño de estos experimentos está orientado hacia la confección de modelizaciones en el dominio del tiempo.
- Como condición adicional, se ha acometido la tarea de modelización con una limitación en el número de datos a 300 minutos de experimento (un dato por minuto). De esta manera, los costes de la calibración, y en general, cualquier proceso relacionado con el tamaño de los modelos se minimizan.
- El comportamiento del sistema frente a los experimentos de Tipo I muestra la característica estática y evidencia la existencia de efectos dinámicos y no lineales importantes que se acentúan al aumentar la concentración. Los errores cometidos al utilizar la característica estática frente a las rápidas variaciones de los experimentos Tipo II son importantes, y justifican modelos más complejos en estos casos.
- Se han obtenido dos tipos de modelo del sistema de medida; directos: modelización desde la entrada hacia la salida e inversos: modelización desde la salida hacia la entrada. Los modelos directos son de aplicación en simulación y control, mientras los modelos inversos abordan el problema de la medida.
- La elaboración de modelos dinámicos lineales mejora el rendimiento de la calibración estática, pero el análisis de los residuos muestra que existe una parte no lineal no modelada.
- El análisis de residuos se ha mostrado como una eficiente herramienta para la validación de los modelos, aportando información valiosa sobre la eficacia del modelo, y evidenciando, en los casos correspondientes, la existencia de información que el modelo no es capaz de tener en cuenta.
- Para la modelización dinámica no lineal se han explorado estimaciones y modelos de diverso origen como: estimadores de correlación, estimación paramétrica,

modelos con series de Volterra, modelos con series de Wiener, modelado por cajas y redes neuronales.

- Las estimaciones basadas en correlación cruzada permiten una evaluación rápida y sencilla de las características, pero requieren una gran cantidad de datos para que el estimador tenga poca varianza. La estimación paramétrica se presenta con capacidad para obtener resultados robustos frente a la escasez de datos.
- Los modelos de cajas, aunque conceptualmente sencillos e interesantes, tienen como problema fundamental la dificultad en la determinación de la topología adecuada. Por otra parte, faltan algoritmos de estimación de sus parámetros eficientes.
- Las series de Volterra estimadas con métodos paramétricas funcionan con buen rendimiento, pero no son óptimas. Frente a ellas, las series de Wiener con funcionales ortogonales al ruido blanco y gaussiano, presentan mejores resultados.
- Las series de Wiener estimadas mediante métodos paramétricos junto con el uso de información proveniente de estimaciones lineales previas para conjeturar su estructura, han mostrado su buen rendimiento y su robustez frente a la escasez de datos y sistemas multientrada. Los resultados pueden considerarse al límite del experimento, y son tan buenos debido a que el sistema es sólo levemente no lineal. Este hecho permite cortar los desarrollos no lineales en el orden 2, donde además, la necesidad de datos es adecuada a la limitación autoimpuesta.
- Las series de Wiener por si solas, adolecen de un método para optimizar su estructura. Sin disponer de la capacidad que ofrece el conocimiento y la información de estimaciones anteriores para realizar conjeturas sobre su estructura, el proceso de estimación mediante series de Wiener puede resultar difícil y poco eficiente.
- Los modelos no lineales muestran un acoplamiento entre las señales de entrada de tolueno y octano. Este fenómeno es, probablemente, debido a un efecto en los propios sensores.
- Las redes neuronales (perceptrones multicapa) son capaces de hallar los modelos adecuados, pero requiere un entrenamiento tedioso para hallar la arquitectura correcta y anular la dependencia de la inicialización de los pesos. Frente a estas, las redes neuronales Radial Basis Functions entrenadas con el algoritmo Orthogonal Least Squares ofrecen también buenos resultados sin exhibir los problemas previamente citados.
- Redes neuronales Radial Basis y series de Wiener estimadas paraméricamente, ofrecen los mejores resultados en los procesos de modelización inversa.
- Ha sido posible recuperar, a partir de las señales de los sensores y mediante diversos procedimientos de estimación, las señales de concentración a la entrada del experimento. Esto se ha hecho con un rendimiento al límite de lo que los experimentos permiten, a pesar de la rápida variación de las señales que intervienen y a pesar de la escasez de datos autoimpuesta. Por tanto, los modelos inversos hallados permiten compensar la no selectividad de los sensores y los efectos dinámicos inducidos por la propia cámara.

