

UNIVERSIDAD DE OVIEDO

Departamento de Ingeniería Eléctrica, Electrónica,  
de Computadores y Sistemas

**TESIS**

**DETECCIÓN E IDENTIFICACIÓN DE FALLOS  
EN PROCESOS INDUSTRIALES MEDIANTE  
TÉCNICAS DE PROCESAMIENTO DIGITAL DE  
SEÑAL Y REDES NEURONALES.  
APLICACIÓN AL MANTENIMIENTO  
PREDICTIVO DE ACCIONAMIENTOS  
ELÉCTRICOS.**

Memoria presentada para la obtención del grado de  
Doctor Ingeniero Industrial

Autor: Ignacio Díaz Blanco

Director: Alberto Diez González

Gijón, Julio de 2000



# Agradecimientos

Debo dar las gracias muy especialmente a Alberto Diez, director y tutor, no sólo en esta Tesis, sino en la orientación de toda mi carrera profesional en la Universidad. El allanó el camino poniendo a mi disposición todos los medios necesarios para llevar a cabo este trabajo; su consejo y apoyo han sido esenciales en mi formación.

Guillermo Ojea ha escuchado pacientemente algunas de mis ideas y ha sabido criticarlas y valorarlas en su medida, lo que ha dado un gran impulso a esta Tesis. Tanto él como José Campomanes, han hecho que el Area de Ingeniería de Sistemas y Automática de la Universidad de Oviedo sea hoy un lugar de trabajo en el que el ambiente, tanto en el ámbito personal como en el laboral, sea inmejorable. Ello, sin duda, ha sido muy positivo para mí.

Agradezco de corazón a Fernando Briz el impulso que me dio en los primeros pasos. El me inició en el campo del Procesamiento Digital de Señal; con él y con Juan Carlos mantuve interesantes discusiones que me enriquecieron enormemente y sembraron la semilla de muchas de las ideas que planteo en este trabajo.

Estoy también especialmente agradecido a Florentino Tuya quien, sobre todo durante los dos años en los que compartimos laboratorio en la EUITI, nos prestó con absoluto desinterés, a mí y a quienes estábamos con él, su enorme experiencia práctica y su ayuda en el trabajo de laboratorio.

Debo expresar también mi agradecimiento a Abel Cuadrado, cuya colaboración fue de gran ayuda para mí. Con él he discutido y probado muchas de las ideas que se exponen en este trabajo.

Podría, con toda seguridad, llenar un capítulo agradeciendo la interminable lista de momentos en los que mis compañeros y amigos del Área me prestaron ayuda, ánimo y apoyo.

No sería justo tampoco olvidar al personal de Aceralia. Gracias a Faustino Obeso, Pablo Quiroga, Loredo, Nicolás, Candás, y otros muchos, cuya ayuda y enorme caudal de experiencia han contribuido en gran medida al desarrollo de este trabajo.

No puede faltar aquí el agradecimiento a mi familia. Doy las gracias a mi madre, por su apoyo incondicional y su cariño, y un recuerdo especial para mi padre, a quien le hubiera hecho una enorme ilusión ver terminado este trabajo.

Finalmente, Sonia. Ella consigue que me olvide de la Tesis y que piense en cosas normales. Sin ella, nada de esto tendría sentido.

# Resumen

En este trabajo se propone una metodología para la supervisión de procesos complejos basada en una proyección o *mapa* no lineal del estado del proceso sobre un espacio de baja dimensionalidad (típicamente 2D) que puede ser representado gráficamente. Bajo ciertas condiciones es posible “cartografiar” este espacio, definiendo sobre él las regiones que corresponden mediante la proyección a diversas condiciones de funcionamiento del proceso. El problema se reduce entonces a determinar visualmente sobre un gráfico las regiones atravesadas en cada momento por la proyección del estado.

Se propone aquí una justificación de esta idea bajo el enfoque de la Teoría de Sistemas, considerando el proceso como un sistema dinámico no lineal cuyo estado describiría unívocamente la situación que atraviesa el proceso en cada instante. Desde esta perspectiva, se conciben los procesos de toma de medidas, extracción de características y reducción de la dimensionalidad como transformaciones sucesivas del espacio de estados sobre espacios de inferior dimensionalidad hasta llegar a una dimensión inferior o igual a 3, que permitiría “visualizar” el proceso, sugiriéndose las condiciones mínimas que éstas deben satisfacer para que la información relativa a la condición del proceso se preserve en cada etapa.

El grueso de esta Tesis se enfoca sobre la etapa de reducción de la dimensionalidad. Existen en el estado del arte diversas técnicas de reducción de la dimensionalidad capaces de describir la estructura subyacente en los datos del proceso mediante unas pocas variables. El Mapa Topológico Autoorganizado (SOM), propuesto por Kohonen hace casi dos décadas, destaca entre ellas por sus propiedades de preservación de la topología y aproximación de la función de densidad y se revela como un excelente método de proyección para utilizar en la supervisión de proce-

sos. Su principal inconveniente radica en el carácter discreto del espacio sobre el que realiza la proyección, dificultando la posibilidad de visualizar derivas, tendencias, ciclos límite, etc. a través de la trayectoria proyectada.

Con el objeto de definir una proyección de carácter continuo se propone en esta Tesis una arquitectura basada en interpolación entre los espacios de entrada y salida del SOM mediante regresión de núcleo. La proyección resultante tiene, globalmente, las mismas propiedades que el SOM, tratándose además de una proyección continua sobre el espacio de visualización.

Basándose en esta idea, se sugiere una generalización de la metodología propuesta por Kohonen, extendiendo al caso continuo algunos de sus elementos, tales como los mapas de distancias, mapas de características y error de modelado, e incorporando algunas técnicas nuevas como la visualización 3D de la geometría de los datos de proceso y de su trayectoria basada en la aplicación de técnicas MDS sobre los elementos del SOM, los mapas de activaciones, que facilitan la determinación de regiones asociadas a diferentes condiciones, o el error de modelado vectorial, que permite determinar la presencia de desviaciones del proceso respecto al modelo, indicando además las variables que más se alejan de sus valores previstos.

Finalmente, el trabajo se completa demostrando la aplicación de la metodología con resultados experimentales en dos casos: 1) monitorización de la condición de un motor asíncrono de 4 kW y 2) monitorización de la condición de un motor de CC de 7000 kW instalado en un tren de laminación de la empresa Aceralia.

# Índice General

<b>1</b>	<b>Introducción y Objetivos</b>	<b>15</b>
1.1	Introducción . . . . .	15
1.2	Principales métodos de Detección y Diagnóstico de Fallos	17
1.2.1	Métodos basados en Umbral. . . . .	17
1.2.2	Métodos basados en Modelos. . . . .	18
1.2.3	Métodos basados en Datos. . . . .	24
1.3	Objetivos de la Tesis . . . . .	26
1.4	Formulación del Problema . . . . .	28
1.4.1	La monitorización de la condición en el marco de la Teoría de Sistemas . . . . .	28
1.4.2	Influencia del sistema de medidas en la observabilidad	30
1.4.3	Proyección del espacio de medidas sobre un espacio visualizable . . . . .	34
<b>2</b>	<b>Técnicas de Extracción de Características</b>	<b>37</b>
2.1	Introducción . . . . .	37
2.2	Selección de las señales más significativas . . . . .	38
2.3	Técnicas de Normalización . . . . .	39
2.3.1	Normalización a $[-1, +1]$ . . . . .	40
2.3.2	Normalización de la varianza . . . . .	40
2.3.3	Normalización PCA . . . . .	40
2.3.4	Esferificación . . . . .	41
2.4	Técnicas basadas en Análisis Espectral . . . . .	41
2.4.1	Introducción . . . . .	41
2.4.2	Métodos no paramétricos . . . . .	44
2.4.3	Métodos Paramétricos . . . . .	45
2.5	Análisis Cepstral . . . . .	48
2.6	Análisis de Estadísticos de Orden Superior (HOSA) . . . . .	49

2.7	Conclusiones . . . . .	52
<b>3</b>	<b>Métodos de reducción de la dimensionalidad</b>	<b>55</b>
3.1	Introducción . . . . .	55
3.2	Análisis de componentes principales (PCA) . . . . .	56
3.3	Análisis de Componentes Principales de Núcleo (Kernel PCA) . . . . .	62
3.4	Métodos de Escalado Multidimensional (MDS) . . . . .	70
3.5	Mapas Topográficos Generativos (GTM) . . . . .	76
3.6	Redes Neuronales FeedForward Autoasociativas . . . . .	79
3.7	Curvas y Superficies Principales . . . . .	79
3.8	El Mapa Topológico Autoorganizado (SOM) . . . . .	81
3.8.1	Introducción . . . . .	81
3.8.2	Algoritmo de Proyección . . . . .	81
3.8.3	Algoritmos de Aprendizaje . . . . .	83
3.8.4	Propiedades del mapa de características . . . . .	87
3.9	Conclusiones . . . . .	93
<b>4</b>	<b>Generación de Mapas Continuos</b>	<b>95</b>
4.1	El problema de la cuantificación del SOM en la Supervisión de Procesos . . . . .	95
4.2	Interpolación mediante Regresión de Núcleo . . . . .	96
4.3	Otras Alternativas . . . . .	100
4.3.1	Modelos Lineales Locales . . . . .	100
4.3.2	Algoritmo PSOM . . . . .	101
4.3.3	Núcleos Hiperelipsoidales . . . . .	101
<b>5</b>	<b>Metodología para Monitorización de la Condición basada en SOM Interpolado</b>	<b>107</b>
5.1	Introducción . . . . .	107
5.2	Proyección del Estado del Proceso . . . . .	108
5.3	Mapas de Características . . . . .	109
5.4	Mapas de Activación . . . . .	110
5.5	Mapas de Distancias Interneuronales . . . . .	113
5.6	Error de Modelado. . . . .	114
5.7	Visualización 3D mediante técnicas MDS . . . . .	118
5.7.1	Introducción . . . . .	118

---

5.7.2	Visualización 3D de la geometría espacial de los datos en continuo . . . . .	118
5.8	Aplicación Práctica en Procesos Industriales . . . . .	121
5.8.1	Generación de Ejemplos . . . . .	121
5.8.2	Base de datos de entrenamiento . . . . .	123
5.8.3	Entrenamiento del Modelo . . . . .	124
5.8.4	Interpretación de los Resultados . . . . .	125
<b>6</b>	<b>Resultados Experimentales</b>	<b>129</b>
6.1	Caso 1: Supervisión del estado de un motor de inducción. Prototipo de Laboratorio. . . . .	129
6.1.1	Descripción del Banco de Pruebas . . . . .	129
6.1.2	Extracción de Características . . . . .	131
6.1.3	Entrenamiento del Modelo . . . . .	133
6.1.4	Generación de mapas . . . . .	134
6.2	Caso 2: Supervisión de un motor de CC de 7000 Kw en un Tren de Laminación Semicontinuo . . . . .	143
6.2.1	Extracción de Características. . . . .	146
6.2.2	Entrenamiento del Modelo . . . . .	156
6.2.3	Interpretación de los resultados . . . . .	156
<b>7</b>	<b>Discusión General y Líneas Futuras</b>	<b>175</b>
7.1	Conclusiones finales . . . . .	175
7.2	Aportaciones . . . . .	177
7.3	Futuras Líneas de Trabajo . . . . .	179
<b>A</b>	<b>Nomenclatura</b>	<b>181</b>



# Índice de Figuras

1.1	Esquema del problema de la Monitorización de la Condición de Procesos . . . . .	35
3.1	Ejemplo de aplicación del método PCA . . . . .	60
3.2	Ejemplo de mal funcionamiento del PCA . . . . .	61
3.3	Nube de puntos en $\mathbb{R}^2$ formada por las medidas $(i_{exc}, i_{ind})$ . . . . .	65
3.4	Componentes Principales obtenidas mediante el método <i>Kernel PCA</i> . . . . .	66
3.5	Evolución temporal de las 4 características principales para los datos de proceso. . . . .	67
3.6	Características principales de los puntos de la trayectoria confrontadas entre sí. . . . .	68
3.7	Ejemplo del método CCA . . . . .	73
3.8	Ejemplo del método CCA . . . . .	74
3.9	Representación del SOM en los espacios de entrada y salida. . . . .	82
3.10	Comparación de SOM y PCA en la reducción de la dimensionalidad. . . . .	91
3.11	Partición del espacio de entrada en regiones de Voronoi. . . . .	92
4.1	Datos de ejemplo en el espacio de entrada. . . . .	103
4.2	Interpolación con núcleos hiperelipsoidales. Comparación con núcleos esféricos. . . . .	104
5.1	Principio de cooperatividad en el SOM. . . . .	112
5.2	Error de modelado vectorial. . . . .	117
5.3	Flujograma de aplicación de la metodología propuesta. . . . .	122
6.1	Foto del banco de pruebas . . . . .	130
6.2	Esquema del experimento. . . . .	131

6.3	Diagrama de bloques de la etapa de extracción de características en el Caso 1. . . . .	132
6.4	Espectrograma (diagrama tiempo-frecuencia) del acelerómetro $a_x$ . . . . .	133
6.5	Retícula $15 \times 15$ utilizada en el experimento del Caso 1. . . . .	134
6.6	Mapa de Distancias Caso 1. . . . .	135
6.7	Mapas de Características. Caso 1. . . . .	137
6.8	Mapas de Activación. Caso 1. . . . .	138
6.9	Error de Modelado. Caso 1. . . . .	140
6.10	Trayectorias proyectadas sobre el mapa de distancias y sobre los mapas de los armónicos 48.8 Hz y 97.6 Hz, junto con un detalle de la trayectoria. . . . .	141
6.11	Proyección 3D de la Trayectoria. Caso 1. . . . .	142
6.12	Medidas de proceso. Caso 2. . . . .	144
6.13	Espectrograma de $i_{ind}(t)$ . . . . .	146
6.14	Detalles de los espectrogramas de $i_{ind}(t)$ , $i_{exc}(t)$ y $V_{ind}(t)$ . . . . .	147
6.15	Evolución temporal de las características de $i_{ind}$ (datos de entrenamiento). . . . .	150
6.16	Evolución temporal de las características de $i_{exc}$ (datos de entrenamiento). . . . .	151
6.17	Evolución temporal de las características de $V_{ind}$ (datos de entrenamiento). . . . .	152
6.18	Evolución temporal de las características de $i_{ind}$ (datos de test). . . . .	153
6.19	Evolución temporal de las características de $i_{exc}$ (datos de test). . . . .	154
6.20	Evolución temporal de las características de $V_{ind}$ (datos de test). . . . .	155
6.21	Mapas de distancias interneuronales. Caso 2. . . . .	157
6.22	Mapas de las 16 características del proceso. . . . .	158
6.23	Conjuntos de datos correspondientes a diversas condiciones de funcionamiento del motor. . . . .	161
6.24	Mapas de Activación del Caso 2. . . . .	163
6.25	Etiquetado de regiones en el espacio de visualización. . . . .	164
6.26	Mapas de activación frente a mapas de características . . . . .	165
6.27	Errores de modelado global y vectorial para los datos de entrenamiento. . . . .	167

---

6.28 Errores de modelado global y vectorial para los datos de test. . . . .	168
6.29 Detalle de los errores de modelado global y vectorial para los conjuntos de entrenamiento y de test. . . . .	169
6.30 Mapa de distancias junto con las proyecciones de los datos de test. . . . .	171
6.31 Mapas de activaciones de diversos tramos en vacío. . . .	172
6.32 Mapa de activaciones de diversos tramos laminando. . .	173



# Capítulo 1

## Introducción y Objetivos

### 1.1 Introducción

Las estrategias de mantenimiento de maquinaria mecánica tienen una importante repercusión económica en cualquier instalación industrial. Una aplicación ineficiente del mantenimiento puede acarrear elevados costes por paralización de la producción, gastos de reparación en la maquinaria y movilización de personal, entre otros. Concretamente, el gasto anual medio realizado por empresas industriales en mantenimiento constituye aproximadamente un 80% de la cantidad anualmente invertida en planta y maquinaria [1]. A pesar de ello, en torno a un 70% del trabajo de mantenimiento llevado a cabo por las empresas puede ser catalogado como “mantenimiento de emergencia” o de carácter correctivo. Tales consideraciones, unidas al hecho de que los procesos y la maquinaria involucradas en la industria actual son cada vez más caros y complejos, demuestran la necesidad de un mantenimiento eficiente.

Algunas de las posibles estrategias de mantenimiento empleadas hoy en día son:

**Mantenimiento correctivo:** Este tipo de mantenimiento se lleva a cabo mediante un reemplazamiento o reparación de los componentes sólo cuando éstos fallan.

**Mantenimiento preventivo:** Consiste en el reemplazamiento sistemático de los componentes conforme a su vida útil o su probabilidad de fallo, determinados en base a la experiencia y la estadística de fallos

de los mismos. Se persigue con ésto, lograr que la probabilidad de fallo en la máquina sea nula o despreciable.

**Mantenimiento predictivo o por condición:** Esta estrategia se basa en la realización de tareas de mantenimiento en función de la condición de la máquina, para lo cual precisa de métodos capaces de determinarla con la mayor precisión y fiabilidad posibles.

El *mantenimiento correctivo* puede acarrear graves problemas, no sólo económicos, tales como paradas no planificadas en la producción, averías tremendamente costosas que pueden tener influencia en otros elementos de la cadena de producción o incluso accidentes laborales. El *mantenimiento preventivo*, por otra parte, puede resultar excesivamente caro al reemplazar componentes que a menudo están operativos o pueden seguir funcionando más tiempo. Evidentemente, el *mantenimiento predictivo* se ajusta a un planteamiento más eficiente, que permite resolver los problemas antes de que éstos se produzcan, lo que redunda en un ahorro económico considerable.

El mantenimiento predictivo depende en gran medida de la aplicación de métodos capaces de determinar con la mayor precisión posible el estado de la máquina y su diagnóstico para poder actuar antes de la aparición de un fallo crítico. La aplicación de este tipo de métodos se denomina habitualmente *monitorización de la condición*<sup>1</sup> o, simplemente, *supervisión*. Es, precisamente, éste el campo en el que se centra el contenido de la presente Tesis.

---

<sup>1</sup>Este término es una traducción literal del anglosajón *condition monitoring*. El diccionario de la Real Academia Española, recoge como acepciones de la palabra *condición*: “Estado, situación especial en que se halla una persona.” así como “Circunstancias que afectan a un proceso o al estado de una persona o cosa.”. A lo largo de esta Tesis se aplicará este término, por extensión, a procesos, entendiendo con él el “estado” o situación actual del mismo.

## 1.2 Principales métodos de Detección y Diagnóstico de Fallos

### 1.2.1 Métodos basados en Umbral.

Quizás, los métodos más elementales –aunque, probablemente, los más utilizados, por su sencillez y bajo coste– en la supervisión de procesos son los *métodos basados en umbral*. Su principio básico consiste en el establecimiento de una serie de *umbrales* o *límites* para ciertas variables del proceso entre los cuales pueden oscilar éstas cuando el proceso se encuentra en condiciones de funcionamiento normales. La violación de alguno de estos umbrales refleja, por tanto, una situación anómala en el proceso que permite actuar. Este tipo de métodos, sin embargo, padece grandes limitaciones:

- Por lo general, presentan un mal compromiso entre márgenes estrechos (falsas alarmas) y márgenes anchos (insensibilidad a fallos).
- A menudo un mismo fallo se manifiesta en varias medidas lo que causa varios disparos [2]. Esto hace complicado el aislamiento de los defectos y la interpretación de las alarmas<sup>2</sup>.
- Perturbaciones ajenas a los fallos pueden causar falsas alarmas que es necesario descartar.

Este tipo de métodos utiliza un precario grado de conocimiento del proceso, de carácter *apriorístico*, a menudo basado en la experiencia, relativo a los umbrales de variación de las variables. No obstante, merece especial mención su aplicación a la monitorización de los espectros de vibración en máquinas eléctricas [1], de uso muy extendido en la industria. Para ello, se predeterminan unos niveles operacionales, generalmente tomando como base la experiencia o los estándares internacionales, construyendo una envolvente del espectro que no debe ser sobrepasada. Pese a su sencillez, el análisis de los niveles del espectro de vibración ofrece

---

<sup>2</sup>Según Sachs [3], los sistemas de instrumentación y control de una planta nuclear pueden tener en torno a unas 2000 alarmas en una misma sala de control además de los displays con datos analógicos. En la simulación de un accidente de pérdida de refrigerante, 500 luces se encendieron o apagaron durante el primer minuto y 800 en el segundo.

una enorme dosis de información sobre el estado de la máquina. Diversos autores han analizado la relación entre diversos tipos de fallos y las frecuencias alteradas en el espectro para las máquinas eléctricas rotativas más frecuentes, excelentemente resumidas por Tavner en [1].

### 1.2.2 Métodos basados en Modelos.

Los métodos basados en modelos se basan en la generación de uno o varios modelos de comportamiento del proceso o de partes del mismo que permiten extraer conclusiones a partir de las discrepancias entre el proceso real y los modelos.

En general, los métodos basados en modelos reúnen importantes ventajas:

- Utilizan información *a priori* del proceso lo que evita la necesidad de extraerla a partir de datos experimentales. Además, pueden incorporar al sistema de diagnóstico información sobre el proceso no disponible a través de datos experimentales.
- La utilización de modelos matemáticos, puede permitir hacer insensible el sistema de diagnóstico ante perturbaciones no medidas [2].
- Mejoran ostensiblemente el compromiso entre márgenes estrechos o amplios para los umbrales de alarma, lo que los hace más selectivos.

Los métodos basados en modelo, sin embargo, presentan algunos inconvenientes:

- La eficacia de este tipo de métodos tiene una fuerte *dependencia del modelo*, [4], [5], [6]. El modelo debe incluir todas las situaciones a estudiar y debe ser capaz de considerar cambios en el punto de trabajo.
- Asimismo, debe considerarse la *sensibilidad* del sistema de diagnóstico frente a errores en el modelado. Debe determinarse el rango de validez del modelo, que podría no ser válido si el proceso está sujeto a variaciones no consideradas en el mismo, [5], [6].

- A menudo, este tipo de métodos presentan una elevada carga computacional, lo que puede dificultar su aplicación en sistemas de diagnóstico on-line y que requieran bajos tiempos de respuesta (e.g. sistemas de vuelo, etc.)
- El desarrollo de los modelos, especialmente en el caso de procesos moderadamente complejos, puede requerir un enorme esfuerzo e incluso resultar imposible o inviable [4]. Esto limita su rango de aplicación a procesos de complejidad moderada o a partes de procesos.
- Por la misma razón, este tipo de métodos son, en ocasiones, poco flexibles y difícilmente adaptables, pudiendo resultar muy costosa la reconfiguración de los modelos ante cambios en la dinámica del proceso.

Existen diversas posibilidades dentro de los métodos basados en modelo. A continuación se citan algunas de las más importantes:

**Métodos de Redundancia Hardware.** Los *métodos de redundancia hardware*, utilizados desde hace décadas, consisten en duplicar o triplicar *físicamente* un componente crítico de un sistema en orden a verificar su correcto funcionamiento a partir las diferencias entre las réplicas. La doble redundancia permite detectar un fallo, aunque no permite discernir cuál es el componente dañado. La redundancia triple (TMP ó *Triple Modular Redundancy*), por el contrario, permite reconocer el componente dañado y además sustituirlo de inmediato por alguno de los dos restantes. Ocasionalmente, en sistemas especialmente críticos, se utiliza redundancia múltiple (cuatro o más réplicas de un mismo componente), utilizando, por ejemplo, reglas de decisión de votado.

**Métodos de Redundancia Analítica.** Los *métodos de redundancia analítica*, consisten en la generación de modelos matemáticos del comportamiento del proceso, o de partes del mismo para hacerlos evolucionar en paralelo con el proceso real. Las desviaciones entre las variables medibles del proceso y las variables generadas por los modelos, denominadas *residuos*, son altamente significativas en cuanto al diagnóstico del proceso, debido a que la información relativa al comportamiento predecible del

mismo está ya desacoplada. Uno de los objetivos principales en este tipo de métodos consiste en la generación de residuos sensibles a un conjunto concreto de fallos e insensibles a otros fallos y al ruido [6]. Este tipo de métodos suele requerir una etapa posterior, que implicaría la *evaluación* de los residuos generados, bien por procedimientos estadísticos o mediante un conjunto de reglas, para la generación de diagnósticos.

Se han propuesto diversos métodos para la generación de residuos en la literatura tales como las *ecuaciones de paridad* [7], [8], [2], los *observadores* [2], o las redes neuronales [9], [10], [11].

**Modelos basados en Redes Neuronales.** Aunque, por lo general, los métodos de redundancia analítica suelen emplear modelos matemáticos o físicos obtenidos a partir de información derivada del conocimiento *a priori* del proceso, la filosofía y metodología propia de éstos permite combinarlos con modelos basados en el aprendizaje a partir de ejemplos, como las redes neuronales, que aportan información *a posteriori*, derivada a partir de los ejemplos. En [9], por ejemplo, Sanz-Bobi y Donaire utilizan el concepto de redundancia analítica para el diagnóstico de fallos de un motor de inducción empleando modelos basados en *redes neuronales feedforward* que relacionan los armónicos de la corriente con la carga, los armónicos de la vibración axial en ambos extremos del motor, etc. analizando posteriormente los residuos generados por las diferencias entre los valores reales y los predichos por los modelos.