- Con este trabajo se inició el desarrollo de la primera estación de test de sensores de gas del departamento de electrónica de la Universitat de Barcelona. Inicialmente, la estructura básica de la estación fue desarrollada por mi. En la actualidad, y gracias a las diferentes aportaciones del grupo de trabajo, la estructura básica inicial, se ha convertido en un completo sistema de medición y test capaz de realizar una gran y variada cantidad de mediciones relativas a gases y/o sustancias volátiles, dotando al departamento de una línea de trabajo nueva pero a la vez potente y bien equipada.
- Cuestiones como el rendimiento de los modelos frente a las derivas temporales no han sido tratadas en este trabajo y han quedado pospuestas para trabajos futuros. En este sentido, y saltando del análisis cuantitativo al reconocimiento cualitativo, la tarea en el campo de clasificación de patrones tampoco ha tenido cabida en este trabajo a pesar de haber trabajado parcialmente en ella durante la realización de esta tesis.

Resumen de artículos y comunicaciones a congresos

Los siguientes artículos y comunicaciones a congresos han sido fruto directo del trabajo realizado en esta tesis:

- ✓ J. Samitier, J.M. López Villegas, S. Marco, L. Cámara, A. Pardo, O. Ruiz, J. R. Morante, “ A new method to analyze signal transients in chemical sensors” *Sensor and Actuators B*, Vol 18-19, No. 1-3, 308-312 (1994)
- ✓ F.Davide, C. DiNatale, A. D’Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, W. Göpel, S. Marco, A. Pardo. “Dynamic calibration of QMB polymer-coated sensors by Wiener kernel estimation,” *Sensors and Actuators B* **26-27**, 275-285 (1995).
- ✓ S. Marco, A. Pardo, F. Davide, C. DiNatale, A. D’Amico, A. Hierlemann, J.Mitrovics, M. Schweizer, U. Weimar, W. Göpel, “Different strategies for identification of gas sensing systems,” *Sensors ans Actuators B* **34**, 213-223 (1996).
- ✓ A.Pardo, S. Marco, J. Samitier “Nonlinear Inverse Dynamic Models of Gas Sensing Systems Based on Chemical Sensor Arrays for Quantitative Measurements”, *IEEE Transactions on Instrumentation and Measurement*, **47, 3**, 644-651 (1998)
- ✓ S. Marco, A.Pardo, A. Ortega, J. Samitier ,“Gas Identification with tin oxide sensor array and Self Organizing Maps: Adaptive correction of sensor drifts” *IEEE Transactions on Instrumentation and Measurement*, **47**, 316-321 (1998)
- ✓ A.Ortega , A.Pardo, S. Marco, A. Cirera, A. Cornet, J.R. Morante, R. Jané, S. Jiménez, E. Castaño, J. Gracia “Smart multisensors integrating methane and carbon monoxide gas detection with improved selectivity by signal processing.” *International Gas research conference, San Diego* (1998)
- ✓ T. Sundic, S. Marco, A.Pardo, A. Ortega, J. Samitier, “Quantitative Signal Processing Algorithms for low Cost Methane and Carbon Monoxide Detectors”, *IEEE Instrumentation and Measurement Technology Conference. Venice* (1999)

Por otra parte, el siguiente listado presenta otras contribuciones que no han sido reflejadas en esta tesis por no pertenecer a su ámbito.

- ✓ S. Marco, A. Pardo, “Activities in Signal Processing and Dynamic Modeling” *Jornadas Hispano-Francesas. Microsistemas i Innovació Tecnològica.*(1995)
- ✓ M. Carmona, S. Marco, H. Nguym, U. Lundquist, A. Pardo, J. Samitier, “A behavioral model for the dynamic simulation of silicon micromachined pumps”. *Mixed Design of Integrated Circuits and Systems – Education of Computer Aided design of Modern Ics and Devices*, 464-469, Lodz (1996)
- ✓ S. Marco, U. Lundquist, M. Carmona , A. Pardo, J. Samitier, ”System Identification for finite element transients” *Micro-Mechanics Europe* 144-147, Barcelona (1996)

- ✓ M. Carmona, S. Marco, H. Nguyen, U. Lundquist, A. Pardo, J. Samitier, "Dynamic simulation of Micromachined Thermo-Pneumatic pumps with behavioral models" XI Conference Design of Integrated Circuits and Systems, Sitges (1996)