Otra aplicación del concepto de redundancia analítica mediante modelos basados en *redes neuronales feedforward autoasociativas* se detalla en [10], donde Petsche et al. entrenan un autoasociador para reconstruir el espectro de un motor sano. Los residuos obtenidos a partir de una medida de la disparidad entre el espectro real y el estimado permiten detectar la presencia de anomalías.

En [11], Wilson e Irwin describen un método basado en redes RBF que permite modelar a partir de ejemplos la geometría de los datos de proceso y proyectar sobre un espacio de visualización. Su método permite también proyectar hacia atrás los puntos del espacio de visualización de nuevo hacia el espacio de los datos, obteniendo así una estimación del dato original según el modelo. La diferencia entre un dato y su “retroproyección”, constituye un residuo de gran utilidad en la detección de anomalías. En esta Tesis se propone una arquitectura –descrita con

más detalle en 3.8– basada en el SOM (*Self Organizing Map*) y regresión de núcleo (en realidad, una forma regularizada de red RBF) para el modelado de los datos de proceso. Los residuos generados por la diferencia entre los datos actuales y la predicción del modelo aportan una información que puede resultar extremadamente valiosa en la detección y diagnóstico de fallos como se describe en 5.6 y en el capítulo 6, de resultados experimentales.

Debe aclararse, no obstante, que aunque las redes neuronales se describen aquí como un modelo más del proceso, lo que permitiría considerarlas en el marco de los *métodos basados en modelos*, el origen de la información que permite construir dicho modelo parte de ejemplos consistentes en datos de proceso, por lo que hemos preferido encajarlas dentro de los *métodos basados en datos* expuestos más adelante.

**Detección de Fallos basada en Estimación de Parámetros.** Aunque este tipo de métodos podría encuadrarse, en cierto modo, dentro de los métodos de redundancia analítica, pueden ser citados aparte al haber recibido considerable atención en la literatura de detección y diagnóstico de máquinas eléctricas. Estos consisten en la estimación de uno o varios parámetros críticos del proceso, que habitualmente tienen un significado físico definido que permite diagnosticar de forma evidente ciertos tipos de anomalías. En [12], por ejemplo, Cabanas et al. proponen una metodología basada en la FFT para la determinación de la impedancia de secuencia inversa, parámetro directamente relacionado con la presencia de cortocircuitos entre espiras del estátor en motores de inducción. En diversos trabajos, como en [13], se ha demostrado que la presencia de determinados defectos en el rotor de un motor asíncrono se traducen en una variación de ciertos parámetros como la inductancia y la resistencia del rotor, lo que sugiere un diagnóstico basado en la estimación de éstas. En esta línea, Beguenane [14] propone un método sencillo pero eficaz para identificar la resistencia del rotor con el objeto de determinar la condición térmica de un motor de inducción. Cho et al. [15], por otra parte, aplican un observador de la resistencia del rotor para la detección de asimetrías rotóricas. Más recientemente, Moreau et al. [16], aplican el mismo concepto observando la resistencia y la inductancia de dispersión del estator para las tres fases así como la resistencia e inductancia del rotor, pudiendo detectar y diagnosticar diversos tipos de fallos e incluso

localizar su procedencia.

**Modelos basados en técnicas de Razonamiento Cualitativo.** Los modelos empleados en los métodos anteriores se denominan *métodos cuantitativos* ya que manejan variables con valores perfectamente definidos. La precisión numérica de los métodos cuantitativos puede presentar diversos inconvenientes:

- A menudo generan flujos de datos enormes, a veces no asimilables por un humano.
- Exigen asumir cantidades exactas y precisas en las variables, lo que no siempre constituye una correcta representación de la realidad.
- Casi siempre requieren una fase posterior que genere información fácilmente asimilable por el hombre (en forma gráfica, mediante sentencias, etc.).

El *Razonamiento Cualitativo* (*QR*, o *Qualitative Reasoning*) persigue el desarrollo de técnicas de representación y razonamiento que permitan extraer conclusiones acerca del comportamiento de sistemas físicos en ausencia de información precisa<sup>3</sup> [17]. El razonamiento cualitativo permite describir procesos físicos sin necesidad de recurrir a valores exactos de las variables. Kuipers, unos de los pioneros en razonamiento cualitativo, describe muy esquemáticamente el razonamiento cualitativo en [18] comparándolo con una bola lanzada hacia arriba. Una descripción cualitativa podría ser:

*La bola sube durante un tiempo, reduciendo su velocidad,  
hasta detenerse y posteriormente desciende, incrementando  
velocidad.*

---

<sup>3</sup>El concepto de *imprecisión* manejado en el Razonamiento Cualitativo, según Iwasaki [17], no debe confundirse con el de *incertidumbre*, manejado en la teoría de la probabilidad y la lógica difusa. Mientras la incertidumbre expresa un grado de creencia o de pertenencia a un conjunto, la imprecisión no es incierta. Así, alguien puede tener la completa certeza (100 %) de que la temperatura del agua de mar en el norte de España durante los meses de verano está entre 5<sup>0</sup>C y 35<sup>0</sup>C. Sin embargo, en Lógica Borrosa, las expresiones pueden ser parcialmente ciertas y en Probabilidad un hecho puede ser más o menos probable.

La descripción omite los detalles, centrándose en el aspecto cualitativo del fenómeno, y no describe la velocidad, posición o aceleración exactas para un instante dado. En el modelado y diagnóstico de procesos, el razonamiento cualitativo cuenta con varias ventajas [17]:

- Capacidad de manejar información incompleta, característica especialmente útil cuando no se dispone de datos cuantitativos precisos del proceso o cuando éstos son muy costosos de obtener.
- Genera predicciones cualitativas, más próximas al entendimiento y forma de pensar humanos.
- El modelado y simulación cualitativos permiten generar todo el abanico de comportamientos del proceso que se derivan del conocimiento impreciso (aunque cierto) disponible sobre éste. Ello permite determinar dinámicamente los rangos de validez de los distintos comportamientos, de manera similar aunque mucho más potente que los métodos de umbral.

El razonamiento cualitativo hace uso del *cálculo cualitativo*, una serie de técnicas paralelas a las del cálculo convencional, que permiten trabajar con información imprecisa. Una línea de investigación de gran interés en este campo es la combinación del razonamiento cualitativo, la inferencia semicuantitativa y la simulación numérica, [19],[18]. En el *razonamiento semicuantitativo*, la imprecisión es descrita habitualmente a través de intervalos para las variables. Según un razonamiento de tipo cuantitativo, por ejemplo, si se sabe que  $z = x + y$ , y las variables  $x$  e  $y$  vienen dadas de forma imprecisa mediante intervalos,  $x \in [a, b]$  é  $y \in [c, d]$  puede hallarse la variable  $z$  con su rango posible de variación,  $z = x \oplus y = [a + c, b + d]$ . Análogamente, pueden definir las operaciones  $\otimes, \ominus$ , etc.

En [20], Vescovi et al. emplean el método NIS (*Numerical Interval Simulation*), un método de simulación que permite simular sistemas dinámicos a partir de variables definidas mediante intervalos, dando como solución variables con sus rangos posibles de variación. Para ello emplea las operaciones aritméticas con intervalos y redefine los métodos de integración numérica (Euler, Runge Kutta, etc.) en términos de intervalos, lo que permite aportar soluciones acotadas por rangos a las ecuaciones diferenciales. Los rangos de las variables devueltas por la simulación varían dinámicamente empleando la información disponible del modelo, lo que

permite una detección de fallos similar a la de umbral, pero mucho más selectiva, ya que los umbrales admisibles son establecidos dinámicamente empleando la información del modelo.

El razonamiento cualitativo, sin embargo, cuenta con desventajas similares a las de los métodos de modelado cuantitativo, principalmente derivadas de la dificultad en el modelado de sistemas complejos.

### 1.2.3 Métodos basados en Datos.

Los *métodos basados en datos*<sup>4</sup>, a diferencia de los *métodos basados en modelos*, no requieren la obtención previa (*a priori*) de un modelo matemático del proceso. Este tipo de métodos, por el contrario, extrae la información sobre el proceso *a posteriori*, a partir de medidas procedentes de éste. Esta información permite elaborar un *mapa* del espacio de medidas sobre un *espacio de decisión* [5].

Los métodos basados en datos constan, habitualmente, de tres partes bien definidas: adquisición de medidas, extracción de características y sistema de decisión. Como se verá más adelante, la elección del sistema de medidas y la extracción de características son de importancia esencial en el rendimiento de este tipo de métodos y de ellas depende gran parte de su éxito. Pese al hecho de que el grueso de la información requerida para modelar el proceso se obtiene *a posteriori*, a partir de ejemplos, en el diseño de las dos primeras etapas –especialmente en la adquisición de medidas– suele ser de vital importancia el empleo de una mínima información *a priori* sobre el proceso, como puede ser el determinar la posición adecuada de los sensores de vibraciones, el tipo de medidas que más información ofrecen sobre los fallos, etc.

La etapa de decisión, finalmente, transforma las características del proceso en un punto del *espacio de decisión*. El espacio de decisión puede ser:

---

<sup>4</sup>Este tipo de métodos son también conocidos como métodos *libres de modelos* o, en inglés, *model-free*. Se ha preferido aquí el término “basados en datos” porque, de hecho, estos métodos también utilizan un modelo. La diferencia fundamental reside en la procedencia de la información para la elaboración del modelo. En los métodos basados en modelos la información es *apriorística*, procedente de la experiencia, del conocimiento previo del proceso o de las leyes de la física, en base a lo cual se elabora un modelo matemático que se supone que se ajusta al proceso. En los métodos basados en datos, la información procede de los datos del proceso, a partir de la cual elaboran un modelo (estadístico) que se ajusta el proceso.

- a) Un conjunto discreto, formado por una colección de *clases* (normal, fallo 1, fallo 2, etc.), para lo que se emplean habitualmente *técnicas de clasificación*,
- b) Un espacio continuo, recurriéndose en este caso a *técnicas de reducción de dimensionalidad*.

Esta última alternativa, relativamente reciente, consiste en la generación de proyecciones o *mapas* no lineales a partir de las medidas del proceso sobre un espacio de visualización (2D ó 3D). La esencia de este tipo de métodos estriba en preservar la información relativa a los distintos modos de funcionamiento del proceso durante la transformación. La consecución de este objetivo permite hacer patentes de forma asombrosamente intuitiva las distintas condiciones del proceso en un gráfico bidimensional o tridimensional en el que se representa una proyección no lineal de la trayectoria del estado del proceso.

El trabajo más significativo dentro de este grupo de métodos es, quizás, la metodología para el análisis y modelado de procesos complejos desarrollada por el grupo de Kohonen, [21], [22] y [23], en la que se inspira buena parte del trabajo presentado en esta Tesis. En esta metodología se aplica el *self organizing map* (SOM)<sup>5</sup>, una red neuronal no supervisada, capaz de “captar” la geometría y estructura de los datos generados por el proceso y que permite visualizar la información adquirida en mapas extraordinariamente intuitivos.

Esta idea ha inspirado diversos trabajos. Así, Göppert et al., [24], [25], [26], demuestran con éxito la aplicación del SOM en la detección y evaluación de la concentración de gases y disolventes orgánicos mediante espectros de interferencia.

En una línea similar, Svensén propone en su tesis doctoral [27] una técnica de proyección denominada GTM (*Generative Topographic Mapping*), de propiedades similares al SOM, aunque fundamentada en principios estadísticos. A diferencia del SOM, el GTM genera mapas continuos, lo que constituye una importante ventaja sobre el SOM. Svensén aplica esta técnica en el análisis de datos simulados de medidas no intrusivas mediante gamma-densitometría en una tubería que transporta una mezcla con proporciones variables de gas, agua y aceite, detectando las

---

<sup>5</sup>En la sección 3.8 y en los capítulos 4 y 5 se analizan en detalle las propiedades del SOM así como su importancia y utilidad en la supervisión de procesos.

concentraciones de los distintos componentes así como las configuraciones multi-fase (mezcla homogénea, anular o estratificada) en las que se produce el transporte.

En [11], Wilson e Irwin describen un método de monitorización de procesos basado en una extensión no lineal del método PCA utilizando funciones de base radial. Las primeras componentes principales permiten describir las características más importantes del proceso y representar su evolución en un espacio de visualización. En definitiva, el método define una proyección continua del espacio de datos sobre el espacio de visualización preservando la información más relevante.

Finalmente, varios resultados de esta Tesis se han presentado en [28] y en [29]. En [28] se realiza un estudio comparativo de diversas estrategias de interpolación aplicadas al SOM, proponiéndose finalmente la regresión de núcleo y demostrándose ésta sobre datos experimentales procedentes del mismo motor de inducción del Caso 1 en el capítulo 6. En [29] se aplica el SOM con interpolación de regresión de núcleo a datos procedentes de un motor de 7000 kW empleado en un tren de laminación de Aceralia.

### 1.3 Objetivos de la Tesis

Parte del desarrollo de esta Tesis se enmarca dentro de un proyecto de investigación europeo (programa CECA) de mantenimiento predictivo que el Área de Ingeniería de Sistemas y Automática de la Universidad de Oviedo está llevando a cabo para grupos transformador-rectificador-motor CC del tren semicontinuo en las instalaciones que Aceralia tiene en su factoría de Avilés. Este proyecto persigue el desarrollo de un sistema de captura y monitorización del estado de los equipos de laminación que permita a los técnicos de mantenimiento visualizar de una forma sencilla el estado de los equipos, así como su evolución y su posible estado de funcionamiento.

El alto grado de complejidad del sistema, unido al elevado número de variables que es necesario analizar, hace complicada una solución basada en reglas o en modelos físicos demandando, por el contrario, métodos de aprendizaje. Al mismo tiempo, el estado del proceso es un ente continuo, que puede derivar de una situación normal a una situación de fallo de manera paulatina, puede mostrar tendencias e incluso puede describir

trayectorias cuya geometría puede aportar una información esencial sobre el proceso. Por las características del problema, parecía necesario orientar el esfuerzo de investigación en el desarrollo de técnicas:

- a) Capaces de *modelar el comportamiento* del proceso a partir de ejemplos (medidas del proceso).
- b) Que permitan evaluar la condición del proceso de *manera continua*.
- c) Capaces de representar la información sobre el estado del proceso *de manera gráfica* o visual.

Esta Tesis se centró en el campo de los *métodos basados en datos*, citados en la sección anterior. Con el objeto de delimitar con más detalle el marco en el que se centra la investigación llevada a cabo en esta Tesis, en la siguiente sección se formula con detalle el problema investigado en esta Tesis.

## 1.4 Formulación del Problema

En esta sección el problema de la monitorización de la condición de procesos se plantea bajo el enfoque de la Teoría de Sistemas. Desde esta óptica, el problema puede formularse en términos de las regiones atravesadas por el estado de un sistema dinámico complejo en el espacio de estados.

Bajo ciertas condiciones, exigidas al sistema de medidas, el problema puede reformularse en el espacio de salida en términos de las regiones atravesadas por el vector de medidas, salvando así el problema de la observabilidad del estado y reduciendo la complejidad y dimensionalidad del problema.

Finalmente, se sugiere el problema fundamental a cuyo análisis y solución se dedica esta Tesis: *la búsqueda de un mapa continuo que, preservando la información sobre la condición, proyecte las medidas sobre un espacio visualizable (2D ó 3D), de manera que el problema pueda seguir formulándose en los mismos términos de trayectoria y regiones.*

En estas condiciones, la totalidad del proceso (sistema de medidas y mapa de visualización) constituye en sí mismo un mapa que proyecta el espacio de estados sobre un espacio de visualización, preservando la topología de las regiones y, en definitiva, la información relativa a la condición.

### 1.4.1 La monitorización de la condición en el marco de la Teoría de Sistemas

**Descripción del Proceso.** En general, una máquina o proceso puede concebirse [30] como un sistema altamente complejo, en el que gran cantidad de variables evolucionan en el tiempo con arreglo a leyes físicas a menudo no lineales y con un alto grado de interacciones. Según la Teoría de Sistemas, tal proceso puede describirse según:

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t), t) \\ y(t) &= g(x(t), u(t), t)\end{aligned}\tag{1.1}$$

donde,

- $x(t)$  es el *vector de estados* o, simplemente, *estado* del proceso y describe por completo la situación del sistema en un instante dado  $t$ .
- El conjunto de todos los estados posibles del proceso se denomina *espacio de estados*  $X \subset \mathbb{R}^n$ . El espacio de estados engloba, por tanto, todas las situaciones posibles en las que puede encontrarse el proceso. Cada punto  $x \in X$  representa una situación única e irrepetible del proceso.
- $u(t)$  es el *vector de entradas*, perteneciente al *espacio de entradas*  $E \subset \mathbb{R}^m$  y considera todos los factores externos que influyen en el proceso.
- $y(t)$  es el *vector de medidas*, perteneciente al *espacio de salidas*  $S \subset \mathbb{R}^p$ , formado por todas las medidas (o transformaciones de éstas) realizadas en el proceso en el instante  $t$ .
- $f$  y  $g$  son en general funciones no lineales; a  $f$  se la conoce como *función de estado* y describe la dinámica del proceso; a  $g$  se la conoce como *función de observación* y describe cómo se manifiesta externamente el proceso a través de las medidas. Estas funciones se suponen suficientemente complejas como para describir todo el abanico de comportamientos posibles del proceso.

**Condición de un Proceso.** Bajo el enfoque de la Teoría de Sistemas, puede definirse la *condición* de un proceso como un conjunto de estados del mismo que se etiquetan bajo el mismo nombre<sup>6</sup>.

Supóngase que existen catalogadas  $K$  condiciones posibles en las que un proceso dado puede funcionar, a saber,  $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_K$ , de manera que el proceso siempre se encuentra en una y sólo una<sup>7</sup> de ellas. El problema

---

<sup>6</sup>Una condición de funcionamiento no tiene por qué implicar un único estado. De hecho, por ejemplo, es frecuente definir las condiciones de funcionamiento mediante rangos para determinadas variables del proceso. Así, por ejemplo, un motor asíncrono puede considerarse en una condición “normal” de funcionamiento en un rango razonable de niveles de carga: se trata de distintos estados de funcionamiento del mismo, pero todos ellos pueden considerarse “normales”.

<sup>7</sup>Si existiese algún estado del proceso compatible con dos o más condiciones simultáneamente, o si existiesen estados no catalogados, se definirían nuevas condicio-

de la monitorización de la condición consiste entonces en determinar en cuál de las  $K$  condiciones se encuentra el proceso en un instante dado  $t$ .

Denotaremos por  $C_i \subset X$  el conjunto de todos los estados  $x(t)$  en los que el proceso puede considerarse funcionando en la condición  $\mathcal{C}_i$

$$C_i = \{x(t) \in X \mid \text{proceso funciona en condición } \mathcal{C}_i\} \quad (1.2)$$

Dado que el proceso no puede estar en dos situaciones  $\mathcal{C}_i$  y  $\mathcal{C}_j$  de manera simultánea, las regiones  $C_i$  serán disjuntas

$$C_i \cap C_j = \emptyset, \quad \forall i \neq j \quad (1.3)$$

Análogamente, el proceso debe estar al menos en alguna de las  $K$  situaciones predefinidas

$$X = \bigcup_{i=1}^K C_i \quad (1.4)$$

En definitiva, los conjuntos  $C_i$  constituyen una *partición* del espacio de estados  $X$ . Bajo este punto de vista, el problema de la monitorización de la condición se desglosa dos subproblemas:

- a) La determinación de las regiones  $C_i$  de  $X$  correspondientes a cada una de las condiciones  $\mathcal{C}_i$ .
- b) La observación del estado  $x(t)$ .

### 1.4.2 Influencia del sistema de medidas en la observabilidad

Evidentemente, el conocimiento del estado  $x(t)$  así como de las regiones  $C_i$  de  $X$  correspondientes a cada una de las condiciones que se pretenden evaluar resolvería el problema de la monitorización de la condición.

---

nes para describirlos unívocamente (e.g. definiendo nuevas condiciones, como “proceso con fallos  $i$  y  $j$ ” ó “proceso en situación no catalogada”).

Lamentablemente no siempre es posible realizar un seguimiento del vector de estado a partir de las medidas disponibles, ya sea por la complejidad de las leyes que rigen el proceso (funciones  $f$  y  $g$ ), ya por no ser éste completamente observable. Sin embargo, el conocimiento completo del vector de estado, aunque deseable, no siempre es imprescindible. Las situaciones del proceso que se pretenden identificar, a menudo se corresponden con regiones bien definidas en el espacio de estados que describen su comportamiento general y que, en cierto modo, se hacen patentes en las medidas [21].

La cantidad de información sobre el estado presente en las medidas depende exclusivamente de la función de observación  $g$ . Esta función está directamente relacionada con el número, calidad, tipo, posición, etc. de cada uno de los sensores aplicados al proceso para la toma de medidas. Si el sistema de medidas del proceso está bien diseñado,  $g$  preservará la información necesaria del estado para determinar su condición. Un diseño adecuado del sistema de medidas del proceso debe reunir dos condiciones básicas:

§1. *Las medidas no deben depender directamente de variables ajenas al estado del proceso.*

En efecto, una dependencia directa<sup>8</sup> de las medidas respecto a variables no ligadas al proceso, a saber, las entradas  $u(t)$  y el tiempo  $t$ , haría posible que un mismo estado de funcionamiento del proceso pudiera dar lugar a medidas distintas<sup>9</sup>. Formalmente, esta condición exige

$$y(t) = g(x(t)) \quad (1.5)$$

<sup>8</sup>Evidentemente, las entradas influyen en la dinámica del proceso y, por tanto  $x$  depende de  $u$  y de  $t$ , por lo que  $y$  también dependería *indirectamente* de éstas. Matizando, entendemos aquí por *dependencia directa* aquella que no se produce a través del propio estado  $x$ .

<sup>9</sup>Un caso muy típico de dependencia de las medidas respecto de las variables de entrada es la presencia del armónico de 50 Hz de alimentación en las medidas de los sensores.

Análogamente, la dependencia de  $t$  tiene que ver a menudo con la degradación de los sensores, su mala repetibilidad, etc. Un ejemplo típico se encuentra en los acelerómetros, en los que un afloje de las fijaciones o una ligera modificación en su emplazamiento pueden afectar drásticamente al sistema de diagnóstico.

es decir, las medidas deben depender del estado del proceso y sólomente de éste.

§2. *Las medidas deben suministrar información suficiente acerca de la condición del proceso.*

Esta condición exige la utilización, en la medida de lo posible, del conocimiento experto (*a priori*) del proceso en el diseño del esquema de medidas. La elección del tipo de señales a medir, ubicación y tipo de los sensores, etc. deben estar orientados a generar medidas significativas en cuanto a los modos de funcionamiento del proceso que se desean supervisar. Desde un punto de vista más formal, se trata de generar regiones disjuntas en el espacio de medidas recorrido por  $y(t)$  que se correspondan con cada una de las diferentes condiciones del proceso a supervisar.

$$g(C_i) \cap g(C_j) = \emptyset, \quad \forall i \neq j \quad (1.6)$$

Así, por ejemplo, en un motor de alterna en el que sólo se tomase la temperatura en los cojinetes (el espacio de medidas sería  $\mathbb{R}^1$ ) podría discriminarse una situación de fallo en los cojinetes frente a una situación normal, ya que ambos darían lugar a rangos de temperatura bien diferenciados (separables). Sin embargo, sería imposible diferenciar distintos tipos de asimetrías o desalineaciones en el eje, puesto que una misma temperatura podría corresponder a varios fallos distintos. Varias de estas clases aparecerían solapadas en el espacio de medidas  $\mathbb{R}^1$  recorrido por la temperatura en los cojinetes, con lo que la información necesaria para determinar la condición se habría perdido.

La verificación de las condiciones anteriores exige, por tanto, que:

- a) Las medidas sean función exclusivamente del estado (ec. 1.5)
- b) Dicha función conserve suficiente información acerca de todos los posibles estados del sistema que se pretenden determinar.

Bajo estas condiciones:

*Se induce en el espacio de salidas  $S$  una partición en  $K$  regiones disjuntas  $C'_i = g(C_i)$  que representan, respectivamente, a las  $K$  condiciones posibles  $C_i$  del motor.*

En definitiva, bajo tales hipótesis debe resultar *posible* (aunque no necesariamente fácil) determinar la condición del proceso a partir de las medidas. Puede concluirse, por tanto, que

*Con un buen sistema de medidas, el problema de la supervisión del estado de un proceso pasa de analizar la evolución del vector de estados, que no es necesariamente observable, a analizar directamente el vector de medidas.*

### 1.4.3 Proyección del espacio de medidas sobre un espacio visualizable

Asumiendo la viabilidad de la monitorización de la condición a partir de las medidas  $y(t)$ , cabe sugerir la posibilidad de generar matemáticamente un nuevo mapa  $\mathcal{S}$  que, verificando las dos condiciones esenciales, a saber, la no dependencia directa de variables ajenas y la preservación de la información relativa a la condición, genere una proyección de  $y(t)$  sobre un espacio visualizable  $V$  de dimension dos o tres.

La primera de las condiciones no plantea problema alguno ya que al tratarse de una expresión matemática obviamente exigiremos que no dependa de variables ajenas a las medidas.

El objetivo de la segunda condición consiste en generar una proyección  $\mathcal{S} : S \rightarrow V$  de manera que *preserve la información* relativa a la condición del proceso, es decir,

$$\mathcal{S}(C'_i) \cap \mathcal{S}(C'_j) = \emptyset \quad (1.7)$$

$$\bigcup_{i=1}^K \mathcal{S}(C'_i) = V \quad (1.8)$$

lo que induciría una nueva partición en  $K$  regiones  $C''_i = \mathcal{S}(C'_i)$  en el espacio de visualización. El problema de determinar la condición del proceso se reduciría entonces a visualizar un “puntero”  $p(t) = \mathcal{S}(y(t))$  sobre el espacio de visualización y determinar sobre cuál de las  $K$  regiones  $\{C''_i\}_{i=1,\dots,K}$  se posa.

Sin embargo, el cumplimiento de esta segunda condición no es, en absoluto, obvio y plantea dos problemas esenciales:

- Por una parte, el mapa  $\mathcal{S}$  debe *preservar la información* relativa a la condición del proceso, conservando intacta la partición de las  $K$  regiones.
- En segundo lugar, es necesario determinar las  $K$  regiones  $C''_i$  de  $V$ , al objeto de poder determinar en cuál de ellas se encuentra el puntero  $p(t)$ .

El primero de los problemas conecta directamente con un grupo de técnicas denominadas en la literatura *técnicas de reducción de la dimensionalidad* cuyo objetivo fundamental estriba en explicar un fenómeno

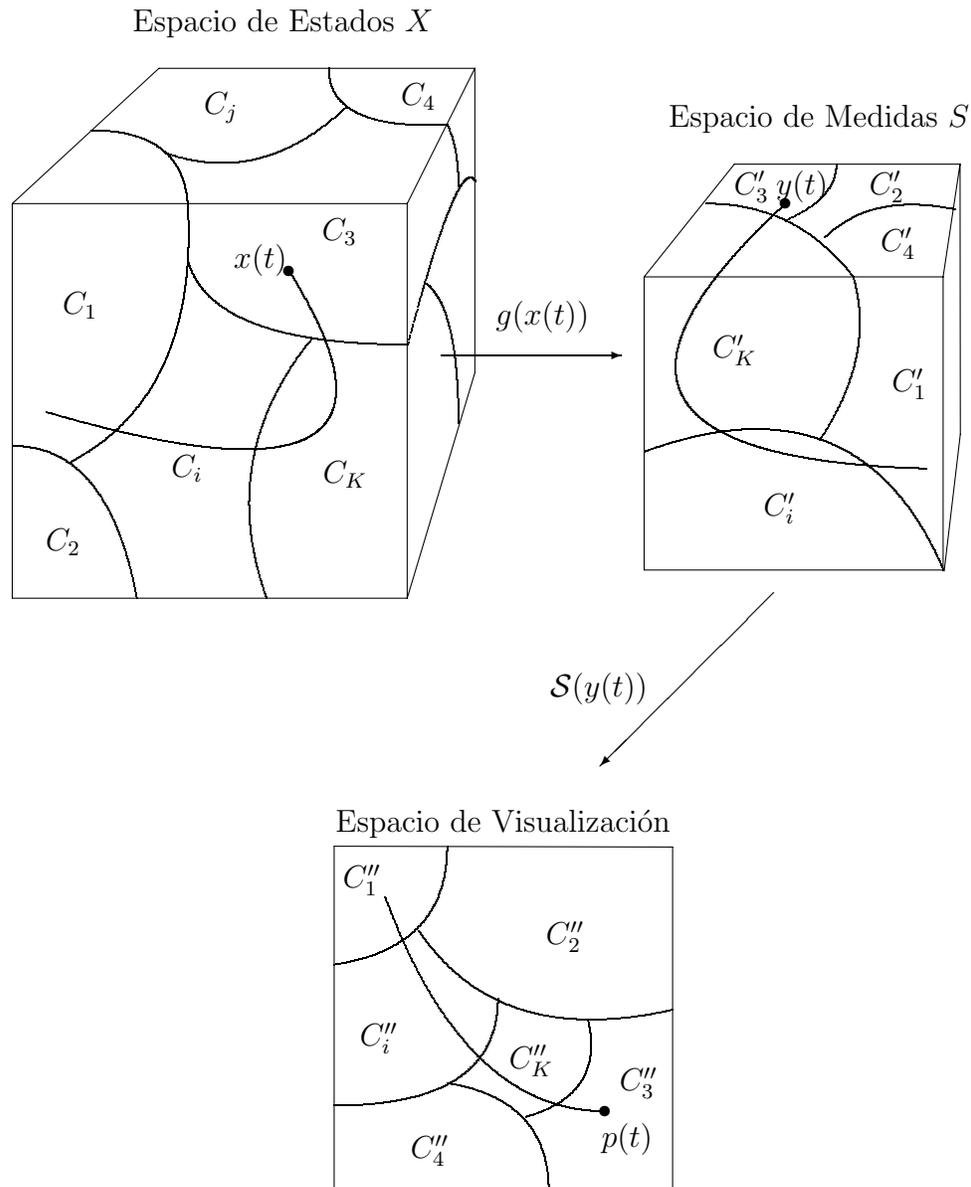


Figura 1.1: El problema de la monitorización de la condición se plantea como en términos de la obtención de un *mapa* entre el espacio de estados y un espacio de visualización (2D o 3D) que preserve la información relativa a la condición del proceso. El mapa es composición de dos mapas: un *mapa hardware*,  $g(\cdot)$ , dado por la ubicación y tipo de sensores y un *mapa software*,  $S(\cdot)$ , de carácter matemático, generado a partir de información disponible del proceso.

dado mediante un número inferior de variables. El segundo de los problemas podría enmarcarse dentro de las *técnicas de clasificación*. El problema de la clasificación, sin embargo, se ve a menudo enormemente simplificado si las técnicas de reducción de dimensionalidad son efectivas.

El contenido de esta Tesis se orienta esencialmente al análisis de éstas últimas. En el capítulo 2 se analizan las principales técnicas de extracción de características, que pueden considerarse como técnicas de reducción de dimensionalidad básicas que consituyen una etapa previa. Estas técnicas utilizan a menudo un conocimiento *a priori* de carácter físico o matemático para extraer las características más relevantes, reduciendo así el número de variables. En el capítulo 3 se abordan las técnicas de reducción de la dimensionalidad propiamente dichas. Estas técnicas suponen habitualmente (aunque no necesariamente) la existencia previa de una etapa de extracción de características y normalmente no asumen conocimientos *a priori* sobre los datos a procesar. Entre ellas se analiza en detalle el Mapa Topológico Autoorganizado (SOM), cuyas propiedades lo hacen muy interesante en relación con el planteamiento aquí propuesto. En el capítulo 4 se estudia la problemática derivada del carácter discreto del algoritmo SOM original y se analizan diversas alternativas para la generación de un SOM continuo, proponiéndose finalmente la interpolación mediante regresión de núcleo. En el capítulo 5 se propone una metodología basada en la generalización continua del SOM propuesta en el capítulo 4

Finalmente, en el capítulo 6 se aplica la metodología propuesta sobre datos reales vibraciones de un motor de laboratorio y sobre datos reales de un motor de laminación de 7000 kW perteneciente a la factoría de Aceralia en Avilés.

# Capítulo 2

## Técnicas de Extracción de Características

### 2.1 Introducción

Una vez resuelto el problema de la *viabilidad* o *posibilidad* de determinar la condición de un proceso a partir de las medidas, nada hace pensar que las regiones generadas en el espacio de medidas sean fáciles de determinar.

Considérese, por ejemplo, la señal  $a(t)$  procedente de un acelerómetro conectado a un motor, muestreada con un periodo de muestreo  $T_m$  convenientemente elegido, siendo el vector de medidas:

$$y(t) = [a(t), a(t - T_m), a(t - 2 \cdot T_m), \dots, a(t - (n - 1) \cdot T_m)]^T$$

Eligiendo un número  $n$  de muestras suficientemente elevado, los elementos del vector  $y(t)$  permiten discriminar situaciones correspondientes a diferentes formas de onda en  $a(t)$ , tales como un estado de funcionamiento normal, presencia de asimetrías mecánicas, fallos en la alimentación de una de las fases, etc. Sin embargo, las trayectorias descritas por el vector  $y(t)$  para cada una de las condiciones pueden tener geometrías muy complejas que distan mucho de ser fácilmente tratables.

Se hace casi imprescindible un proceso, posterior a las medidas, que transforme las medidas del proceso en otro conjunto de variables o *vector de características* que verifique, además de las citadas en el capítulo anterior, las siguientes condiciones:

- a) Debe explicar el fenómeno a analizar con el mínimo número posible de variables (principio de *empaquetamiento de la información*).
- b) Las características deben ser, en la medida de lo posible, invariantes para un mismo estado o condición del proceso.

Nótese que esto no exige que cada característica tome un valor diferente para cada condición del proceso, lo que representaría una exigencia más fuerte, sino que sólo permite cambios de éstas cuando la condición del proceso cambia. Bajo esta hipótesis, las regiones asociadas a cada condición ocupadas por el *vector de características* del proceso, se reducirían a puntos (nubes de puntos, si se consideran factores aleatorios) en el *espacio de características*.

En el caso anterior, por ejemplo, la simple aplicación de una DFT<sup>1</sup> con un enventanado al vector  $y(t)$ , permite desacoplar en cierta medida la variación periódica de  $y(t)$ , dando lugar a un vector de características poco dependiente del tiempo pero dependiente de la forma de onda de  $a(t)$ . En otras palabras, permite obtener puntos en el espacio de características relativamente invariantes para una misma condición del proceso.

Las transformaciones de las medidas orientadas a verificar las hipótesis citadas más arriba no son, en absoluto, evidentes y son objeto de una enorme y variada cantidad de técnicas conocidas por el nombre de *extracción de características*, orientadas a determinar, mediante transformaciones matemáticas de las medidas, un conjunto de características (*vector de características*) idóneo para su posterior tratamiento mediante un sistema de inferencia. A lo largo del capítulo nos limitaremos a destacar las técnicas más importantes, centrandó más la atención sobre aquéllas orientadas al mantenimiento predictivo de accionamientos eléctricos.

## 2.2 Selección de las señales más significativas

Una de las formas más elementales –aunque no por ello innecesarias– de extracción de características consiste en la selección, basada en la experiencia y el conocimiento de la física del proceso, de las variables más

---

<sup>1</sup>*Discrete Fourier Transform*, descrita más adelante en este mismo capítulo

significativas en relación a las características que se pretenden determinar.

Dentro del campo de las máquinas eléctricas, las *vibraciones* de una máquina aportan gran cantidad de información sobre el estado de la máquina y han sido, por ello, objeto de intenso estudio. Muchos tipos de defectos frecuentes en maquinaria producen cambios en las vibraciones que, convenientemente detectados y clasificados, permiten elaborar un diagnóstico [1] [5] [31]; el esfuerzo científico orientado al análisis de vibraciones como elemento de diagnóstico es tal que casi podría afirmarse que ésta es la señal más utilizada para determinar la condición de la máquina.

Asimismo, la *temperatura* en determinados puntos suele ser otro de los parámetros afectados por fallos de muy diversos tipos [1] [21]; la termografía explota al máximo la influencia de los fallos en la temperatura para determinar el origen y causa de éstos [32].

Otra de las señales que revelan fallos incipientes en máquinas eléctricas es la forma de onda de las *corrientes* absorbidas, que también ha sido objeto de intenso estudio, revelando multitud de anomalías, como roturas en las barras de motores de jaula [13] [33], excentricidades en el rotor [34], [33], o asimetrías rotóricas [33] entre otras.

Otras señales que también han demostrado resultar útiles en la determinación de la condición de la máquina son el *flujo magnético axial* [1], [33], [35] y el número de *partículas* metálicas en el lubricante [31].

## 2.3 Técnicas de Normalización

Las técnicas de normalización son, en general, transformaciones muy sencillas de los datos (habitualmente, traslaciones, cambios de escala y, a lo sumo, rotaciones) que van dirigidas a homogeneizar los datos de manera que sus rangos de variación sean similares. Su utilidad se hace patente cuando existen variables con rangos de variación que difieren varios órdenes de magnitud. En estos casos, los errores relativos cometidos por algoritmos de etapas posteriores pueden ser de tal magnitud que incluso pueden llegar a ignorar información útil aportada por las medidas o características de variación más pequeña.

### 2.3.1 Normalización a $[-1, +1]$

Este tipo de normalización consiste en eliminar la media de cada característica y variar la escala de manera que todas las características varíen en el intervalo  $[-1, +1]$ . Concretamente, para  $N$  muestras de una característica dada  $x$ , y denotando  $\bar{x} = \frac{1}{N} \sum_k x_k$ , se tiene:

$$x'_k = \frac{x_k - \bar{x}}{\max\{x_k\} - \min\{x_k\}} \quad k = 1, \dots, N \quad (2.1)$$

### 2.3.2 Normalización de la varianza

Este tipo de normalización busca hacer que cada característica tenga media nula y desviación típica 1. Denotando  $\bar{\sigma} = \sqrt{\frac{1}{N-1} \sum_k (x_k - \bar{x})^2}$  se tiene

$$x'_k = \frac{x_k - \bar{x}}{\bar{\sigma}} \quad k = 1, \dots, N \quad (2.2)$$

### 2.3.3 Normalización PCA

Este tipo de normalización, más complejo, implica una rotación del vector de características, transformándolos de manera que todos ellos estén incorrelados. Dado un conjunto de vectores de características  $\mathbf{x}_k$  cuya media ha sido eliminada por substracción, la matriz de correlación  $\mathbf{R}_{xx}$  se define como

$$\mathbf{R}_{xx} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k \mathbf{x}_k^T \quad (2.3)$$

Sea  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  la matriz cuyas columnas son los vectores propios de  $\mathbf{R}_{xx}$ , y sea  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  la matriz diagonal formada con los valores propios. Entonces

$$\mathbf{x}'_k = \mathbf{U}^T \mathbf{x}_k \quad (2.4)$$

Este método puede ser utilizado también para eliminar componentes cuya contribución es pequeña o nula. En la sección 3.2 se describe este método con más detalle.

### 2.3.4 Esferificación

Este método es una combinación de la normalización PCA y la normalización de la varianza. Suponiendo vectores de características  $\mathbf{x}_k$  de media nula,

$$\mathbf{x}'_k = \Lambda^{-1/2} \mathbf{U}^T \mathbf{x}_k \quad (2.5)$$

Esta transformación da como resultado, además de la incorrelación de las características, una matriz de correlaciones igual a la matriz identidad. Aplicada a una normal multidimensional daría como resultado una distribución normal  $N(\mathbf{0}, \mathbf{I})$ , cuya simetría es esférica en  $\mathbb{R}^n$ .

## 2.4 Técnicas basadas en Análisis Espectral

### 2.4.1 Introducción

El análisis espectral es, probablemente, una de las técnicas de análisis más extendidas en el mantenimiento predictivo de máquinas eléctricas. Gran parte de las máquinas eléctricas son de funcionamiento cíclico, y contienen elementos rotatorios, que generan medidas que son periódicas mientras no varían las condiciones de funcionamiento de la máquina.

El análisis en el dominio de la frecuencia, aplicado a señales periódicas, permite describir señales cambiantes en el tiempo mediante características (frecuencias y potencia de la señal a dichas frecuencias) que son estáticas bajo las mismas condiciones de la máquina. En cierto modo, este tipo de análisis permite “desacoplar” las variaciones temporales de carácter periódico sustituyéndolas por una descripción basada en frecuencias y sus potencias con lo que, si la forma de onda no cambia, dicha descripción tampoco. El método estándar para determinar la estructura periódica de una señal es el *espectro de potencias* [31].

Sea  $F$  la frecuencia,  $F_m$  la frecuencia de muestreo y  $f = F/F_m$  la frecuencia normalizada. La transformada de Fourier aplicable a secuencias discretas infinitas, aperiódicas, de energía finita, se conoce como *Transformada de Fourier Discreta en el Tiempo* (DTFT):

$$X(f) = \sum_{k=-\infty}^{\infty} x_k e^{-j2\pi f k} \quad (2.6)$$

expresión que es periódica en  $f$ , con periodo 1.

Una forma de calcular la expresión anterior en un subconjunto de puntos equiespaciados del intervalo de frecuencias  $[0, 1)$  en el que se mueve  $f$  es la DFT (*Discrete Fourier Transform*). Dado un registro de  $N$  datos de una señal,  $\{x_k\}_{k=0, \dots, N-1}$ , se define la DFT como:

$$X(n) = \sum_{k=0}^{N-1} x_k e^{-j2\pi kn/N} \quad (2.7)$$

El cálculo de la DFT tiene un coste computacional elevado, del orden  $\mathcal{O}(n^2)$ ; sin embargo, existe un algoritmo muy extendido, conocido como FFT (*Fast Fourier Transform*), que explota las propiedades de simetría de la DFT, con un coste  $\mathcal{O}(n \log(n))$ .

Existen varios problemas prácticos a la hora de realizar el análisis frecuencial mediante la DFT. El principal problema reside en la disponibilidad de un registro limitado de datos. La aplicación de la DFT asume que la señal discreta  $x_k$  es periódica de periodo  $N$  lo que, en general, no es cierto. La aplicación directa de la DFT equivale a la aplicación de una *ventana rectangular* sobre la totalidad de la señal discreta lo que origina la aparición de lóbulos colaterales en la representación frecuencial, efecto denominado *derrame espectral* o *fenómeno de Gibbs*, muy bien documentado en la literatura (cfr. [36], [37], [38], entre otros). En procesos industriales es típica la presencia de ruido y componentes aleatorias en la señal el cual se hace patente en el resultado de la DFT,  $X(n)$ . Un problema adicional es el de la fase de las componentes periódicas de la señal. La fase de éstas varía de forma irregular en cada ventana de datos<sup>2</sup> con lo cual la fase de las componentes representadas en  $X(n)$  no es significativa.

Considerada  $x_k$  como una variable aleatoria, la DFT no es sino un estimador del espectro y el resultado de su aplicación es también una variable aleatoria. Surge, por tanto, la necesidad de aplicar estimadores del verdadero espectro de la señal con características deseables (insesgados,

---

<sup>2</sup>A menos que las ventanas se tomen de manera sincronizada con las componentes periódicas. Esta situación es la que se persigue con el *muestreo sincronizado*, en el que el inicio de la toma de cada ventana de datos se *sincroniza* con un evento periódico del proceso, con el que se relacionan las componentes más importantes a analizar. Un caso típico es la toma de medidas sincronizada con el paso por cero del rotor en un motor eléctrico.

consistentes, etc.). El problema se formula, pues, como un problema de estimación del espectro de  $x_k$ . El *espectro de potencia* de  $x_k$  se define como:

$$S_{xx}(f) = \sum_{k=-\infty}^{\infty} r_{xx}(k)e^{-2\pi kf} = E[X(f)X^*(f)] = E[|X(f)|^2] \quad (2.8)$$

donde  $r_{xx}(k) = E[x_{k+l}x_k], \forall k$  es la *secuencia de autocorrelación* de la señal  $\{x_k\}$ . Nótese que la secuencia de autocorrelación,  $-y$ , por tanto, la expresión (2.8)– no tiene carácter aleatorio. La cantidad  $|X(f)|^2$  representa la distribución de energía de señal en función de la frecuencia, de ahí que también se conozca la definición dada en (2.8) como *densidad espectral de energía (o de potencia)*.

Conocido el espectro de potencia, es posible determinar una colección de características  $\{A_1, \dots, A_m\}$  basándose en las amplitudes del espectro (densidades de potencia espectral) a una serie de frecuencias  $\{f_1, \dots, f_m\}$  en las que se sabe (por un conocimiento experto *a priori*) que el proceso se manifiesta de manera significativa respecto a los fallos o situaciones a detectar:

$$A_i = S_{xx}(f_i), \quad i = 1, \dots, m \quad (2.9)$$

Existen en la literatura numerosos métodos orientados a estimar  $S_{xx}(f)$  a partir de un registro limitado de datos. Gran parte de éstos se basan en la aplicación de ventanas distintas a la rectangular, cuyos efectos colaterales en el espectro son menores, así como en el promediado o filtrado de los sucesivos espectros  $X(f)$  obtenidos para cada ventana. Este tipo de métodos se conocen como *métodos no paramétricos*. Finalmente, existe otro gran grupo de métodos denominados *metodos paramétricos* en el que los datos a analizar se asumen generados según un modelo conocido cuyos parámetros se deben ajustar. Una comparativa exhaustiva de éstos está fuera del marco de esta tesis, pudiendo remitirse el lector a obras consolidadas en la materia (cfr. [36], [37], [38], etc.); no obstante, a continuación se describen brevemente los métodos más importantes dentro de ambas categorías.

## 2.4.2 Métodos no paramétricos

### Método de Bartlett

Este método también es conocido como *promediado del periodograma*. El periodograma es una estimación del espectro de potencia construido a partir de un registro limitado de datos:

$$\hat{S}_{xx}(f) = \frac{1}{L} \left| \sum_{k=0}^{L-1} x_k e^{-j2\pi f k} \right|^2 \quad (2.10)$$

El método de Bartlett se desarrolla en tres pasos:

- a) Divide la secuencia  $\{x_k\}_{k=0, \dots, N-1}$  en  $K$  segmentos no solapados de longitud  $M$ :

$$\{x_k\}^{(i)} = x_{k+iM} \quad i = 0, \dots, K-1; \quad k = 0, \dots, M-1 \quad (2.11)$$

- b) Evalúa el periodograma  $\hat{S}^{(i)}(f)$  de cada uno de los segmentos  
 c) Promedia los periodogramas

$$\hat{S}^B(f) = \frac{1}{K} \sum_{i=0}^{K-1} \hat{S}^{(i)}(f) \quad (2.12)$$

Este método permite reducir la varianza de  $\hat{S}^B(f)$  a costa de una reducción en la resolución espectral debida al uso de registros con tan sólo  $M$  muestras.

### Método de Welch

Este método se conoce también como *promediado de periodogramas modificados*. El método es una generalización del método de Bartlett, difiriendo de éste tan sólo en dos aspectos:

- a) La subdivisión en tramos de datos se realiza con solapamiento.  
 b) Se utiliza una ventana  $w_k$  para mejorar las propiedades espectrales del periodograma dando lugar al *periodograma modificado*:

$$\hat{S}_{xx}(f) = \frac{1}{L} \left| \sum_{k=0}^{L-1} x_k w_k e^{-j2\pi f k} \right|^2 \quad (2.13)$$

Respecto al método de Bartlett, este método mejora el compromiso entre resolución espectral y la varianza de la estimación.

### Método de Blackman-Tukey

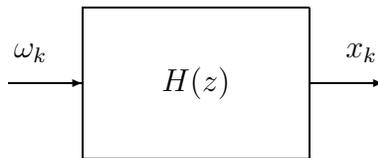
Este método se conoce también como *suavizado del periodograma*. La estimación del espectro dada por este método es la siguiente:

$$\hat{S}_{xx}^{BT}(f) = \sum_{m=-(M-1)}^{M-1} r_{xx}(m)w_m e^{-j2\pi f m} \quad (2.14)$$

donde  $r_{xx}(m)$  es la *secuencia de autocorrelación* que se define como  $r_{xx}(m) = E[x_{k+m}x_k]$ , pudiendo obtenerse una estimación  $\hat{r}_{xx}(m)$  de ésta por diversos métodos (cfr. [37], [36]). El método de Blackman-Tukey ofrece mejor compromiso entre varianza y resolución espectral que los métodos anteriores.

### 2.4.3 Métodos Paramétricos

Este tipo de métodos supone los datos a analizar  $\{x_k\}$  generados a partir de un modelo paramétrico cuya estructura (tipo y número de parámetros) se conoce pero cuyos parámetros hay que determinar. El modelo más típico es un filtro lineal  $H(z) = B(z)/A(z)$  alimentado por ruido blanco  $\omega_k$ , tal y como se aprecia en la figura



La ecuación en diferencias del modelo es

$$x_k = - \sum_{i=1}^p a_i x_{k-i} + \sum_{i=0}^q b_i \omega_{k-i} \quad (2.15)$$

Un proceso generado de acuerdo con (2.15) se conoce, en general, como *proceso autorregresivo de media móvil de orden  $p, q$* , denotándose por  $\text{ARMA}(p, q)$ . En particular, los modelos con  $q = 0$  y  $b_0 = 1$  constituyen una importante subclase de éstos y se conocen como *modelos autorregresivos* denotándose por  $\text{AR}(q)$ . Si se considera a las señales  $x_k$  y  $\omega_k$  como procesos aleatorios estacionarios, el teorema de *Wiener-Khintchine* que la densidad espectral de potencia de  $x_k$

$$S_{xx}(f) = S_{\omega\omega}(f)|H(f)|^2 \quad (2.16)$$

dado que  $\omega_k$  es un proceso aleatorio de ruido blanco se tiene que  $S_{\omega\omega}(f) = \sigma_\omega^2$ , de donde se deduce que,

$$S_{xx}(f) = \sigma_\omega^2 |H(f)|^2 \quad (2.17)$$

La ecuación anterior revela que el espectro de potencia de  $x_k$  puede determinarse sencillamente a partir de los parámetros  $a_k, b_k$  del modelo  $H(z)$ . Existen diversas técnicas para estimar los parámetros  $a_k$  y  $b_k$  conocido el orden  $(p, q)$  del modelo.

Los modelos  $\text{AR}(q)$  presentan una especial facilidad para el cálculo de los parámetros planteándose el problema en términos de sencillas ecuaciones lineales y son, con diferencia, los más utilizados. Para su resolución destacan, por su simplicidad, el *método de Yule-Walker*, que conduce a la resolución de un sencillo sistema de ecuaciones, y el *método de Burg*, computacionalmente más eficiente y que da lugar a modelos  $\text{AR}$  estables (cfr. [37]).

Los modelos  $\text{ARMA}(p, q)$ , aunque no tan sencillos, permiten mejorar la estima espectral utilizando menos parámetros en el proceso. La obtención de sus coeficientes se lleva a cabo en dos partes:

- a) Estimación de la parte autorregresiva ( $\text{AR}$ ) utilizando un método mínimo-cuadrático que conduce al cálculo de una pseudoinversa (*método Yule-Walker modificado*).
- b) Estimación de la parte de media móvil ( $\text{MA}$ ) aprovechando el conocimiento de la parte  $\text{AR}$  ya estimada, generando una secuencia  $v_k$

$$v_k = x_k + \sum_{i=1}^p \hat{a}_k x_{k-i} \quad (2.18)$$

Dado que  $x_k$  es un proceso ARMA( $p, q$ ) generado por  $H(z) = B(z)/A(z)$ , aproximadamente  $v_k$  es un proceso MA( $q$ ) generado por el proceso  $B(z)$ .

**Métodos Recursivos:** Los metodos anteriormente descritos trabajan en bloque, es decir, requieren la disponibilidad de todos los datos  $\{x_k\}_{k=0, \dots, N-1}$  para la estimación del espectro de potencias. Existen, no obstante, muchas situaciones en las que los datos son suministrados uno a uno. Aunque la aplicación de algoritmos de bloque es posible e incluso frecuente, en ciertos casos es posible aprovechar los cálculos hechos para los  $N$  datos anteriores cuando llega el dato  $N + 1$  con un coste computacional muy inferior al de la aplicación de métodos en bloque. Este tipo de métodos se basa en algoritmos de estimación mínimo-cuadráticos de carácter recursivo, como es el caso del algoritmo RLS (cfr. [39] para un estudio detallado). El algoritmo RLS resuelve el siguiente problema de regresión

$$x_k = \psi_0(k)\theta_0 + \dots + \psi_{n-1}(k)\theta_{n-1} + e_k = \psi^T(k) \cdot \theta \quad (2.19)$$

donde  $\psi(k)$  es un vector con  $n$  datos conocidos en el instante  $k$  y  $\theta$  es un vector con  $n$  parámetros que minimiza el error cuadrático cometido.

La estimación de modelos autorregresivos AR puede ser formulada como un problema de regresión lineal. El modelo AR( $p$ ) responde a la siguiente ecuación en diferencias

$$x_k = -(a_1x_{k-1} + \dots + a_px_{k-p}) \quad (2.20)$$

que tiene la misma estructura de (2.19) con

$$\psi(k) = [x_{k-1}, \dots, x_{k-p}]^T \quad (2.21)$$

$$\theta(k) = [a_1, \dots, a_n]^T \quad (2.22)$$

donde, evidentemente,  $\psi(k)$  es conocido en el instante  $k$ . El algoritmo RLS permite, además, considerar señales no estacionarias mediante un parámetro denominado *factor de olvido* que refleja el peso asignado a datos anteriores en el cálculo del modelo en el instante  $k$ . En [40] Ljung et al. hacen un completo estudio de la capacidad de adaptación de diversos algoritmos mínimo-cuadráticos para el seguimiento de sistemas variantes.

## 2.5 Análisis Cepstral

Una variante, heredada del campo de reconocimiento del habla, es el análisis basado en el *cepstrum*. El cepstrum  $c_x(n)$  se define como la transformada inversa de Fourier del logaritmo de la transformada directa de la señal, a saber,

$$X(\theta) = \sum_{k=-\infty}^{\infty} x_k e^{-j\theta k} \quad (2.23)$$

$$c_x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln X(\theta) e^{j\theta n} \quad (2.24)$$

El cepstrum permite el análisis de ciertos efectos no lineales. El cepstrum es una transformación *homomórfica*, es decir, una transformación que cumple un principio generalizado de superposición aplicable a sistemas no lineales (*sistemas homomórficos*) [38][37]. En el caso de la convolución, por ejemplo, si una señal  $\{x_k\}$  atraviesa un filtro  $H(z)$  dando lugar a  $\{y_k\}$ , se verifica

$$y_k = h_k * x_k \quad (2.25)$$

$$Y(z) = H(z)X(z) \quad (2.26)$$

$$\ln Y(z) = \ln H(z) + \ln X(z) \quad (2.27)$$

$$c_y(n) = c_h(n) + c_x(n) \quad (2.28)$$

es decir, una operación no lineal como “\*” entre dos secuencias se transforma en una operación lineal “+” entre las transformadas. La multiplicación y la convolución entre señales de diferentes procedencias es un fenómeno que se presenta a menudo en el análisis de vibraciones de maquinaria como consecuencia de acoplamientos no lineales entre dos o más elementos. La aplicación de una transformación homomórfica como el cepstrum facilita la separación o desacoplamiento entre los fenómenos que interactúan vía convolución o multiplicación al convertirse éstas en sumas en el dominio cepstral.

En [1], Tavner cita como aplicación típica del cepstrum el análisis de cajas de engranajes, cuyos espectros de vibración presentan multitud de familias de bandas laterales debido a la gran variedad de frecuencias moduladoras y moduladas que interactúan en este tipo de sistemas. La

modulación es una interacción multiplicativa que el cepstrum trata adecuadamente por su carácter de transformación homomórfica, desacoplando las señales que interactúan. Desde otro punto de vista, el cepstrum desacopla la periodicidad debida a las bandas laterales en el espectro de frecuencia –efectúa la transformación inversa de Fourier del logaritmo del espectro– presentando la información en una escala temporal<sup>3</sup> en la que las múltiples bandas laterales en el espectro originan picos claramente diferenciados en el cepstro.

La detección de fallos en rodamientos es también un problema en el que se dan interacciones entre señales. Concretamente, en rodamientos defectuosos, suelen generarse pulsos de *shock* a ciertas frecuencias típicas<sup>4</sup> que excitan los modos de vibración natural de los distintos elementos, lo que origina interacciones de tipo convolutivo. En [41], Barkov describe diversos métodos para la evaluación de la condición en rodamientos, entre los que destaca el cepstrum como un método efectivo para este tipo de problema.

## 2.6 Análisis de Estadísticos de Orden Superior (HOSA)

Las técnicas de Análisis Espectral vistas en la sección previa consideran la señal  $\{x_k\}$  como de carácter aleatorio y los métodos descritos se centran en el análisis de propiedades estadísticas de órdenes uno y dos de la señal (e.g. secuencia de autocorrelación) así como sus transformadas de Fourier (e.g. espectro de potencia). Este tipo de métodos tiene ciertas limitaciones. El espectro de potencia, por ejemplo, no aporta información acerca de la fase [42][43]. Asimismo, existen situaciones en las que dos o más armónicos interactúan entre sí, se acoplan, aportando energías a las frecuencias sumas y diferencias de las originales, fenómeno conocido como *acoplamiento cuadrático de fase* [42][31]. Otra limitación sería del análisis

<sup>3</sup>En la literatura se utiliza a veces el término *quefreny* (palabra *frequency* con algunas letras al revés para identificar esta metodología con el análisis de espectros). El propio término *cepstrum* deriva de permutar algunas letras en la palabra *spectrum*.

<sup>4</sup>Estas frecuencias dependen de qué superficie presente irregularidades en el cojinete. Son típicas frecuencias de fallo la frecuencia de giro de los elementos rodantes (BSF), la frecuencia de paso de las bolas por las cara exterior e interior del cojinete (BPFO, BPFI) y la frecuencia de giro del tren (FTF).

espectral clásico reside en el hecho de que no proporciona información acerca de la *gaussianidad* de las señales [42], a saber, del grado de ajuste a una normal, lo que puede aportar información relevante sobre el origen de ésta.

Una extensión natural del análisis espectral surge a partir del análisis de estadísticos de órdenes superiores. El análisis de estadísticos de orden superior se basa en unos estadísticos denominados *cumulantes*. Los cumulantes se definen, para variables aleatorias  $x_i$  de media nula, como [43]:

$$\text{cum}(x_1, x_2) = E[x_1 x_2] \quad (2.29)$$

$$\text{cum}(x_1, x_2, x_3) = E[x_1 x_2 x_3] \quad (2.30)$$

$$\begin{aligned} \text{cum}(x_1, x_2, x_3, x_4) &= E[x_1 x_2 x_3 x_4] - \\ &\quad - E[x_1 x_2] E[x_3 x_4] \\ &\quad - E[x_1 x_3] E[x_2 x_4] \\ &\quad - E[x_1 x_4] E[x_2 x_3] \end{aligned} \quad (2.31)$$

Para señales  $\{x_k\}$  estacionarias, de media nula, los cumulantes de órdenes 2, 3 y 4 se definen directamente a partir de las expresiones anteriores (cfr. [43], [31]):

$$C_{2,x}(\tau) = E[x_k x_{k+\tau}] \quad (2.32)$$

$$C_{3,x}(\tau_1, \tau_2) = E[x_k x_{k+\tau_1} x_{k+\tau_2}] \quad (2.33)$$

$$\begin{aligned} C_{4,x}(\tau_1, \tau_2, \tau_3) &= E[x_k x_{k+\tau_1} x_{k+\tau_2} x_{k+\tau_3}] \\ &\quad - C_{2,x}(\tau_1) C_{2,x}(\tau_2 - \tau_3) \\ &\quad - C_{2,x}(\tau_2) C_{2,x}(\tau_3 - \tau_1) \\ &\quad - C_{2,x}(\tau_3) C_{2,x}(\tau_1 - \tau_2) \end{aligned} \quad (2.34)$$

Nótese que la definición de  $C_{2,x}(\tau)$  coincide con la de la secuencia de autocorrelación, en la que se basa el análisis espectral clásico. De manera paralela al análisis espectral, se definen el *biespectro* y el *triespectro* –y, en general, de forma análoga, los *poliespectros*– a partir de los cumulantes  $C_{3,x}(\tau_1, \tau_2)$  y  $C_{4,x}(\tau_1, \tau_2, \tau_3)$  según:

$$B_x(f_1, f_2) = \int \int C_{3,x}(\tau_1, \tau_2) e^{-j2\pi(f_1\tau_1 + f_2\tau_2)} d\tau_1 d\tau_2 \quad (2.35)$$

$$T_x(f_1, f_2, f_3) = \int \int \int C_{4,x}(\tau_1, \tau_2, \tau_3) e^{-j2\pi(f_1\tau_1 + f_2\tau_2 + f_3\tau_3)} d\tau_1 d\tau_2 d\tau_3$$

Ambos admiten expresiones más eficientes para su cálculo [31]:

$$B_x(f_1, f_2) = E[X(f_1) \cdot X(f_2) \cdot X^*(f_1 + f_2)] \quad (2.36)$$

$$T_x(f_1, f_2, f_3) = E[X(f_1) \cdot X(f_2) \cdot X(f_3) \cdot X^*(f_1 + f_2 + f_3)] \quad (2.37)$$

Una importantísima propiedad de los cumulantes y que se aplica a los poliespectros es que los cumulantes se anulan a partir del orden dos, cuando el proceso  $\{x_k\}$  es gaussiano, es decir, son *ciegos* frente a procesos gaussianos. Así, los espectros de orden superior, proveen una medida de la *no-gaussianidad* de un proceso [44]. Al mismo tiempo, permiten elevar considerablemente la relación señal ruido cuando las señales a analizar están corrompidas por ruido gaussiano. Ambas propiedades son muy deseables en la extracción de características de procesos industriales.

Tradicionalmente, se han venido utilizando de manera implícita estadísticos de ordenes superiores para la extracción de características en señales de vibración. La *kurtosis* y el *sesgo* se derivan directamente  $C_{2,x}(0)$ ,  $C_{3,x}(0, 0)$  y  $C_{4,x}(0, 0, 0)$ , que se denominan *cumulantes de desplazamiento cero* (en inglés, *zero-lag cumulants*). Este tipo de características ignora la información temporal que porta la señal, aportando sólo información acerca de cómo está distribuida. En [41], Barkov menciona la kurtosis de las señales de vibración como una característica útil en la detección de defectos en los cojinetes. Análogamente, McCormick [31] utiliza los cumulantes *zero-lag* en las vibraciones horizontal y vertical de los cojinetes de un motor eléctrico para determinar las características de fallos de tipo mecánico tales como defectos de alineación, excentricidades, etc. Los cumulantes con desplazamientos y sus poliespectros, no obstante, aportan mucha más información sobre la estructura de la señal y los procesos que la generan, lo que hace que éstos constituyan buenas características [31]. Murray y Penman, por otra parte, demuestran en [42], también para un caso de vibraciones de un motor eléctrico, cómo componentes presentes en un fallo aislado se combinan con componentes en otro fallo produciendo características adicionales en el biespectro.

Finalmente, citar que al igual que en el caso de las correlaciones cruzadas, los cumulantes pueden ser aplicados entre distintas secuencias aplicando 2.29–2.31. También son utilizados en algunas aplicaciones los *policepstros* [43][44] en un intento de conjugar las propiedades del cepstrum con las de los estadísticos de orden superior.

## 2.7 Conclusiones

En este capítulo se ha pretendido realizar un esfuerzo de síntesis en la descripción de las técnicas de extracción de características más empleadas en el estado del arte de la monitorización de la condición de procesos y, más concretamente, de máquinas eléctricas.

Este tipo de técnicas, constituyen transformaciones matemáticas sencillas que permiten simplificar enormemente la geometría de las regiones asociadas a las distintas condiciones de la máquina. La gran mayoría de las técnicas de extracción de características utilizadas en máquinas eléctricas, emplean, de un modo u otro, transformaciones frecuenciales –el cepstrum y las técnicas HOSA también hacen uso de la transformada de Fourier–. La transformada de Fourier puede verse como una transformación de coordenadas en el espacio de las funciones, constituyendo, de hecho, una proyección sobre una base de funciones ortogonales (familia de funciones  $e^{j\omega t}$  para distintos valores de  $\omega$ ).

La enorme ventaja que plantea esta transformación puede resumirse, a nuestro juicio, en dos puntos:

- Dado que en la inmensa mayoría de las máquinas eléctricas la etiología de los fenómenos involucrados es de carácter senoidal, una proyección sobre una base de funciones senoidales<sup>5</sup> favorece enormemente el *principio de empaquetamiento de la información*, citado al principio del capítulo. De esta manera, las variables de interés, casi siempre senoides, son representadas por un único coeficiente complejo.
- Las variaciones periódicas que sufren los elementos de una ventana de datos temporales –como el vector  $y(t)$  del ejemplo al inicio del capítulo– se manifiestan en variaciones de la fase de los coeficientes complejos en el dominio de la frecuencia, lo que permite eliminar fácilmente esta variación tomando como características exclusivamente los módulos de éstos. Esto favorece la *invarianza de las características* cuando el proceso no varía su condición.

---

<sup>5</sup>Por brevedad, se emplea aquí el término “senoides” indistintamente para hacer referencia a senoides o a exponenciales complejas. Una senoide –teorema de Leblanc, geometría hiperbólica– no es sino una composición de dos exponenciales complejas que giran en sentidos opuestos.

En el contexto de lo expuesto en las secciones 1.4 y 2.1, las técnicas de extracción de características permiten transformar, de una manera relativamente sencilla, regiones con geometrías muy complejas, en regiones más simples y descritas con arreglo a una representación más adecuada, tanto desde el punto de vista de la economía en el número de variables necesarias para describir una condición, como por su invarianza ante una misma condición.



# Capítulo 3

## Métodos de reducción de la dimensionalidad

### 3.1 Introducción

El objetivo básico de este capítulo es la descripción de métodos de generación de nuevas características mediante transformaciones de las características existentes o de los datos de medida del proceso. La idea clave en este tipo de métodos consiste en transformar un conjunto dado de medidas o características del proceso o sistema a analizar, en otro conjunto de características, más reducido pero igualmente significativo en cuanto al análisis que se pretende realizar. Este tipo de métodos explotan, por lo general, las correlaciones existentes entre las variables del proceso. Si dos o más variables están correladas, entonces alguna o varias de ellas pueden ser explicadas en términos de las demás. Los métodos de reducción de la dimensionalidad persiguen el aprovechamiento de estas relaciones para expresar con unas pocas variables o características, muy significativas, lo que antes se expresaba con un número superior de variables. Este tipo de métodos [45] pueden dar lugar a características con un alto grado de *empaquetamiento de la información* en comparación con los datos originales, lo que facilita en gran medida las tareas de clasificación e inferencia posteriores.

## 3.2 Análisis de componentes principales (PCA)

El método de análisis de componentes principales (en inglés *Principal Component Analysis* PCA) es uno de los métodos más simples y utilizados de reducción de la dimensionalidad. Básicamente, las componentes principales [46] son una colección de variables  $y_i$ , combinación lineal de las variables observadas  $x_i$ , con la propiedad de tener varianza máxima y estar incorreladas. Estas propiedades de las componentes principales permiten explicar, mediante unas pocas componentes, la mayor parte de la información que contienen las variables observadas. A continuación se expone este método con más detalle.

Sea  $\mathbf{x}$  un vector de  $n$  medidas o características de media nula <sup>1</sup>. Sea  $\mathbf{R}_{xx}$  su matriz de correlación,

$$\mathbf{R}_{xx} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}(k)\mathbf{x}(k)^T \quad (3.1)$$

Sean  $\{\mathbf{u}_k\}_{k=1,\dots,N}$  los vectores propios de  $\mathbf{R}_{xx}$ , y sean  $\{\lambda_k\}_{k=1,\dots,N}$  los valores propios que, al tratarse de una matriz simétrica serán todos reales y positivos. Según el Algebra Lineal,  $\mathbf{R}_{xx}$  admite la siguiente descomposición:

$$\mathbf{R}_{xx} = \mathbf{U}\Lambda\mathbf{U}^T \quad (3.2)$$

donde  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_N]$  y  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$ . La ecuación (3.2) puede desarrollarse en una suma,

$$\mathbf{R}_{xx} = \sum_{k=1}^N \lambda_k \cdot \mathbf{u}_k \mathbf{u}_k^T \quad (3.3)$$

si los valores propios están ordenados  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$  entonces puede aproximarse  $\mathbf{R}_{xx}$  de la siguiente manera:

$$\mathbf{R}_{xx} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \dots + \lambda_p \mathbf{u}_p \mathbf{u}_p^T + \varepsilon \quad (3.4)$$

---

<sup>1</sup>La suposición de media nula no resta generalidad al argumento. En cualquier momento se puede considerar  $\mathbf{x} - E[\mathbf{x}]$

donde  $\varepsilon$  es un término residual. Los valores propios  $\lambda_i$  en la ecuación (3.4) describen las contribuciones de cada una de las direcciones principales  $\mathbf{u}_i$  en la formación de la matriz de correlaciones. La descomposición (3.4), revela que la matriz de correlaciones  $\mathbf{R}_{xx}$  puede ser descompuesta en términos de las contribuciones de la variabilidad  $\lambda_i$  de  $\mathbf{x}$  en cada dirección principal  $\mathbf{u}_i$ , y permite considerar las  $p$  direcciones que más contribuyen a su formación.

Geoméricamente, las  $p$  componentes principales pueden obtenerse hallando las coordenadas del vector  $\mathbf{x}$  sobre la base definida por los  $p$  vectores principales. Las coordenadas serán las proyecciones de  $\mathbf{x}$  sobre los vectores principales  $\mathbf{u}_i$ :

$$y_i = \text{Proy}_{\mathbf{u}_i} \mathbf{x} = \mathbf{u}_i^T \mathbf{x} \quad (3.5)$$

Si se define el vector  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$ , puede aplicarse en bloque (3.5) llegando a la siguiente expresión

$$\mathbf{y} = \mathbf{U}^T \mathbf{x} \quad (3.6)$$

Esta expresión se conoce también por el nombre de transformación de *Karhunen-Loève (KL)* y su uso está muy extendido en reconocimiento de patrones [45][47]. Por ser  $\mathbf{R}_{xx}$  simétrica,  $\mathbf{U}$  es ortogonal,  $\mathbf{U}^{-1} = \mathbf{U}^T$ , con lo cual

$$\mathbf{R}_{yy} = \mathbf{U}^T \mathbf{R}_{xx} \mathbf{U} = \mathbf{U}^T \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \mathbf{U} = \mathbf{\Lambda} \quad (3.7)$$

Por la ortogonalidad de  $\mathbf{U}$  puede también obtenerse  $x$  en función de  $y_i$  y  $\mathbf{u}_i$ ,

$$\mathbf{x} = \mathbf{U} \mathbf{y} = \sum_{k=1}^N y_k \mathbf{u}_k \quad (3.8)$$

Las expresiones (3.5), (3.6) y (3.7) permiten hacer las siguientes consideraciones:

- Dado que  $\Lambda$  es diagonal, las variables  $y_i$  son mutuamente incorreladas, con lo que la información que aportan está exenta de redundancias debidas a correlaciones, siendo sus respectivas varianzas  $\lambda_i$ .
- Los vectores  $\mathbf{u}_1, \dots, \mathbf{u}_p$  indican las  $p$  direcciones de máxima variabilidad y son los que más contribuyen a la formación de la matriz  $\mathbf{R}_{xx}$ .
- Mediante las  $p$  variables  $\{y_i\}_{i=1, \dots, p}$  es posible resumir las características más significativas de  $\mathbf{x}$
- El vector  $\mathbf{x}$  puede aproximarse de forma óptima utilizando la información de las  $p$  componentes principales según,

$$\hat{\mathbf{x}} = \sum_{k=1}^p (\mathbf{u}_k^T \mathbf{x}) \mathbf{u}_k = \sum_{k=1}^p y_k \mathbf{u}_k \quad (3.9)$$

pudiendo comprobarse fácilmente de (3.9) que el error cometido es:

$$E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2] = \sum_{k=p+1}^N \mathbf{u}_k^T \lambda_k \mathbf{u}_k = \sum_{k=p+1}^N \lambda_k \quad (3.10)$$

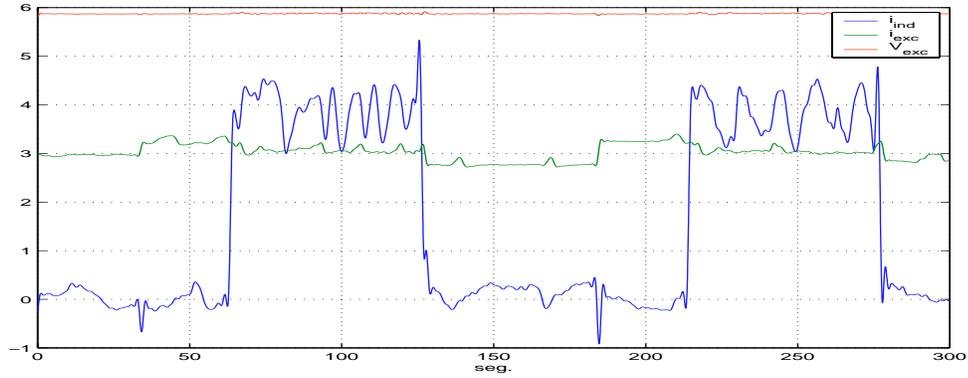
Este error es el mínimo posible siendo  $\lambda_{p+1}, \dots, \lambda_N$  los valores propios más pequeños. Cuanto más próximos sean éstos valores a cero, más efectiva será la reducción de la dimensionalidad en el sentido de que preservará mayor información sobre los datos de entrada [47].

**Ejemplo de aplicación de componentes principales.** En la figura 3.1 se ilustra la aplicación del método PCA a datos reales de proceso en la factoría de Aeralia. Los datos proceden de un motor CC de 7000 Kw, instalado en un tren de laminación, y corresponden a la corriente de excitación  $i_{exc}$ , corriente de inducido  $i_{ind}$  y a la tensión de inducido  $V_{ind}$  durante dos operaciones de laminado. En la representación 3D pueden apreciarse, claramente diferenciadas, tres nubes de puntos que, en

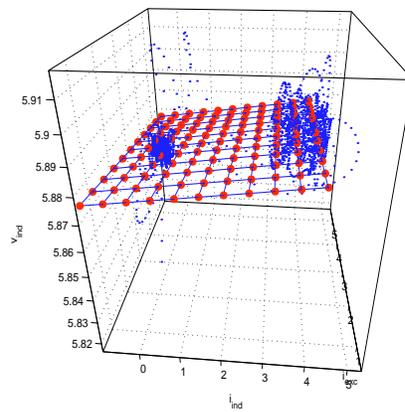
principio, corresponden a tres situaciones estables por las que atraviesa el proceso. Una observación minuciosa de la proyección 2D en el plano principal muestra que, en realidad existen más estados. Los dos estados más evidentes corresponden al motor girando en vacío y al motor laminando (incremento de la  $i_{exc}$  debido a la demanda de par), mientras que el resto de los estados –pequeños subgrupos dentro de la agrupación de la izquierda– corresponden a otras situaciones en las que los lazos de regulación implementados en el accionador del motor debilitan o fortalecen el campo para regular la velocidad.

**Inconvenientes del método PCA.** El método PCA, sin embargo, no siempre es idóneo para el reconocimiento de patrones. Ciertas geometrías de los datos de entrada, aunque sean linealmente separables, pueden dar lugar a solapamientos entre clases cuando se utilizan las componentes principales ocultando, por tanto, importante información para la separación de clases, tal y como se aprecia en la figura 3.2. Este problema se deriva del hecho de que la transformación de Karhunen-Loève, aunque es óptima respecto a la minimización del error de aproximación, no es óptima con respecto a la separabilidad de clases [45].

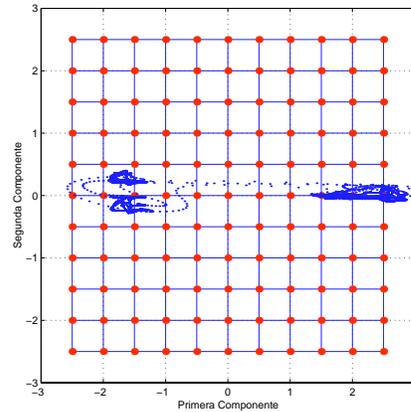
En el caso, más general, de clases no linealmente separables es evidente que las componentes principales no aportan ninguna solución dado que tan sólo representan una transformación lineal ortogonal (giros y cambios de escala) de los datos.



(a) Corriente de excitación  $i_{exc}$ , corriente de inducido  $i_{ind}$  y tensión de inducido  $V_{ind}$  durante dos operaciones de laminado en un motor CC de 7000 Kw.

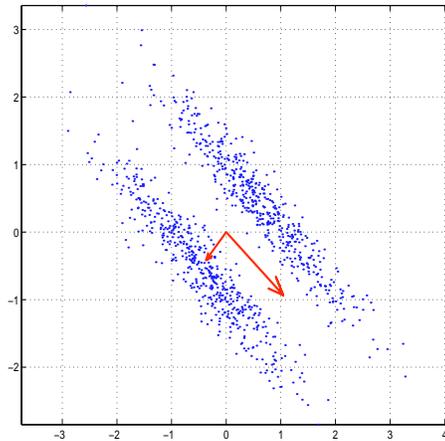


(b) Datos  $i_{exc}, i_{ind}, V_{ind}$  representados como nubes de puntos en  $\mathbb{R}^3$  junto con el subespacio engendrado por las direcciones principales  $\mathbf{u}_1, \mathbf{u}_2$

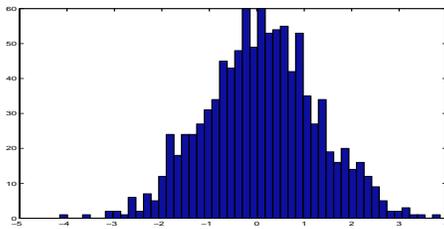


(c) Componentes Principales  $y_1, y_2$

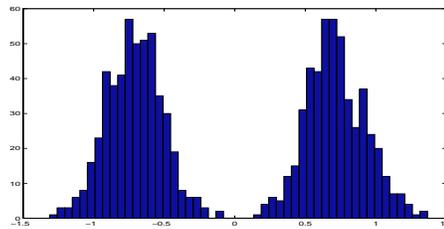
Figura 3.1: Reducción de la dimensionalidad mediante análisis de componentes principales. Los datos observados  $i_{exc}, i_{ind}$  y  $V_{exc}$  varían principalmente en dos direcciones del espacio. Una tercera componente, la de menor relevancia, reflejaría el “grosor” de la nube de puntos.



(a) Datos en el espacio de entrada



(b) Histograma de la primera componente principal.



(c) Histograma de la segunda componente principal.

Figura 3.2: Ejemplo en el que el PCA no realiza una reducción de la dimensionalidad óptima para el reconocimiento de patrones. La primera componente principal no aporta información para discriminar las dos clases mientras que la segunda componente permite la clasificación.

### 3.3 Análisis de Componentes Principales de Núcleo (Kernel PCA)

El análisis de componentes principales de núcleo es una técnica muy reciente debida fundamentalmente a Schölkopf [48][49]. La idea clave de esta técnica consiste en realizar un análisis PCA lineal en un *espacio de características* de dimensión generalmente muy superior a la del espacio de entrada (incluso de dimensión infinita) determinadas *no linealmente* a partir del espacio de entrada. Este método permite considerar, por ejemplo, características basadas en correlaciones de órdenes superiores entre las variables de entrada. Debido a la naturaleza no lineal de la relación entre el espacio de entrada y el espacio de características el método kernel PCA es *no lineal*.

Sea una función  $\psi(\mathbf{x})$  tal que  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , donde  $n$  es la dimensión del espacio de entrada y  $m$  es la dimensión del espacio de características. Puede definirse una matriz de correlación en el espacio de características de manera similar a la expresión (3.1) de la sección 3.2:

$$\mathbf{R}_{\psi\psi} = \frac{1}{N} \sum_{k=1}^N \psi(\mathbf{x}(k)) \cdot \psi(\mathbf{x}(k))^T \quad (3.11)$$

La descomposición en valores y vectores propios de  $\mathbf{R}_{\psi\psi}$ , de forma idéntica al método PCA da lugar a valores propios  $\{\lambda_i\}_{i=1,\dots,m}$  y vectores propios  $\{\mathbf{u}_i\}_{i=1,\dots,m}$  en el espacio de características  $C$ . Estos son las soluciones de,

$$\mathbf{R}_{\psi\psi} \mathbf{u} = \lambda \mathbf{u} \quad (3.12)$$

Los vectores  $\{\mathbf{u}_i\}_{i=1,\dots,m}$  engendran el subespacio generado por todos los  $\psi(\mathbf{x}(k))$ . Existe, por tanto, un conjunto de valores  $\alpha_k$  tales que,

$$\mathbf{u} = \sum_{k=1}^N \alpha_k \psi(\mathbf{x}(k)) \quad (3.13)$$

Combinando las ecuaciones (3.11), (3.12) y (3.13) se obtiene

$$\sum_{i=1}^N \sum_{j=1}^N \alpha_j \psi(\mathbf{x}(i)) K(\mathbf{x}(i), \mathbf{x}(j)) = N \lambda \sum_{j=1}^N \alpha_j \psi(\mathbf{x}(j)) \quad (3.14)$$

donde  $K(\mathbf{x}, \mathbf{y})$  es un núcleo de producto interno definido en términos de los vectores de características mediante

$$K(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{x})^T \psi(\mathbf{y}) = \psi(\mathbf{x}) \cdot \psi(\mathbf{y}) \quad (3.15)$$

Premultiplicando la ecuación (3.14) por  $\psi(\mathbf{x}(k))^T$  se llega a

$$\sum_{i=1}^N \sum_{j=1}^N \alpha_j K(\mathbf{x}(k), \mathbf{x}(i)) K(\mathbf{x}(i), \mathbf{x}(j)) = N\lambda \sum_{j=1}^N \alpha_j K(\mathbf{x}(k), \mathbf{x}(j)) \quad (3.16)$$

que, definiendo la *matriz de núcleo*,  $\mathbf{K} = [K(\mathbf{x}(i), \mathbf{x}(j))]_{i,j=1,\dots,N}$ , y el vector  $\alpha = [\alpha_1, \dots, \alpha_N]^T$  puede expresarse matricialmente de forma más compacta,

$$\mathbf{K}^2 \alpha = N\lambda \mathbf{K} \alpha \quad (3.17)$$

La expresión anterior admite la siguiente simplificación,

$$\mathbf{K} \alpha = N\lambda \alpha \quad (3.18)$$

Asumiendo  $\lambda_1 \geq \lambda_2, \dots, \lambda_N$  los vectores propios  $\mathbf{u}$  pueden obtenerse fácilmente a partir de los  $\alpha$  imponiendo

$$\mathbf{u}_k^T \mathbf{u}_k = 1, \quad k = 1, \dots, p \quad (3.19)$$

siendo  $\lambda_p$  el menor valor propio no nulo. Utilizando (3.12) y (3.13) puede comprobarse que la condición de normalización anterior equivale a

$$\alpha_k^T \alpha_k = \frac{1}{\lambda_k}, \quad k = 1, \dots, p \quad (3.20)$$

Las proyecciones  $\psi(\mathbf{x})$  del vector de características de un punto  $\mathbf{x}$  del espacio de entrada sobre los vectores propios dan lugar a las componentes principales:

$$\begin{aligned} y_k(x) &= \mathbf{u}_k^T \cdot \psi(\mathbf{x}) \\ &= \sum_{j=1}^N \alpha_j^k \psi(\mathbf{x}(j))^T \psi(\mathbf{x}) \\ &= \sum_{j=1}^N \alpha_j^k K(\mathbf{x}(j), \mathbf{x}), \quad k = 1, \dots, p \end{aligned} \quad (3.21)$$

donde, como puede comprobarse, se han obtenido éstas en términos del núcleo  $K(\cdot, \cdot)$ .

Las ecuaciones (3.18), (3.20) y (3.21) ofrecen un potente método para la extracción de las características más significativas de un espacio de entrada dado basándose en dos principios: transformación no lineal a un espacio de características y análisis PCA lineal en el espacio de características.

Nótese [48], [50], que  $\psi(\cdot)$  puede ser cualquier mapa no lineal arbitrario sobre un espacio de características posiblemente de alta dimensionalidad. Concretamente, por ejemplo,  $\psi(\cdot)$  puede ser constituido por los valores de todos los monomios de orden  $d$  en las componentes  $x_i$  de  $\mathbf{x}$ . Los productos escalares del tipo  $\psi(\mathbf{x}) \cdot \psi(\mathbf{y})$  pueden tener en este caso un coste prohibitivo al ser la dimensión de  $\psi(\mathbf{x})$  igual a  $\frac{(n+d+1)!}{d!(n-1)!$ , que crece a un ritmo de  $n^d$ . Scholkopf, [49], cita como ejemplo datos de entrada procedentes de imágenes de  $16 \times 16$  pixels ( $n = 256$ ) para un valor de  $d = 5$  lo que da lugar a un espacio de características de  $10^{10}$  dimensiones.

La solución a este problema estriba en el hecho de que para el análisis PCA en el espacio de características tan sólo es preciso conocer los productos escalares entre los vectores de características, *sin necesidad de conocer éstos explícitamente*. El análisis anterior expresa el análisis de componentes principales en el espacio de características (de dimensionalidad posiblemente elevada o incluso infinita) mediante la evaluación de núcleos

$$K(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{x}) \cdot \psi(\mathbf{y}) \quad (3.22)$$

en los que no es necesaria explícitamente la evaluación de  $\psi(\mathbf{x})$ . En [49] se demuestra de manera sencilla que el núcleo

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d \quad (3.23)$$

equivale al producto  $\psi(\mathbf{x}) \cdot \psi(\mathbf{y})$  en el ejemplo anterior, donde  $\psi(\cdot)$  transforma  $\mathbf{x}$  sobre el espacio de características de todos los monomios de grado  $d$  formados con los elementos de  $\mathbf{x}$ , reduciendo drásticamente el coste computacional debido a la explosión combinatoria.

Otros núcleos frecuentemente utilizados [49] son los de funciones de base radial

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad (3.24)$$

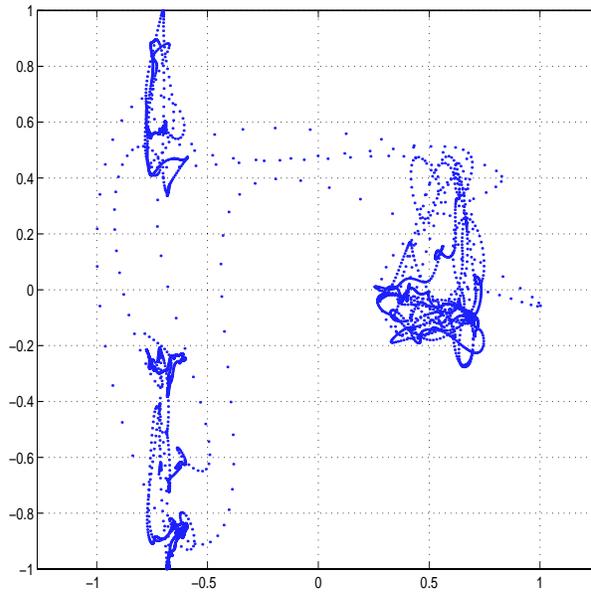


Figura 3.3: Nube de puntos en  $\mathbb{R}^2$  formada por las medidas  $(i_{exc}, i_{ind})$ .

y los núcleos sigmoidales

$$K(\mathbf{x}, \mathbf{y}) = \tanh(\kappa(\mathbf{x} \cdot \mathbf{y}) + \Theta) \quad (3.25)$$

La cuestión relativa al núcleo que debe aplicarse para cada problema concreto es aún un problema abierto [49].

En el siguiente ejemplo se muestra el funcionamiento del método descrito. En la figura 3.3 se las mismas medidas del motor de 7000 Kw utilizadas en la sección anterior. Por conveniencia en la representación se han utilizado tan sólo las dos variables más significativas,  $i_{exc}$ ,  $i_{ind}$  y se han normalizado (mediante desplazamiento y variación de escala) de manera que varíen entre +1 y -1.

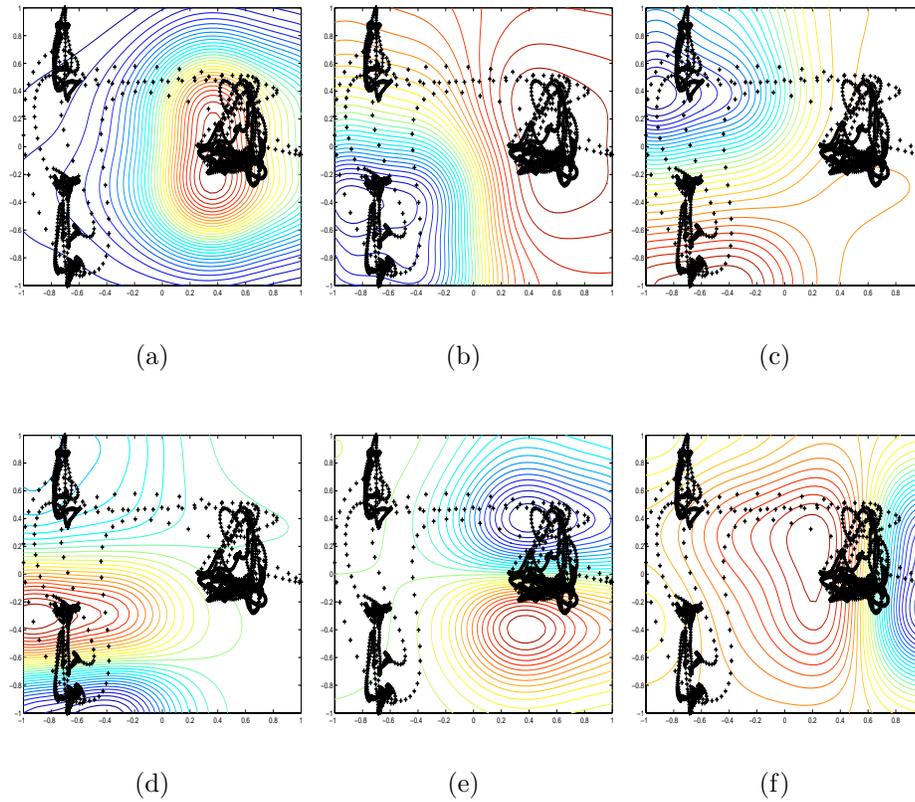


Figura 3.4: Componentes Principales obtenidas mediante el método *Kernel PCA*.

En la figura 3.4 se describe el valor de las 6 componentes principales (“altura” representada por superficies de nivel), correspondientes a cada punto del plano en el que se encuentran los datos a analizar. Puede apreciarse cómo las primeras componentes (subfiguras a,b,c) “captan” las estructuras o agrupaciones más importantes, mientras que las componentes de menor peso (d,e,f) revelan detalles más sutiles. En la figura 3.5 se muestra la evolución de las 4 características principales para los puntos de la trayectoria que describe el proceso, donde puede apreciarse cómo las primeras características segmentan las situaciones más sobresalientes, mientras que las características menos predominantes revelan aspectos más concretos. En la figura 3.6 pueden apreciarse representadas una frente a otra las características más importantes de los puntos de la

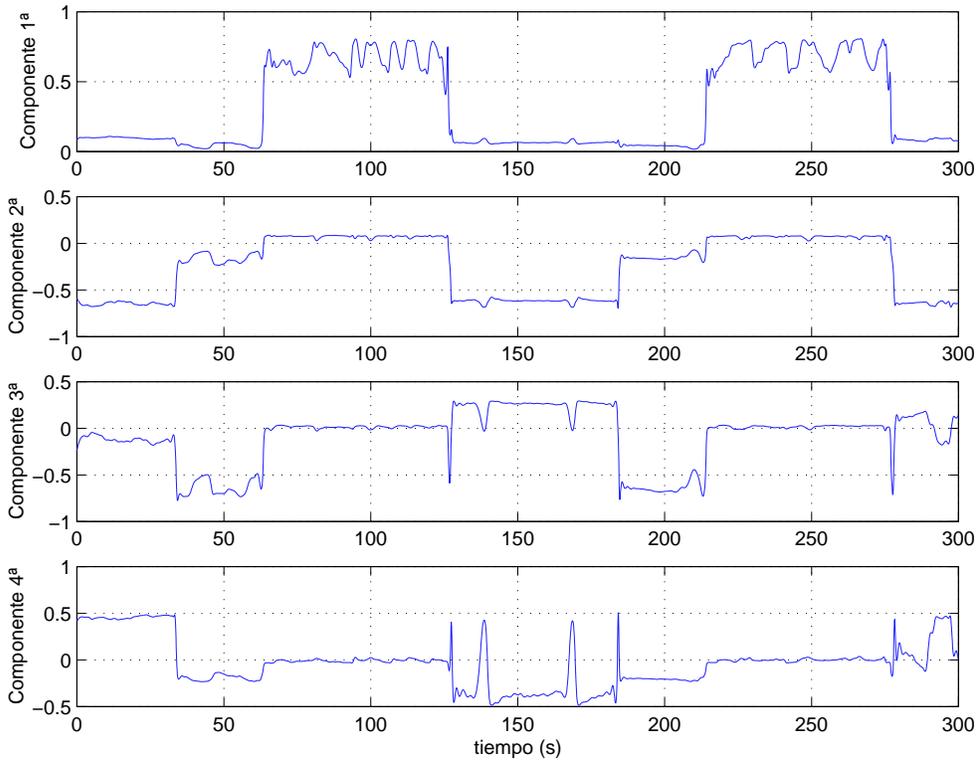


Figura 3.5: Evolución temporal de las 4 características principales para los datos de proceso.

trayectoria.

**Ventajas e inconvenientes del Kernel PCA.** El principal problema de este método es que el tamaño de la matriz  $\mathbf{K}$  crece con el número de observaciones, lo que puede hacerlo inviable para un número elevado de muestras. Además, su interpretación no es evidente, siendo la elección del núcleo un problema aún poco claro. No obstante, la idea en la que se basa este método está recibiendo una considerable atención en los últimos años. La transformación a un espacio de mayor dimensionalidad permite utilizar clasificadores con una menor dimensión VC, utilizando un conjunto reducido de los datos de entrada denominados vectores de soporte, lo que permite controlar la generalización, lo que ha dado pie a un nuevo paradigma dentro de las redes neuronales denominado aprendizaje

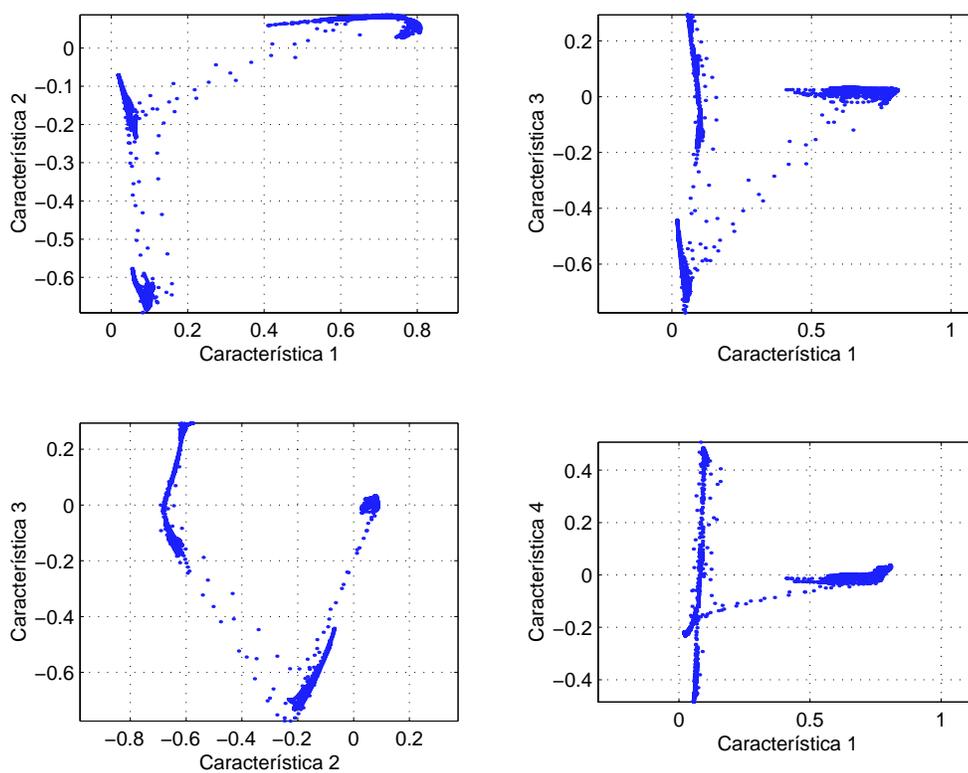


Figura 3.6: Características principales de los puntos de la trayectoria confrontadas entre sí.

basado en vectores de soporte [48].

### 3.4 Métodos de Escalado Multidimensional (MDS)

Los métodos MDS (en inglés, *Multidimensional Scaling*) constituyen un grupo de métodos cuyo objetivo es generar una colección de puntos en un espacio imagen cuyas distancias mutuas se asemejen lo más posible a las del espacio origen.

El punto de partida de los algoritmos MDS es una matriz formada por las distancias mutuas entre los vectores  $\{\mathbf{x}_k\}_{k=1,\dots,N}$  del espacio de entrada  $\mathbb{R}^n$ ,

$$\mathbf{X}_{ij} = d_x(\mathbf{x}_i, \mathbf{x}_j) \quad (3.26)$$

La distancia  $d_x(\cdot, \cdot)$ , en principio, no tiene por qué restringirse a la euclídea, por lo que debe entenderse ésta en un sentido genérico. El objetivo de este tipo de métodos consiste en generar una colección de puntos  $\{\mathbf{y}_k\}_{k=1,\dots,N}$  de un espacio de dimensión inferior,  $\mathbb{R}^p$ , de manera que su matriz de distancias mutuas

$$\mathbf{Y}_{ij} = d_y(\mathbf{y}_i, \mathbf{y}_j) \quad (3.27)$$

se asemeje lo más posible a  $\mathbf{X}_{ij}$ . La matriz de distancias mutuas de un conjunto de puntos define su topología, es decir, la forma en que éstos están dispuestos, su geometría. La posibilidad de determinar una colección de puntos en un espacio 2D (ó 3D) cuya matriz de distancias mutuas se asemeje a la de los datos de entrada puede aportar gran cantidad de información sobre éstos.

Un aspecto particularmente evidente de su utilidad es la detección, a la vista de los puntos proyectados, de *clusters* o agrupaciones de datos en el espacio de entrada, así como el grado de similitud entre unas y otras agrupaciones. Además, la conservación de distancias mutuas permite revelar disposiciones geométricas más complejas en el espacio de entrada (e.g. anillos, esferas, tubos, etc.) siempre que éstas sean topológicamente realizables en el espacio de visualización.

La determinación de las proyecciones  $\mathbf{y}_k$  en este tipo de algoritmos se lleva a cabo a través de la minimización de funciones de coste que

penalizan disparidades entre las matrices de distancias mutuas. Es, precisamente, en dichas funciones de coste, donde residen las diferencias entre los distintos tipos de algoritmos MDS.

La forma más básica de este tipo de algoritmos es el *MDS métrico*, cuya función de coste es

$$E = \sum_{i \neq j} [d_x(\mathbf{x}_i, \mathbf{x}_j) - d_y(\mathbf{y}_i, \mathbf{y}_j)]^2 \quad (3.28)$$

Este método da el mismo peso a todas las disparidades entre distancias mutuas que puedan darse entre los datos de entrada y los de salida. Sin embargo, no siempre es posible reducir a cero el coste dado por (3.28) dado que en un espacio de entrada existen topologías que no pueden tener lugar en espacios de inferior dimensionalidad. A menudo, suele ser preferible un mayor grado de cumplimiento de la igualdad de distancias mutuas para distancias pequeñas, esto es, preservar la topología *localmente*. Esta es precisamente la base de otros algoritmos tales como los *mapas de Sammon* (cfr. [22][51][27]) y el *Análisis de Componentes Curvilíneas (CCA)* [52].

El mapa de Sammon, muy similar al MDS métrico, minimiza la siguiente expresión,

$$E = \sum_{i \neq j} \frac{[d_x(\mathbf{x}_i, \mathbf{x}_j) - d_y(\mathbf{y}_i, \mathbf{y}_j)]^2}{d_x(\mathbf{x}_i, \mathbf{x}_j)} \quad (3.29)$$

donde, como puede comprobarse, la única diferencia estriba en el hecho de que la exigencia de conservación de distancias tiene más peso cuando éstas son cortas. El Análisis de Componentes Curvilíneas (CCA), propuesto por Demartines et al. [52] generaliza a los anteriores, permitiendo elegir la función de ponderación. Su función de coste es,

$$E = \sum_{i \neq j} [d_x(\mathbf{x}_i, \mathbf{x}_j) - d_y(\mathbf{y}_i, \mathbf{y}_j)]^2 \cdot F(d_y(\mathbf{y}_i, \mathbf{y}_j), \lambda_y) \quad (3.30)$$

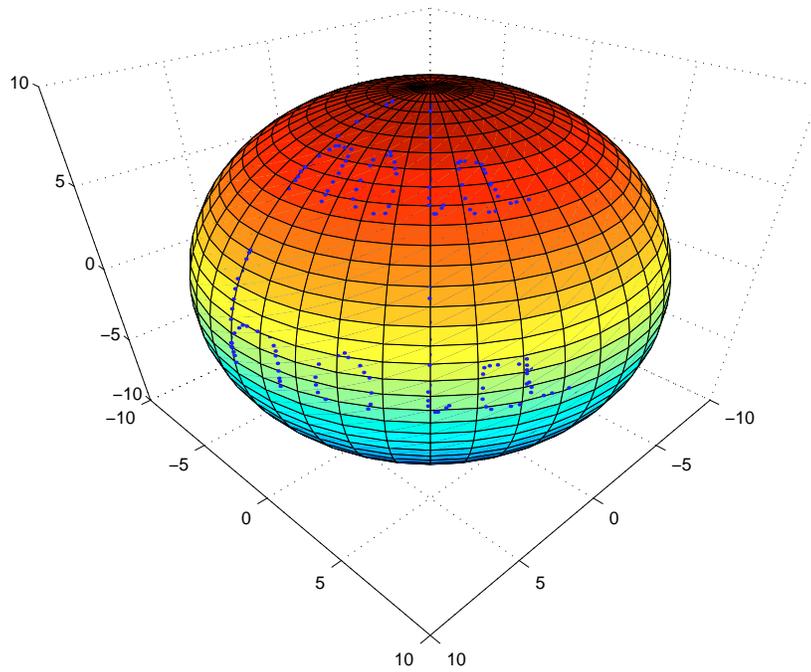
donde  $F(d_y(\mathbf{y}_i, \mathbf{y}_j), \lambda_y)$  es generalmente una función monótona decreciente para favorecer la conservación de la topología localmente y  $\lambda_y$  es

un parámetro ajustable relacionado con el radio de acción de la función. El uso de esta función de coste [52], permite el desdoblamiento de estructuras de datos altamente no lineales e incluso cerradas sobre sí mismas. El algoritmo de minimización ofrece rendimientos similares, según [52], varias decenas de veces más rápido que los algoritmos tipo MDS métrico o Sammon.

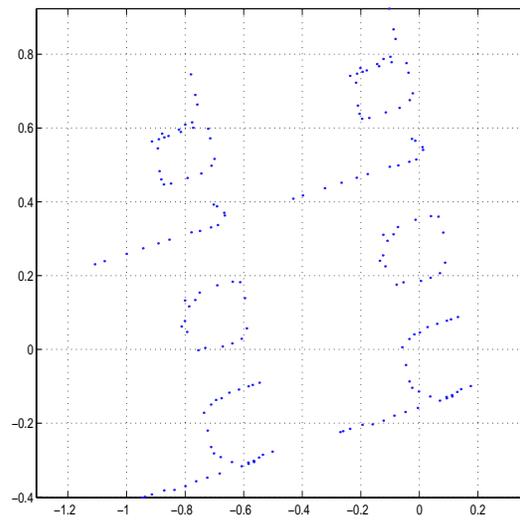
**Ejemplo del algoritmo CCA.** En el siguiente ejemplo se muestra la eficacia del Análisis de Componentes Curvilíneas en el desdoblamiento de datos altamente curvados. En la figura 3.7(a), se muestran los datos del espacio de entrada. Se trata de puntos de  $\mathbb{R}^3$  que componen la palabra “hola” escrita sobre una superficie esférica. La aplicación el algoritmo CCA, con un tiempo de 30 segundos en un Pentium 200 dio lugar al resultado de la figura 3.7(b). El resultado coincide de forma muy aproximada con el original salvo rotaciones o traslaciones, debido a la invarianza de las matrices de distancia ante este tipo de transformaciones.

**Aplicación del algoritmo CCA sobre datos de proceso** El algoritmo CCA fue aplicado a los mismos datos de proceso utilizados en secciones anteriores, formados por  $i_{exc}, i_{ind}, V_{ind}$  de un motor de 7000 Kw del tren reversible de Aceralia. Los resultados se muestran en la figura 3.8. Como puede apreciarse, el algoritmo CCA reproduce en 2D una estructura de puntos similar en su topología (distancias mutuas) a la existente en 3D, reflejando las agrupaciones existentes en el espacio de entrada. El resultado es similar (salvo giro y traslación) al del método PCA. Ello es debido a que una de las componentes,  $V_{ind}$  apenas varía, con lo que los datos están distribuidos en el espacio origen casi en un plano, sin curvatura, lo que en cierta medida refleja que el análisis CCA generaliza al PCA.

**Inconvenientes de los métodos MDS** Uno de los principales problemas de los métodos MDS es el de la complejidad de los algoritmos de minimización. Este tipo de algoritmos requieren, en general, la evaluación de las distancias mutuas, con lo que su complejidad es, al menos, de tipo  $\mathcal{O}(N^2)$ . Para volúmenes de datos grandes, el coste computacional se vuelve prohibitivo y más aún si el algoritmo de proyección debe aplicarse sobre un flujo de datos en tiempo real. El algoritmo de Demartines et al.

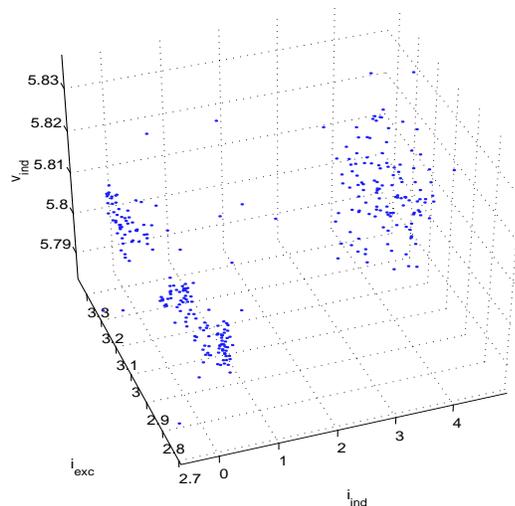


(a) Datos en el espacio de entrada. Los datos están altamente curvados (situados sobre una superficie esférica).

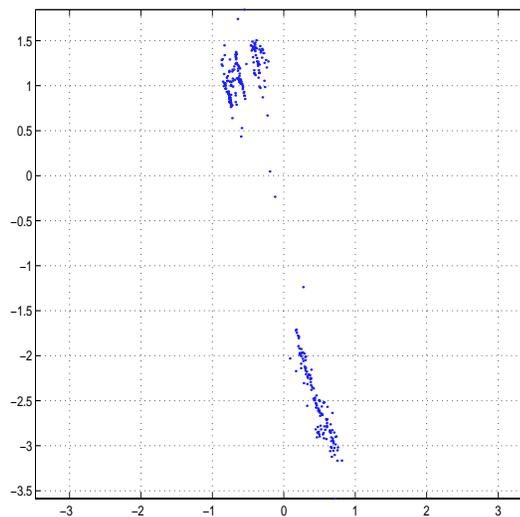


(b) Resultado del algoritmo CCA.

Figura 3.7: Ejemplo del método CCA.



(a) Datos en el espacio de entrada.



(b) Resultado del algoritmo CCA. El algoritmo CCA reproduce en 2D una estructura de puntos similar en su topología (distancias mutuas) a la existente en 3D, reflejando las agrupaciones existentes en el espacio de entrada.

Figura 3.8: Ejemplo del método CCA.

[52], reduce este tiempo a  $\mathcal{O}(N)$  para un ciclo de adaptación durante el mapeo, sin embargo no aclara cuántos pasos de adaptación son necesarios en cada punto proyectado.

Los algoritmos MDS, además, no aportan una expresión funcional explícita del mapa, tal y como lo hacen los métodos PCA y Kernel PCA vistos en secciones anteriores, sino que los puntos proyectados se obtienen mediante un proceso iterativo.

### 3.5 Mapas Topográficos Generativos (GTM)

Este método, debido a Svensén [27], trata de explicar los datos observados en el espacio de entrada en términos de datos en un *espacio latente* de menor dimensionalidad, mediante un modelo probabilístico *generativo* (es decir, un modelo que “genera” los datos de entrada, cuyo comportamiento se pretende explicar, a partir de los datos del espacio latente).

El GTM define una función paramétrica entre el espacio latente de dimensión  $p$  y el espacio de entrada, de dimensión  $n$ :

$$\mathbf{y}(\mathbf{x}, \mathbf{W}) : \mathbb{R}^p \longrightarrow \mathbb{R}^n \quad (3.31)$$

La función  $\mathbf{y}(\mathbf{x}, \mathbf{W})$ , puede ser cualquier función continua y diferenciable, aunque una elección adecuada de ésta facilita una estimación de los parámetros  $\mathbf{W}$ . Svensén asume que los datos del espacio de entrada  $\mathbf{t}$  son explicados en términos de los puntos del espacio latente mediante

$$p(\mathbf{t}|\mathbf{x}, \mathbf{W}, \beta) = \left(\frac{\beta}{2\pi}\right)^{-n/2} \exp\left[-\frac{\beta}{2} \cdot \|\mathbf{t} - \mathbf{y}(\mathbf{x}, \mathbf{W})\|^2\right] \quad (3.32)$$

según esto, puede expresarse la función de densidad de los datos de entrada según

$$p(\mathbf{t}|\mathbf{W}, \beta) = \int_{\mathbb{R}^p} p(\mathbf{t}|\mathbf{x}, \mathbf{W}, \beta)p(\mathbf{x}) \cdot d\mathbf{x} \quad (3.33)$$

La expresión anterior hace patente el carácter generativo del modelo, el cual *genera* o *explica* la función de densidad del espacio de entrada a partir de las variables latentes según  $\mathbf{y}(\mathbf{x}, \mathbf{W})$ .

Los parámetros  $\mathbf{W}$  y  $\beta$  se obtienen fácilmente por el método de máxima verosimilitud cuando dicha función se elige adecuadamente<sup>2</sup>, lo que da lugar a un algoritmo iterativo (cfr. [27], pp. 36 y ss.) relativamente sencillo de implementar.

<sup>2</sup>Svensén toma una combinación lineal de funciones de tres tipos: funciones gaussianas de base radial  $\phi_i(\mathbf{x}) = \exp(-\|\mathbf{x} - \mu_i\|^2/\sigma)$ , funciones identidad  $\phi_i(\mathbf{x}) = \mathbf{x}$  y constantes  $\phi(\mathbf{x}) = \mathbf{1}$ , todas ellas con derivadas muy simples, lo que facilita la aplicación del método de máxima verosimilitud. En la tesis doctoral de Svensén [27] se muestra una derivación completa del algoritmo.

**Visualización de los datos.** Las probabilidades condicionadas de cada uno de los puntos del espacio latente permiten visualizar los niveles de activación de cada punto del espacio latente por los puntos del espacio de entrada. La probabilidad condicionada de la variable latente  $\mathbf{x}$  dado un punto del espacio de entrada  $\mathbf{t}$  puede obtenerse a través del teorema de Bayes:

$$p(\mathbf{x}|\mathbf{t}) = \frac{p(\mathbf{t}|\mathbf{x}, \mathbf{W}, \beta)p(\mathbf{x})}{\int p(\mathbf{t}|\mathbf{x}, \mathbf{W}, \beta)p(\mathbf{x})d\mathbf{x}} \quad (3.34)$$

No obstante, en el método se trabaja con puntos  $\mathbf{x}_k$  uniformemente distribuidos (en forma de retícula rectangular) en el espacio latente, con lo que la expresión anterior queda

$$p(\mathbf{x}_k|\mathbf{t}) = \frac{p(\mathbf{t}|\mathbf{x}_k, \mathbf{W}, \beta)p(\mathbf{x}_k)}{\sum_{k'} p(\mathbf{t}|\mathbf{x}_{k'}, \mathbf{W}, \beta)p(\mathbf{x}_{k'})} \quad (3.35)$$

Habitualmente, suele escogerse una dimensión  $p = 2$  ó  $p = 3$  para el espacio latente con el objeto de poder visualizar los datos. En el caso 2D es posible asignar un nivel de pseudocolor correspondiente a la probabilidad condicionada  $p(\mathbf{x}_k|\mathbf{t})$  a cada punto de la retícula  $\mathbf{x}_k$  conformándose así una imagen con los niveles de probabilidad de cada punto del espacio latente, lo que permite revelar detalles sobre la estructura de los datos, agrupaciones, etc.

**Determinación de las Componentes Principales.** Las componentes principales de un punto  $\mathbf{t}_n$  del espacio de entrada, pueden obtenerse fácilmente utilizando la media de la distribución a posteriori:

$$\bar{\mathbf{x}}_n = \sum_k \mathbf{x}_k p(\mathbf{x}_k|\mathbf{t}_n) \quad (3.36)$$

o bien la moda

$$\mathbf{x}_n^{\text{moda}} = \arg \max_{\mathbf{x}_k} \{p(\mathbf{x}_k|\mathbf{t}_n)\} \quad (3.37)$$

**Ventajas e inconvenientes.** El GTM surge como una alternativa al SOM (que se verá más adelante, en la sección 3.8) derivada a partir de principios estadísticos lo que le aporta un valor añadido. Este método permite generar un mapa continuo que relaciona los espacios de entrada y visualización con propiedades similares a las del SOM. El método de convergencia, sin embargo, no es tan simple y la clase de mapas funcionales que puede emplear está limitada por la necesidad de determinar el gradiente.

El SOM es más flexible. Admite de forma sencilla generalizaciones en cuanto a la geometría del mapa aproximante (anillos, cilindros, toros, esferas, cubos, hipercubos, etc.) mediante sencillas modificaciones en topología y/o la distancia definidas en la retícula del espacio imagen. Pese a su carácter discreto, el SOM admite generalizaciones al caso continuo como la que se propone en esta Tesis, basada en una forma especial de regularización de las redes de base radial, conocida como GRNN, (*Generalized Regression Neural Network*), que se fundamenta en la regresión de núcleo.

## 3.6 Redes Neuronales FeedForward Autoasociativas

Una técnica alternativa a las anteriores (cfr. [53]) para la determinación de las componentes principales en un conjunto de datos multidimensional es la aplicación de redes feedforward autoasociativas.

La clave de este método consiste en entrenar una red tipo feedforward de varias capas con una de ellas constituyendo un “cuello de botella” (es decir, con un número de neuronas  $p$  inferior a la dimensión  $n$  del espacio de entrada) para reproducir la función identidad. En el proceso de entrenamiento, la red ajustará los pesos para reproducir en la salida los mismos vectores que en la entrada con un error cuadrático mínimo. Dado que la información debe fluir a través de la capa “cuello de botella”, puede considerarse que dicha información está presente en las neuronas de dicha capa.

Este método, sin embargo, carece aún de criterios objetivos que permitan determinar la estructura óptima de la red (número de capas ocultas, nodos en cada capa, etc.) exigiendo a menudo procedimientos de carácter heurístico a veces poco rigurosos (con frecuencia se usa el método de *prueba y error* para determinar su estructura), así como técnicas para podado de la red o el uso de vectores de validación y de test para evitar el sobreentrenamiento.

## 3.7 Curvas y Superficies Principales

Las curvas principales (Hastie et al. [54]) constituyen otra generalización del método PCA. Las curvas principales vienen definidas por la propiedad de que cada punto de la curva es el promedio de todos los puntos de la distribución que se proyectan sobre ella. Matemáticamente:

$$\mathbf{f}(x) = E[\mathbf{t} | \text{Proy}_{\mathbf{f}}(\mathbf{t}) = x] \quad (3.38)$$

Intuitivamente, estas curvas se ajustan a la distribución de manera que pasan por el “centro” de ésta. Hastie et al. [54] demostraron, bajo hipótesis poco restrictivas, que las curvas (o superficies)  $\mathbf{f}(x)$  minimizan la esperanza del error de aproximación

$$E[\|t - \text{Proy}_f(t)\|^2] \tag{3.39}$$

propiedad análoga a la del método PCA.

## 3.8 El Mapa Topológico Autoorganizado (SOM)

### 3.8.1 Introducción

El mapa topológico autoorganizado (en inglés, *self-organizing map* o, más brevemente, SOM) constituye una potente herramienta para la visualización de datos estadísticos multidimensionales. Básicamente [21], el SOM convierte relaciones estadísticas complejas y no lineales en sencillas relaciones geométricas en un mapa de baja dimensionalidad, habitualmente de dimensión inferior o igual a tres para permitir su visualización.

Una propiedad clave del SOM es la *conservación de la topología*. El SOM realiza una proyección no lineal de los datos a analizar sobre un espacio de baja dimensionalidad de manera que puntos próximos en el espacio de entrada se proyecten en puntos que también son próximos en el espacio de salida. Esta propiedad le permite revelar de manera visual las agrupaciones o *clusters* y las relaciones existentes en los datos a analizar. Tanto la reducción de la dimensionalidad como la conservación de la topología, hacen de este algoritmo una herramienta de enorme utilidad en la monitorización del estado de una máquina o proceso industrial.

A lo largo del capítulo se irán describiendo en detalle todas estas propiedades así como los algoritmos de proyección y entrenamiento del SOM.

### 3.8.2 Algoritmo de Proyección

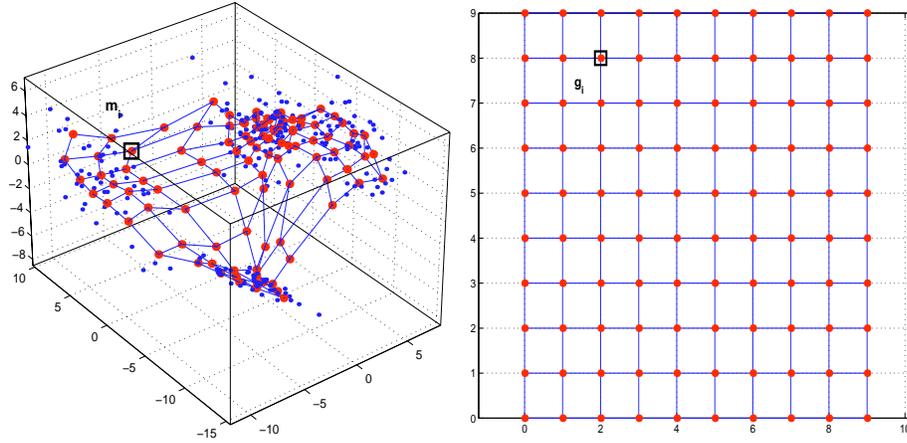
El algoritmo de proyección del SOM o *mapa de características* puede describirse como una proyección no lineal, ordenada y suave de un espacio de entrada multidimensional continuo sobre un espacio de salida discreto de baja dimensionalidad (habitualmente bidimensional).

Sea  $X$  un conjunto de puntos (datos) en el espacio de entrada  $\mathbb{R}^n$ . Cada elemento de  $X$  consta de  $n$  variables,

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T \quad (3.40)$$

El mapa de características  $S(\cdot)$  queda definido mediante una colección de  $K$  vectores  $\{\mathbf{m}_i\}_{i=1\dots K}$  del espacio de entrada, llamados *vectores de codificación*<sup>3</sup>, y otros  $K$  vectores  $\{\mathbf{g}_i\}_{i=1\dots K}$  del espacio de salida  $\mathbb{R}^p$ ,

<sup>3</sup>Dependiendo del contexto, los vectores  $\mathbf{m}_i$  pueden recibir también el nombre de



(a) Vectores de codificación  $\mathbf{m}_i$  en el espacio de entrada.

(b) Retícula formada por los  $\mathbf{g}_i$  en el espacio de salida.

Figura 3.9: Vectores de codificación  $\mathbf{m}_i$  y de retícula  $\mathbf{g}_i$  después de un entrenamiento con los datos del espacio de entrada (puntos azules).

habitualmente dispuestos ocupando las posiciones de una retícula rectangular, tal y como se aprecia en el ejemplo de la figura 3.9.

La imagen mediante el mapa de características  $S(\cdot)$  de cada punto del espacio de entrada sobre el espacio de salida se define como aquél punto de la retícula,  $\mathbf{g}_c$ , cuyo homólogo  $\mathbf{m}_c$  resulta ser el elemento más próximo a  $\mathbf{x}$  en el espacio de entrada de acuerdo con alguna medida de distancia  $d(\cdot, \cdot)$  definida en  $\mathbb{R}^n$ . Matemáticamente,

$$S(\mathbf{x}) = \mathbf{g}_c \quad \text{donde} \quad c = \arg \min_i \{d(\mathbf{x}, \mathbf{m}_i)\} \quad (3.41)$$

En la figura 3.9 se muestran los vectores de codificación  $\mathbf{m}_i$  en el espacio de entrada y sus correspondientes vectores de retícula en el espacio de salida tras un entrenamiento con los datos del espacio de entrada

---

*vectores de pesos sinápticos*; esta nomenclatura hace hincapié en su condición de red neuronal. En esta misma línea, se suele llamar también *neurona* o *unidad* al  $i$ -ésimo elemento de la red,  $\{\mathbf{m}_i, \mathbf{g}_i\}$ , formado por una posición en la retícula y un vector de pesos.

(puntos azules). En la figura se aprecia cómo la topología de la retícula (rectangular de  $10 \times 10$  en el dibujo) es la misma en ambos espacios a la vez que la posición de las unidades se adapta a la distribución de los datos en el espacio de entrada. Mediante un rectángulo se señala la correspondencia entre el vector de codificación  $\mathbf{m}_i$  y su correspondiente posición en la retícula  $\mathbf{g}_i$ .

Para que el algoritmo de proyección definido en (3.41) tenga alguna utilidad, es necesario definir los vectores de codificación  $\mathbf{m}_i$  de manera que la proyección describa adecuadamente la estructura subyacente en la distribución de los datos del espacio de entrada, como se puede apreciar en el ejemplo de la figura 3.9. A continuación se describen los algoritmos de aprendizaje más típicos para determinar los vectores de codificación.

### 3.8.3 Algoritmos de Aprendizaje

#### Cuantificación Vectorial (VQ)

Una forma de determinar los vectores de codificación es la aplicación de algoritmos clásicos de cuantificación vectorial (VQ). Este tipo de algoritmos buscan determinar los vectores  $\mathbf{m}_i$  de manera que su distribución de probabilidad se parezca lo más posible a la de los datos de entrada, es decir, buscan obtener una colección de puntos  $\mathbf{m}_i$  que sean representativos de la distribución de los datos de entrada. Para ello minimizan el *error de cuantificación promediado* [21]

$$E = \int f[d(\mathbf{x}, \mathbf{m}_c)]p(\mathbf{x})d\mathbf{x} \quad (3.42)$$

donde  $f$  es cierta función monótona creciente de la distancia  $d(\cdot, \cdot)$ . La función de coste (3.42) evalúa la distancia promediada de todos los vectores  $\mathbf{x}$  a sus vectores de codificación más próximos para una elección dada de los  $\mathbf{m}_i$ , en otras palabras, evalúa el error de cuantificación global cometido al aproximar cada elemento del espacio de entrada por el vector de codificación que lo representa.

Cuando el número de vectores de codificación coincide con el número de agrupaciones o *clusters* existentes en el espacio de entrada, los algoritmos VQ tienden a asociar un vector de codificación a cada cluster, dicho de otra forma, detectan agrupaciones.

Cuando se emplea un número suficientemente elevado, los vectores de codificación obtenidos mediante algoritmos VQ tienden a generar una distribución equiprobable de activaciones para cada elemento  $\mathbf{g}_i$  de la retícula, alojando mas elementos  $\mathbf{m}_i$  en regiones del espacio de entrada en las que mayor densidad de ejemplos hay y viceversa.

No obstante, los elementos son activados de manera arbitraria en la retícula ya que cualquier permutación de los índices de los vectores de codificación da lugar al mismo valor del error de cuantificación (3.42). Dicho de otra forma, la proyección *no es ordenada*, en el sentido de que datos próximos en el espacio de entrada no se corresponden, en general, con proyecciones próximas en la retícula del el espacio de salida.

Existen diversos algoritmos que permiten minimizar (3.42). Una solución que permite minimizar  $E = \int \|\mathbf{x} - \mathbf{m}_c\|^2 \cdot p(\mathbf{x})d\mathbf{x}$  puede derivarse (cfr. [55]) dando lugar a un algoritmo del tipo *descenso del gradiente*:

$$c = \arg \min_i \|\mathbf{x} - \mathbf{m}_i(t)\| \quad (3.43)$$

$$\mathbf{m}_c(t+1) = \mathbf{m}_c(t) + \alpha(t)[\mathbf{x} - \mathbf{m}_c(t)] \quad (3.44)$$

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t), \quad \forall i \neq c \quad (3.45)$$

donde el coeficiente de aprendizaje,  $\alpha(t)$ , es monótonamente decreciente con el tiempo.

Un método computacionalmente más eficiente y extensamente utilizado en la literatura es el *k-means* [55] [45] [47]. El método puede resumirse en los siguientes pasos:

- a) Elegir una colección arbitraria de vectores de codificación  $\mathbf{m}_i$
- b) Para cada vector de codificación  $\mathbf{m}_i$  generar una lista  $L_i$  con todos los datos  $\mathbf{x}$  para los cuales  $\mathbf{m}_i$  es el vector de codificación más próximo.
- c) Asignar a cada vector de codificación  $\mathbf{m}_i$  la media de los datos que forman su lista  $L_i$ .
- d) Repetir los pasos b) y c) hasta la convergencia.

Los dos algoritmos descritos son de carácter competitivo. En ambos, los elementos del SOM compiten entre sí para adaptarse. En el primero de

los algoritmos, se aprecia este hecho muy claramente: para cada iteración se adapta el elemento que vence. En el método k-means, la generación de las listas en base a las cuales se adapta cada elemento tiene también carácter competitivo.

### Algoritmo de Aprendizaje de Kohonen

La minimización de la función de coste (3.42), como se dijo en la sección anterior, no produce mapas ordenados en el sentido de que elementos  $\mathbf{g}_i$  próximos en el espacio de salida no se corresponden en general con elementos  $\mathbf{m}_i$  próximos en el espacio de entrada. De hecho, la solución a (3.42) padece una importante indeterminación al ser igualmente válida cualquier permutación entre los vectores de codificación hallados. Kohonen [55][21][22] propone para ello la introducción de una ligera modificación en (3.42): un *núcleo de suavizado local*,  $h_{ci}$  función de la distancia en el espacio de salida entre las unidades  $\mathbf{g}_c$  y  $\mathbf{g}_i$ :

$$E = \int h_{ci} \cdot f[d(\mathbf{x}, \mathbf{m}_c)]p(\mathbf{x})d\mathbf{x} \quad (3.46)$$

La introducción del núcleo  $h_{ci}$  induce un orden en los vectores de codificación obtenidos al minimizar (3.46). El carácter totalmente competitivo de los algoritmos de aprendizaje tipo VQ se combina ahora con un carácter *cooperativo*: no sólo debe mejorar la unidad vencedora sino también sus vecinas. El comportamiento del sistema es similar al de una “red elástica” en la que el núcleo  $h_{ci}$  ejerce el efecto de elemento de ligazón entre las unidades de la red.

Aunque no se conoce una solución directa para la minimización de (3.46), existen varios algoritmos que permiten obtener una solución aproximada para las  $\mathbf{m}_i$ . El algoritmo del SOM, propuesto por Kohonen [22] es el siguiente:

$$c = \arg \min_i \|\mathbf{x} - \mathbf{m}_i(t)\| \quad (3.47)$$

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \alpha(t)h_{ci}(t)[\mathbf{x}(t) - \mathbf{m}_i(t)] \quad (3.48)$$

El núcleo  $h_{ci}$  es, en general, una función decreciente con la distancia definida en la retícula del espacio de salida. Dos formas habituales para el

núcleo suelen ser las vecindades de tipo *burbuja* y *gaussiana*. La vecindad de tipo burbuja se define como:

$$h_{ci} = \begin{cases} 1 & \text{Si } d_g(\mathbf{g}_c, \mathbf{g}_i) \leq N_c \\ 0 & \text{en otro caso} \end{cases} \quad (3.49)$$

donde  $d_g(\cdot, \cdot)$  es una medida distancia definida en la retícula del espacio de salida<sup>4</sup> y  $N_c$  define la amplitud del campo receptivo, representando la distancia en la retícula a la neurona vencedora a partir de la cual las neuronas no responden al estímulo. La vecindad *gaussiana* se define como:

$$h_{ci} = e^{-\frac{d_g(c,i)^2}{\sigma^2}} \quad (3.50)$$

que define un campo receptivo cuyo tamaño viene dado por  $\sigma$  y cuya intensidad decrece monótonamente en función de la distancia  $d_g(\cdot, \cdot)$  definida en la retícula a la neurona vencedora.

A diferencia de la vecindad de tipo burbuja, la vecindad gaussiana hace que *todas* las neuronas del SOM se adapten –con mayor o menor intensidad– cada vez que llega un estímulo, decreciendo además la velocidad de adaptación de forma continua con la distancia a la ganadora. Esto mejora las propiedades de autoorganización y hace que el aprendizaje se asemeje más a los procesos biológicos de autoorganización. Por otro lado, la vecindad de tipo burbuja puede ser computacionalmente más eficiente en algunos casos al no requerir la adaptación de todas las neuronas con cada estímulo.

**Entrenamiento del SOM en modo batch.** Al igual que en el caso de los algoritmos VQ, existe una versión *batch* computacionalmente más eficiente:

- a) Se obtienen los índices de las unidades vencedoras para cada dato de ejemplo  $\mathbf{x}(k)$

$$c(k) = \arg \min_i \|\mathbf{x}(k) - m_i(t)\|, \quad k = 1, 2, \dots, N$$

---

<sup>4</sup>Medidas de distancia típicas son las de tipo enlace (número mínimo de enlaces a recorrer para ir de la unidad  $c$  a la unidad  $i$  en la retícula), tipo caja, euclídea, etc.

b) Se obtienen los nuevos  $\mathbf{m}_i$  mediante la media ponderada

$$\mathbf{m}_i(t+1) = \frac{\sum_k h_{c(k)i}(t) \cdot \mathbf{x}(k)}{\sum_k h_{c(k)i}(t)} \quad (3.51)$$

c) Repetir los pasos a) y b) decreciendo suavemente con el tiempo la amplitud del campo receptivo de  $h_{ci}$  hasta la convergencia.

Nótese que el algoritmo *k-means* se obtiene como caso particular de éste cuando el campo receptivo se reduce a un punto:

$$h_{ci} = \delta(c, i) = \begin{cases} 1 & \text{Si } i = c \\ 0 & \text{en otro caso} \end{cases} \quad (3.52)$$

### 3.8.4 Propiedades del mapa de características

El mapa de características definido en 3.8.2 revela importantes propiedades estadísticas de la distribución de los datos en el espacio de entrada.

**Aproximación de la función de densidad de los datos.** En [22] se sugiere la idea de que los vectores de codificación se distribuyen en el espacio de entrada con mayor probabilidad en regiones en las que la función de densidad de probabilidad de los datos,  $p(x)$ , es mayor y viceversa. De acuerdo con esto, el SOM dedica más neuronas a aquellas regiones del espacio de entrada con mayor probabilidad de aparición de datos. Dado que en el espacio de salida estas neuronas se disponen de forma regular, es evidente que regiones con más densidad de probabilidad se proyectan sobre zonas más amplias. En definitiva, el área dedicada a la representación de una característica determinada es, en cierta medida, proporcional a la frecuencia estadística de aparición de dicha característica.

Se define [47] el *factor de magnificación*  $m(\mathbf{x})$  como el número de neuronas por cada diferencial de volumen  $d\mathbf{x}$  del espacio de entrada. El factor de magnificación es tal que integrado en todo el espacio de entrada da el número total de neuronas del SOM:

$$\int_{\mathbb{R}^n} m(\mathbf{x}) \cdot d\mathbf{x} = K \quad (3.53)$$

La idea sugerida por Kohonen, expresada matemáticamente, sería

$$m(\mathbf{x}) \propto p(x) \tag{3.54}$$

En general, la relación (3.54) no se verifica exactamente. En el caso de mapas de características bidimensionales es difícil establecer una relación entre el factor de magnificación y la función de densidad. Ritter et al, la analizaron para el caso unidimensional mostrando que el factor de magnificación tiende asintóticamente a  $m(\mathbf{x}) \propto p(x)^r$ , donde el exponente  $r$  depende del número de neuronas que forman la vecindad de la vencedora y se encuentra aproximadamente entre  $r = 1/3$  y  $r = 2/3$ . No obstante, aunque la proporcionalidad no se verifica, la afirmación de que estímulos frecuentes inducen áreas mayores para su representación sigue siendo válida.

Existen, sin embargo diversas soluciones propuestas en la literatura científica dirigidas a la corrección del factor de magnificación. Bauer et al [56] proponen una modificación del algoritmo original del SOM basada en la modificación del coeficiente de aprendizaje,  $\alpha(t)$ , para cada neurona en función de la densidad de estímulos en la posición de cada neurona  $\mathbf{m}_i$ .

$$\alpha(t) = \alpha_0 P(\mathbf{m}_i)^m \tag{3.55}$$

siendo  $m$  un parámetro de diseño. En particular, para conseguir una proporcionalidad directa ( $r = 1$ ), Bauer et al. proponen  $m = 0.5$ .

La capacidad de aproximar la función de densidad de los datos tiene fuertes connotaciones biológicas. Desde un enfoque biológico, el SOM distribuye la información sensorial recibida en áreas de un mapa cuyos tamaños guardan relación directa con la frecuencia de los estímulos asociados a cada área. Se sabe [55] [22] que la información sensorial se organiza en el cerebro en mapas, esto es, las representaciones internas de información en el cerebro son organizadas, por lo general, espacialmente.

**Ordenación topológica.** Los vectores de codificación del SOM, tras el proceso de adaptación, se distribuyen a lo largo del espacio de entrada de manera que, además de distribuirse de forma similar a como lo hacen los datos, lo hacen de forma ordenada en el sentido de que las relaciones

de vecindad definidas en el espacio de salida (conexiones en la retícula) prevalecen en los vectores de codificación del espacio de entrada tras la convergencia. Puede afirmarse que:

*Nodos  $\mathbf{g}_i$ ,  $\mathbf{g}_j$  próximos en la retícula definida en el espacio de salida, se corresponden con vectores  $\mathbf{m}_i$ ,  $\mathbf{m}_j$  que también son próximos en el espacio de entrada.*

El SOM, tras la convergencia, puede verse como una deformación continua de los nodos de la retícula, como si se tratase de una *red elástica* que se extiende a lo largo de los datos del espacio de entrada, comprimiéndose en zonas de alta densidad y estirándose en zonas de baja densidad. En la figura 3.9 puede apreciarse visualmente este hecho.

La propiedad de ordenación topológica es una consecuencia directa del núcleo  $h_{ci}$  en la ecuación (3.48). Este factor implica el movimiento de adaptación, no sólo de la neurona vencedora  $\mathbf{m}_c$  sino también (aunque no necesariamente en la misma medida) de las neuronas próximas a ella en la retícula del espacio de salida. En definitiva, hace que neuronas próximas en la retícula reciban tratamientos similares durante el entrenamiento.

Esta propiedad se asemeja a un cierto tipo de continuidad del mapa de características al asegurar que la activación de neuronas próximas en la retícula corresponde a estímulos similares. A diferencia de los algoritmos VQ, puramente competitivos, el mapa de características no sólo detecta agrupaciones de estímulos sino que los visualiza próximos o lejanos en función de su grado de similitud.

En el caso concreto de trayectorias en el espacio de entrada (espacio de características), las trayectorias proyectadas por el mapa  $S(\cdot)$  son el resultado de una deformación seguida de una cuantificación sobre el espacio de la retícula. Si la conservación de la topología no se verificase (caso de algoritmos VQ competitivos puros), la trayectoria sufriría un desmenuzamiento y se traduciría en una secuencia aparentemente aleatoria de activaciones en los nodos de la retícula.

**Reducción de la Dimensionalidad.** Esta propiedad es una consecuencia directa de las anteriores. La capacidad de aproximar la geometría de la función de densidad conservando a la vez la topología definida en una retícula de dimensión  $p$  implican que el mapa resultante  $S(\cdot)$  resume en  $p$  coordenadas (las coordenadas o posiciones de la retícula) las

$n$  características del espacio de entrada. Esta propiedad está estrechamente relacionada con la idea del análisis de componentes principales (PCA). En ésta última técnica, un hiperplano de dimensión  $p$  se ajusta mínimo-cuadráticamente a la estructura de los datos; las proyecciones de los datos sobre el hiperplano determinan las  $p$  componentes principales.

En el caso del SOM, se realiza una proyección sobre una hipersuperficie de dimensión  $p$ , que se deforma ajustándose a los datos. Se trata, en cierto sentido de un análisis de componentes principales no lineal. Este tipo de análisis tiene importantes ventajas sobre el PCA clásico. El método PCA da resultados óptimos cuando la distribución de los datos es lineal (normal multivariante, por ejemplo); sin embargo, cuando las componentes están relacionadas de manera no lineal las PCA dan resultados poco satisfactorios. En la figura 3.10 se aprecia claramente. Los datos de ejemplo tienen claramente una componente principal (“largo” de la curva) y una secundaria (“ancho” de la curva). El análisis PCA es incapaz de reconocerlas mientras que el SOM detecta la componente más significativa sin problemas.

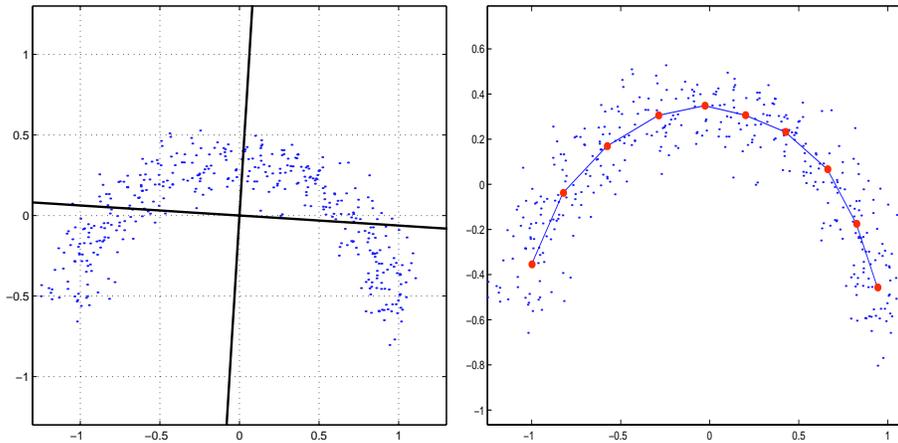
**Aproximación de los Datos de Entrada** Los vectores de codificación constituyen una buena aproximación de los datos del espacio de entrada, precisamente por la propiedad del SOM de aproximación de la función de densidad del espacio de entrada.

La función de coste (3.42) (y en un sentido más general, la función (3.46)) también conocida como *distorsión esperada* indica la esperanza del error cometido por sustituir un dato  $\mathbf{x}$  del espacio de entrada por su vector de codificación más próximo.

Esta propiedad permite codificar los datos de entrada dando lugar a una representación mucho más compacta. El proceso de sustituir cada dato  $\mathbf{x}$  por el vector de codificación más próximo es, en realidad un proceso de *cuantificación* en el espacio de entrada, similar al que ocurre al codificar un valor analógico mediante uno digital. El error cometido se conoce como *error de cuantificación*

$$q(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_c\|, \quad c = \arg \min_i \{\|\mathbf{x} - \mathbf{m}_i\|\} \quad (3.56)$$

El algoritmo (3.41) define una partición del espacio de entrada en  $K$  regiones de Voronoi,  $V_i \subset \mathbb{R}^n$ , donde cada  $V_i$  se define como aquellos puntos  $\mathbf{x}$  cuya distancia a  $\mathbf{m}_i$  es menor que la distancia a  $\mathbf{m}_j$  para cualquier



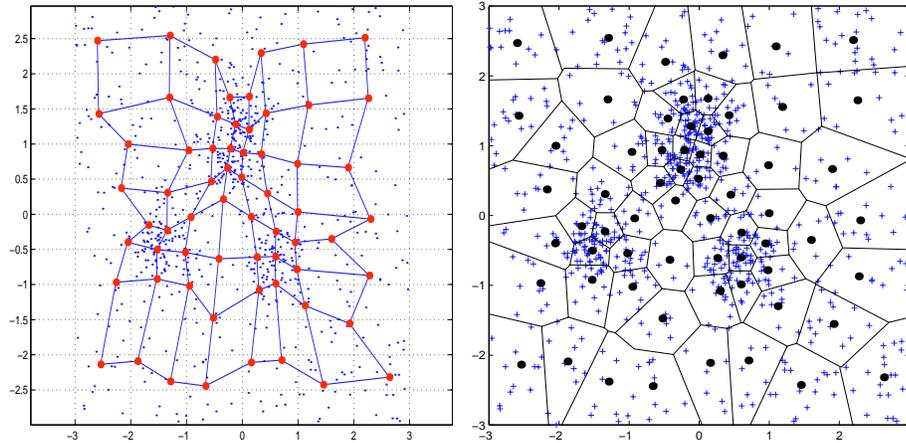
(a) Análisis lineal de componentes principales (PCA)

(b) Análisis no lineal de componentes principales mediante SOM

Figura 3.10: Comparación de SOM y PCA en la reducción de la dimensionalidad. Como puede apreciarse en la figura, el SOM ( $10 \times 1$ ) “capta” la estructura inherente a los datos y la resume en una dimensión (en este caso, la única coordenada de la retícula). El PCA, por el contrario, busca una tendencia lineal, lo que resulta insuficiente en datos de geometría no lineal.

$j \neq i$ . Nótese, que todos los puntos pertenecientes a una misma región de Voronoi  $V_i$  se proyectarán en la misma posición  $\mathbf{g}_i$  en el espacio de salida.

En la figura 3.11 se aprecian las regiones de Voronoi generadas por un SOM de  $7 \times 7$  en un espacio de entrada de dimensión 2. La propiedad de aproximación de la función de densidad hace que zonas con mayor densidad de estímulos sean representadas con más precisión que zonas poco pobladas. Esta propiedad es utilizada en algoritmos de compresión de datos e imágenes. En estos algoritmos se sustituyen bloques de datos por los vectores de codificación más próximos. Como ejemplo, utilizando 256 vectores, es posible codificar un bloque de  $n$  datos en coma flotante (vector de entrada  $\mathbf{x}$  de dimensión  $n$ ) mediante un sólo byte lo que permite llegar a razones de compresión del orden de varias decenas a uno, en función de la precisión deseada.



(a) Aproximación de datos 2D mediante SOM

(b) Partición del espacio de entrada en regiones de Voronoi

Figura 3.11: Partición del espacio de entrada en regiones de Voronoi. El SOM representa con mayor precisión las regiones con mayor densidad de estímulos, minimizando el error de cuantificación.

El error de cuantificación es mínimo cuando la función de densidad de los datos de entrada es similar a la utilizada durante el entrenamiento del SOM para determinar los vectores de codificación  $\mathbf{m}_i$ . Cuando la situación cambia, es decir, cuando los datos son generados con arreglo a otra ley o distribución, el error de cuantificación aumenta, lo que provee un mecanismo para determinar el grado de ajuste de la situación actual respecto a las situaciones utilizadas en el entrenamiento.

Este indicador resulta de gran valor en la monitorización de la condición de procesos, ya que un incremento del mismo revela una situación poco frecuente o imprevista. En [21], Kohonen describe una aplicación real en un sistema de administración de anestesia a un paciente en la que un incremento del error de cuantificación, junto con la observación de la trayectoria proyectada por el SOM, revela una obstrucción en el subsistema de intubación. En esta Tesis se utiliza, como se verá más adelante, este error como indicador de anomalías en la máquina.

## 3.9 Conclusiones

A lo largo de este capítulo se han analizado diversas técnicas de reducción de la dimensionalidad. Todas ellas modelan, de una u otra forma, las correlaciones existentes entre las variables con el objeto de eliminar las redundancias a las que dan lugar. El método PCA es una de las técnicas más empleadas por su efectividad y sencillez, aunque puede resultar ineficaz en el caso de datos no linealmente correlados.

El método Kernel PCA surge como una generalización del PCA. Este método se basa en proyectar los datos mediante un mapa no lineal sobre un espacio de elevada dimensionalidad en el que el problema puede tratarse por métodos lineales sin la necesidad de trabajar explícitamente en dicho espacio gracias al uso de núcleos. Sin embargo, puede ser computacionalmente ineficiente para muestras grandes, al crecer el tamaño de la matriz  $\mathbf{K}$  con el tamaño de la muestra. Por otra parte, la no existencia de criterios objetivos para la elección del núcleo y su dificultad de interpretación pueden ser un problema para su aplicación en la monitorización de la condición de procesos.

Los métodos MDS permiten generar proyecciones de los datos de entrada en un espacio de baja dimensionalidad preservando las distancias mutuas, lo que permite visualizar las relaciones geométricas más importantes, pero utilizan algoritmos que consumen mucha memoria y demandan un alto número de operaciones; además, no definen explícitamente un mapa entre ambos espacios, lo que los hace poco aptos para aplicaciones en tiempo real.

Otros métodos, como las redes neuronales feedforward autoasociativas, aunque tienen muy buenos tiempos de ejecución y dan buenos resultados, son más oscuros en cuanto a la interpretación, y carecen de criterios para la elección de la topología más adecuada de la red, por lo que su entrenamiento suele requerir a menudo un procedimiento heurístico.

El GTM, por el contrario, se deriva a partir de principios estadísticos y genera explícitamente un mapa funcional entre el espacio de proyección (espacio latente) y el espacio de entrada que maximiza la verosimilitud de la muestra. Este mapa es continuo y tiene propiedades similares a las del SOM.

El SOM obtiene resultados similares aunque por un procedimiento extraordinariamente simple y robusto, admitiendo además de forma natural generalizaciones en cuanto a la geometría de la superficie que aproxima

los datos del espacio de entrada, sin más que modificar la topología definida en la retícula. Pese a su carácter discreto, la generalización al caso continuo, como se verá en el siguiente capítulo, puede obtenerse empleando técnicas de regresión de núcleo. La arquitectura resultante muestra rápidos tiempos de ejecución al definir también un mapa funcional de forma explícita, lo que unido a su mayor facilidad de interpretación y flexibilidad la hacen idónea como técnica de proyección para su aplicación en la supervisión de procesos.

# Capítulo 4

## Generación de Mapas Continuos

### 4.1 El problema de la cuantificación del SOM en la Supervisión de Procesos

El algoritmo de proyección del SOM,  $S(\mathbf{x})$ , definido en (3.41), en la sección 3.8.2, plantea evidentes carencias debido al proceso de discretización que tiene lugar. El error cometido al representar un vector  $\mathbf{x}$  mediante un vector de codificación del SOM,  $\mathbf{m}_c$ , puede considerarse debido a dos factores:

- a) Un error de *cuantificación*, derivado del hecho de dividir el espacio de entrada en una colección finita de regiones de Voronoi,  $\{V_i\}_{i=1,\dots,K}$ , asociadas a cada uno de los vectores de codificación  $\{\mathbf{m}_i\}_{i=1,\dots,K}$ .
- b) Un error de *aproximación* o *modelado* que considera el resto de los factores que influyen en el error, derivados, por ejemplo, de valores subóptimos de los  $\mathbf{m}_i$ , o de una dimensionalidad intrínseca de los datos superior a la del modelo, que reflejan la no idoneidad del modelo para adaptarse a la geometría de los datos de entrada.

En la monitorización y supervisión del estado de un proceso, el error de cuantificación constituye una importante desventaja por varias razones. Muchos procesos industriales son, en esencia, continuos, y su estado

exhibe tendencias, derivas, ciclos límite, etc. que, a menudo, juegan un papel crucial a la hora de determinar la condición de los mismos. Cuando se aplica el mapa (3.41) a la trayectoria descrita por un vector de características o medidas del proceso,  $\mathbf{f}(t)$ , da lugar a una trayectoria proyectada sobre un espacio de visualización de carácter discontinuo:

$$\mathbf{f}_g(t) = S(\mathbf{f}(t)) \quad (4.1)$$

Esta trayectoria, aunque puede describir el comportamiento del proceso en grandes líneas, plantea dificultades para reflejar pequeños cambios en la condición de la máquina que podrían revelar información vital sobre la marcha del proceso. Así, por ejemplo, supóngase el proceso en un estado determinado  $\mathbf{x}(t)$ . Si en un momento determinado –por ejemplo, por un incremento paulatino de una de sus variables– el estado del proceso sufre una ligera desviación que se traduzca en el vector de características

$$\mathbf{f}(t + \Delta t) = \mathbf{f}(t) + \Delta\mathbf{f}(t) \quad (4.2)$$

la situación no se apreciará en la trayectoria proyectada  $\mathbf{f}_g(t)$  si el nuevo valor del vector de características,  $\mathbf{f}(t + \Delta t)$ , permanece aún en la región de Voronoi anterior. Dicho de otra forma, una perturbación  $\Delta\mathbf{f}(t)$  en el vector de características no se traducirá en una perturbación  $\Delta\mathbf{f}_g(t)$  en la trayectoria visualizada.

El problema de la cuantificación podría ser parcialmente solucionado incrementando el número de neuronas, lo que reduce el tamaño de las regiones de Voronoi y, por consiguiente, la probabilidad de pasar por alto alteraciones incipientes en el proceso. Sin embargo, esto redundaría en un mayor coste computacional y, aunque en menor medida, el problema subsiste.

## 4.2 Interpolación mediante Regresión de Núcleo

El problema de la cuantificación puede paliarse mediante la generación de un mapa continuo utilizando técnicas de interpolación multidimensional. Una extensión natural del algoritmo de proyección del SOM puede expresarse utilizando el *Estimador de Regresión Nadaraya-Watson (NWR)*:

$$\mathcal{S}(\mathbf{x}|\mathbf{x}_i, y_i) = \frac{\sum_i \Phi(\|\mathbf{x} - \mathbf{x}_i\|)y_i}{\sum_i \Phi(\|\mathbf{x} - \mathbf{x}_i\|)} \quad (4.3)$$

donde el núcleo  $\Phi(\mathbf{x})$  es una función continua, acotada y real de  $\mathbf{x}$ , y simétrica en torno al origen, en el que toma su máximo valor [47]. Una de las formas más habituales es el núcleo gaussiano de base esférica:

$$\Phi(\mathbf{x}) = e^{-\frac{1}{2}\|\mathbf{x}\|^2/\sigma^2} \quad (4.4)$$

que, en su forma más general, viene dada por,

$$\Phi(\mathbf{x}) = e^{-\frac{1}{2}\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}} \quad (4.5)$$

donde  $\mathbf{C}$  es una matriz simétrica que define un hiperelipsoide en el espacio de  $\mathbf{x}$ .

La expresión (4.3) recibe también el nombre de GRNN (*Generalized Regression Neural Network*) dado que puede ser implementada mediante una arquitectura masivamente paralela de neuronas similares y aproxima la superficie de regresión de los datos a medida que el número de ejemplos se aproxima al infinito. La regresión de núcleo es una técnica firmemente asentada en diversas áreas como en predicción de series temporales por su sencillez y eficacia. En [57], Yee y Haykin demuestran que ésta equivale a una forma regularizada de redes de base radial (RBF), lo que permite encuadrarla dentro de éste último tipo de redes.

El estimador de Nadaraya-Watson tiene una sólida base estadística. Puede demostrarse, [58], [47], que éste es un estimador consistente de la función  $f(\mathbf{x})$  para el siguiente problema de regresión:

$$y_i = f(\mathbf{x}_i) + \epsilon, \quad i = 1 \dots, N \quad (4.6)$$

donde la solución óptima para este problema es la esperanza de  $y$  condicionada a  $\mathbf{x}$ , es decir,

$$f(\mathbf{x}) = E[y|\mathbf{x}] = \int_{\mathbb{R}^p} y p(y|\mathbf{x}) dy = \int_{\mathbb{R}^p} \frac{y \cdot p(\mathbf{x}, y)}{p(\mathbf{x})} dy \quad (4.7)$$

donde  $p(y|\mathbf{x})$  es la densidad de probabilidad condicional de un vector  $y$  en el espacio de salida  $\mathbb{R}^p$  (en nuestro caso, el espacio de visualización) dado un vector  $\mathbf{x}$  en el espacio de entrada (espacio de características y/o medidas)  $\mathbb{R}^n$ . Como puede comprobarse, el estimador Nadaraya-Watson no requiere el conocimiento de  $p(\mathbf{x}, y)$  ni el de  $p(\mathbf{x})$ , utilizando tan sólo una muestra  $\{(\mathbf{x}_i, y_i)\}_{i=1, \dots, N}$  representativa de la función de densidad conjunta  $p(\mathbf{x}, y)$ .

En [22] (ver también sección 3.8.4 en esta tesis), Kohonen sugiere la idea de que los vectores de codificación se distribuyen en el espacio de entrada con más densidad en aquellos puntos en los que existe mayor densidad de ejemplos o estímulos. Dicho de otra forma, se atribuye al SOM cierta capacidad de aproximar la función de densidad del espacio de entrada. Existen, sin embargo, evidencias de que esto no es estrictamente cierto. Algunos autores [56], [59], incluso el propio Kohonen [22] reconocen que la aproximación no es exacta.

Una magnitud que describe adecuadamente el grado de aproximación es el *factor de magnificación*  $m(\mathbf{x})$ , que representa la densidad de neuronas en un entorno suficientemente pequeño de  $\mathbf{x}$  en el espacio de entrada. De acuerdo con esta definición, una aproximación ideal de la función de densidad de los datos de entrada implicaría

$$m(\mathbf{x}) \propto p(\mathbf{x}) \quad (4.8)$$

en cuyo caso se alcanzaría un mapa de máxima entropía y todas las neuronas se excitarían con la misma probabilidad siempre que los vectores de entrada  $\mathbf{x}$  sigan distribuyéndose con arreglo a la función de densidad  $p(\mathbf{x})$ .

Sin embargo, aún no está claro bajo qué condiciones es cierta la proporcionalidad directa, existiendo diversos resultados en los que se demuestra, bajo ciertas condiciones, que

$$m(\mathbf{x}) \propto p(\mathbf{x})^\mu \quad (4.9)$$

donde, en general,  $0 < \mu < 1$ . Esto significa que regiones en el espacio de entrada con bajas densidades son sobrerrepresentadas por el SOM mientras que, por el contrario, regiones con altas densidades de estímulos son infrarrepresentadas.

Existen, no obstante, trabajos en los que se proponen ligeras modificaciones en el algoritmo de entrenamiento del SOM orientadas a corregir la distorsión del factor de magnificación. DeSieno [59] propone el denominado “algoritmo de consciencia”, que considera el número de veces que cada neurona vence en orden a modificar el proceso competitivo dando menos oportunidades de ganar a aquellas neuronas con mayor tasa de estímulos recibidos. Bauer et al [56] sugieren, por otra parte, un método basado en la adaptación del coeficiente de aprendizaje –denotado por  $\alpha(t)$  en la ecuación (3.48)– en el SOM.

Si  $m(\mathbf{x}) \propto p(\mathbf{x})$ , en el supuesto de que los estímulos  $\mathbf{x}$  se ajusten a  $p(\mathbf{x})$ , todas las neuronas se activarán con la misma probabilidad. Esto conduce a una distribución espacialmente uniforme (una binomial, al tratarse de un caso discreto) en el espacio de visualización, siempre que los elementos de la rejilla estén distribuidos de forma regular y equiespaciada en el espacio de visualización. Bajo esta hipótesis se dan tres circunstancias:

- a) Los vectores de codificación  $\mathbf{m}_i$  constituyen una muestra cuya distribución en el espacio de características es proporcional a la de los estímulos que llegan del proceso,  $p(\mathbf{x})$ .
- b) Los elementos de la rejilla  $\mathbf{g}_i$  constituyen una muestra distribuida uniformemente en el espacio de visualización.
- c) La correspondencia entre cada  $\mathbf{g}_i$  y cada  $\mathbf{m}_i$  no es arbitraria sino que, dada la capacidad del SOM de aproximar la topología, es tal que elementos próximos en la rejilla se corresponden con elementos próximos en el espacio de características<sup>1</sup>. Bajo otro punto de vista, la preservación de la topología hace que el SOM elija, entre todas las permutaciones posibles que caben para asignar cada posición  $\mathbf{g}_i$  con cada vector de codificación  $\mathbf{m}_i$ , aquella que menor distorsión provocaría a la hora de extrapolar al caso continuo.

Según lo visto anteriormente, parece plausible la idea de que los elementos  $(\mathbf{m}_i, \mathbf{g}_i) \in \mathbb{R}^{n+p}$  constituyen muestras representativas para aproximar una función o mapa continuo que, aplicado a una variable aleatoria continua de distribución uniforme en el espacio de visualización generaría una variable aleatoria cuya distribución se aproximaría a  $p(\mathbf{x})$ . La propiedad de consistencia (i.e. asintóticamente insesgado según crece el tamaño de la muestra) del estimador de Nadaraya-Watson junto con lo indicado

<sup>1</sup>Esta propiedad evoca, en cierto sentido, la definición de *continuidad uniforme*:

*Una función  $f$  se dice uniformemente continua en  $D$  si para todo  $\epsilon > 0$  existe un  $\delta > 0$  tal que para todo  $x, x' \in D$  tales que  $|x - x'| < \delta$  se verifica  $|f(x) - f(x')| < \epsilon$*

es decir, si dos puntos son próximos en el espacio origen, lo son también sus homólogos en el espacio imagen. Esto, a nuestro juicio, parece evidenciar la idoneidad de las muestras  $\mathbf{m}_i, \mathbf{g}_i$  que genera el SOM con vistas a obtener mediante interpolación la función continua de la que supuestamente proceden.

anteriormente sugiere la elección de este estimador para la generación de mapas continuos basados en el SOM:

$$\mathcal{S}(\mathbf{x}|\mathbf{m}_i, \mathbf{g}_i) = \frac{\sum_i \Phi(\|\mathbf{x} - \mathbf{m}_i\|)\mathbf{g}_i}{\sum_i \Phi(\|\mathbf{x} - \mathbf{m}_i\|)} \quad (4.10)$$

Dado que este modelo continuo deriva del SOM y de la aplicación de técnicas de regresión de núcleo (*Kernel Regression*), con el objeto de abreviar, se utilizará el término KR-SOM para describir la aplicación combinada de ambas técnicas.

### 4.3 Otras Alternativas

La obtención de un mapa continuo del espacio de entrada del SOM sobre el espacio de salida a partir de los vectores de codificación  $\mathbf{m}_i$  en el espacio de entrada y sus correspondientes nodos  $\mathbf{g}_i$  en la retícula del espacio de salida constituye un caso de interpolación multidimensional con nodos irregularmente espaciados. La obtención de un mapa en estas condiciones que verifique ciertas condiciones de regularidad y suavidad y al mismo tiempo constituya una buena aproximación no es sencillo, dedicándose aún hoy en día un considerable esfuerzo investigador en este sentido.

El método de regresión de núcleo, sugerido en esta tesis, además de las justificaciones que se detallan en el apartado anterior, destaca por su simplicidad, y su comportamiento asintóticamente óptimo cuando crece el número de neuronas. No obstante, existen en la literatura otras alternativas orientadas a obtener una generalización continua del SOM mediante interpolación que merece la pena destacar.

#### 4.3.1 Modelos Lineales Locales

En [24] y [26], Goppert y Rosenstiel emplean una aproximación de tipo iterativo basada en modelos locales. Cada modelo local está constituido por las  $k$  neuronas mejor adaptadas al vector  $\mathbf{x}$  o bien las  $k$  neuronas más próximas a la vencedora según la topología definida en el SOM. Un algoritmo iterativo sobre el modelo local correspondiente al vector  $\mathbf{x}$  del espacio de entrada permite obtener simultáneamente su aproximación  $\hat{\mathbf{x}}$  así como su homólogo  $\hat{y}$  en el espacio de salida (visualización). Este

método no es óptimo y requiere varias iteraciones por cada vector  $\mathbf{x}$  a proyectar. Por otra parte, el vector  $\mathbf{x}$  induce modelos locales diferentes en función de las neuronas que le sean más próximas lo que da lugar a mapas que no son continuos (los mapas son, en realidad, continuos a trozos). Los propios autores han propuesto modificaciones posteriores basadas en el promediado ponderado de varias soluciones aunque sólo paliar en parte el problema añadiendo complejidad y más parámetros al algoritmo.

### 4.3.2 Algoritmo PSOM

Una solución elegante, aportada por Ritter [60], [91], [92] es el PSOM. El PSOM (*Parameterized SOM*), utiliza un conjunto de funciones base<sup>2</sup> en el espacio de la retícula y construye a partir de ellas una hipersuperficie suave que aproxima la geometría de las  $\mathbf{m}_i$ , lo que define un mapa del espacio de salida (rejilla equiespaciada) sobre el espacio de entrada. Este mapa es de gran calidad, precisamente por la regularidad de los nodos de las funciones base, estructurados de forma equiespaciada. El paso inverso, es decir, la proyección de puntos  $\mathbf{x}$  del espacio de entrada sobre el espacio de salida (espacio de visualización) se hace mediante iteración, utilizando métodos de optimización de segundo orden tipo *Levenberg-Marquardt*. La solución aportada por Ritter es de gran elegancia y resuelve de manera muy eficaz el enorme problema planteado por la dificultad de interpolar con datos irregularmente espaciados en el espacio origen. Sin embargo, el algoritmo *Levenberg-Marquardt* es de carácter iterativo, e implica además el cálculo del gradiente y la matriz Hessiana del error lo que puede hacerlo inviable computacionalmente.

### 4.3.3 Núcleos Hiperelipsoidales

Otra línea de investigación estrechamente ligada a la interpolación por regresión de núcleo se centra en la optimización de los parámetros de las funciones base. Los núcleos esféricos, como los utilizados en el estimador NWRE, funcionan bien con datos distribuidos con cierta uniformidad pero se muestran ineficientes cuando la distribución de éstos es poco regular.

---

<sup>2</sup>Ritter sugiere como funciones base la extensión multidimensional de los conocidos polinomios de Lagrange, mediante el producto de polinomios definidos para cada dimensión.

Cuando la distribución estadística de los datos en el espacio de entrada es poco regular, los núcleos esféricos pueden llegar a ser activados por datos que deberían activar unidades vecinas o, por el contrario, pueden no ser suficientemente activados por datos del espacio de entrada que deberían activarlos. En este caso, a medida que se incrementa el número de núcleos en el estimador NWRE la dependencia de los parámetros del núcleo es menor (propiedad de consistencia), pudiendo conseguirse un buen estimador utilizando un número suficiente de núcleos esféricos a costa de una mayor carga computacional.

Durante el desarrollo de esta tesis se estudió la posibilidad de utilizar como núcleos funciones de base radial (RBF's) de simetría hiperelipsoidal. Parece razonable, en este sentido, la idea de que la utilización de núcleos hiperelipsoidales permite una mayor adaptación de las funciones base a la geometría de la función de densidad, pudiendo aliviar, en parte, los problemas comentados anteriormente. No obstante, el problema es complejo y hasta donde se ha podido indagar durante el desarrollo de esta tesis, varias cuestiones quedaron abiertas. En primer lugar, no existe, en principio, un criterio objetivo que permita determinar la bondad de un determinado juego de matrices de covarianzas para las funciones base. Un aspecto clave, a nuestro juicio, reside en la función de densidad del espacio de entrada. En un principio pareció sugerente la idea aportada por Tipping y Bishop en [61], donde proponen “capturar” la complejidad de la distribución de los datos mediante una combinación de modelos estadísticos locales lineales PCA. Cada uno de estos modelos describiría la distribución localmente según una normal multivariante de media  $\mathbf{c}_i$  y matriz de covarianzas  $\mathbf{C}_i$ . La idea, por tanto, consiste en considerar funciones base con centros  $\mathbf{c}_i$  y covarianzas  $\mathbf{C}_i$  tales que aproximan mediante mezcla gaussiana la función de densidad de los ejemplos en el espacio de entrada.

Probada esta idea para varios casos, los resultados no fueron totalmente satisfactorios fundamentalmente por dos razones:

- a) Se empleó el método EM (*Expectation Maximization*), utilizado por Tipping y Bishop para el cálculo de las matrices de covarianzas de cada unidad, tomando como centros los  $\mathbf{m}_i$  del SOM. Como pudo comprobarse a través de varios ejemplos, en ocasiones da lugar a soluciones degeneradas (matrices de covarianzas con valores propios muy dispares) para algunas de las unidades, lo que implica una

discriminación absoluta en cuanto a la activación de dicha unidad por parte de datos del espacio de entrada situados en las direcciones menos preferentes.

- b) Por otra parte, aun en casos en los que las matrices son balanceadas, los resultados no siempre fueron claramente mejores que con el uso de núcleos esféricos. Pudo comprobarse que las direcciones preferentes lo eran demasiado siendo demasiado activadas por datos que, aunque alejados, caían en dichas direcciones.

En la figura adjunta, se muestra uno de los casos en los que se probó la eficacia de los núcleos hiperelipsoidales. La figura muestra datos procedentes de una distribución generada sintéticamente deformando una uniforme mediante una transformación no lineal. Junto a ella se muestra cómo los vectores del SOM se ajustan a la distribución, aproximando la función de densidad.

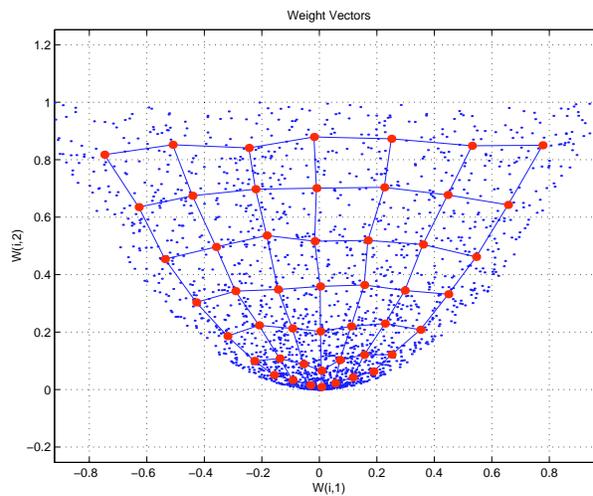


Figura 4.1: Datos de ejemplo en el espacio de entrada.

Tras aplicar el método EM, y realizar las proyecciones de cada uno de los puntos mediante el estimador NWRE empleando los tres tipos de núcleos (esféricos iguales, esféricos obtenidos con EM e hiperelipsoidales obtenidos con EM), los resultados obtenidos fueron los que se muestran en la figura 4.2:

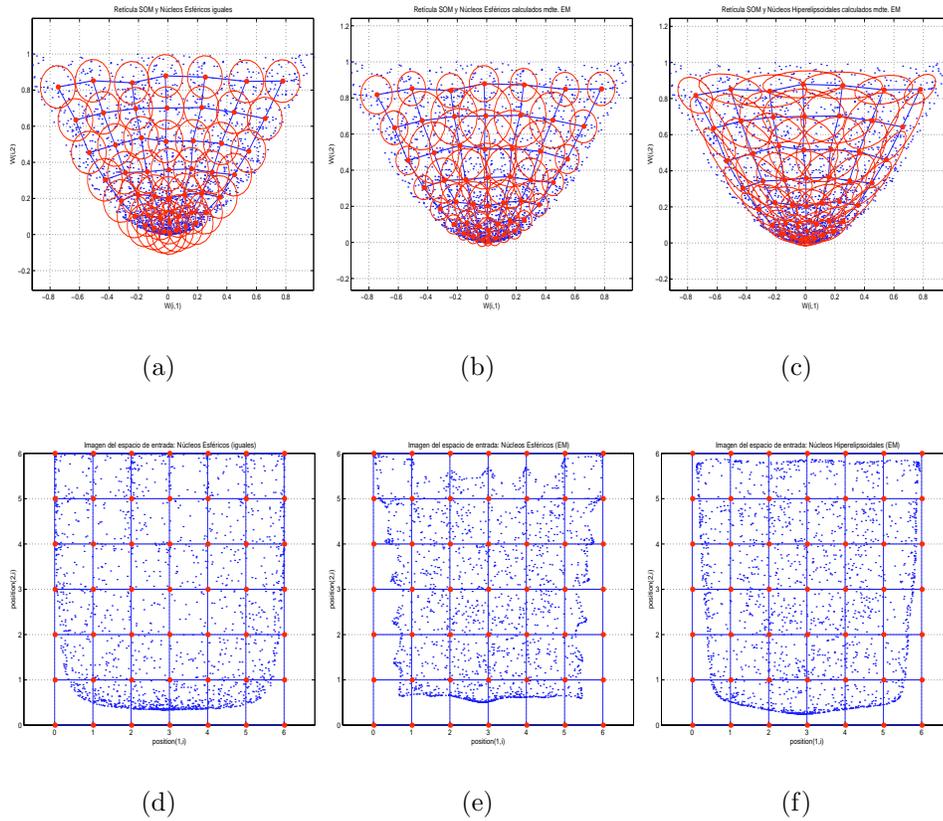


Figura 4.2: Comparación del uso de núcleos esféricos, núcleos esféricos distintos (uno por neurona) y núcleos hiperelipsoidales. Como puede apreciarse, los resultados en las proyecciones no mejoran mucho los resultados arrojados por los núcleos esféricos.

El objetivo ideal sería conseguir una distribución uniforme en el espacio de visualización, lo que implicaría que el mapa no lineal generado es el de máxima entropía. Esta situación implicaría que el mapa continuo generado habría hecho un desdoblamiento óptimo de la función de densidad.

Puede apreciarse que el uso de núcleos esféricos distintos parece que incluso empeora la situación. Aparecen muchas zonas con alta densidad de puntos proyectados junto con zonas vacías. En los núcleos hiperelipsoidales, parece apreciarse una ligera mejoría respecto al uso de núcleos esféricos iguales. Mientras que en éstos últimos se produce una concen-

tración de puntos proyectados debida a la no adecuación del tamaño de los núcleos en la zona de alta densidad de los datos de entrada, en el caso de núcleos hiperelipsoidales, la adaptación del tamaño de los núcleos es mejor y los puntos proyectados son algo más uniformes. Un problema que puede apreciarse en los tres casos es el efecto borde, debido al hecho de que el SOM no ubica neuronas fuera de la región en la que están los datos, lo que origina que los puntos periféricos se proyecten en las regiones periféricas del espacio de visualización.

Pese a la leve mejoría con el uso de núcleos hiperelipsoidales, no parece ser rentable el esfuerzo requerido para determinar los núcleos en relación con el beneficio obtenido. No obstante, lo que sí parece indudable es que la mejor solución no es necesariamente la de núcleos esféricos, siendo, a nuestro juicio, de enorme interés el estudio de un criterio que permita establecer una configuración óptima de los parámetros de las funciones base.



# Capítulo 5

## Metodología para Monitorización de la Condición basada en SOM Interpolado

### 5.1 Introducción

En los capítulos anteriores, de acuerdo con el enfoque dado en la sección 1.4, en la que se formula el problema esencial de esta Tesis, se han descrito diversas técnicas orientadas a reducir la dimensionalidad preservando la información relativa a la condición del proceso.

La reducción de la dimensionalidad se realiza en varios pasos sucesivos: planificación del sistema de medidas (ubicación, tipo, número de sensores, etc.), técnicas de extracción de características y técnicas específicas de reducción de la dimensionalidad. A lo largo de este proceso, el grado de conocimiento *a priori* empleado en cada paso decrece mientras que el nivel de abstracción crece. Si los pasos adoptados verifican ciertas condiciones, como se describe en 1.4, el resultado final es un mapa que proyecta el espacio de estados sobre un espacio de visualización que, a efectos de la determinación de la condición, puede considerarse equivalente al espacio de estados.

A lo largo de este capítulo, se describe una metodología para la aplicación del SOM interpolado (KR-SOM) que presupone ya realizadas las

etapas anteriores, a saber, el diseño del sistema de medidas y la extracción de las características más importantes. Aunque la proyección de la trayectoria de estado sobre el espacio de visualización es fundamental en la metodología propuesta en esta Tesis, de nada serviría sin el conocimiento de la “orografía” de este espacio, es decir, la información que permite relacionar la condición del proceso con la posición del puntero en el espacio de visualización.

El presente capítulo, además de describir la forma de aplicar el KR-SOM para la proyección, tiene como objetivo proponer varias técnicas para extraer la información sobre la citada “orografía” que se encuentra latente en los vectores de codificación  $\mathbf{m}_i$  del SOM.

## 5.2 Proyección del Estado del Proceso

El algoritmo de proyección tiene una enorme importancia en la monitorización y supervisión del estado de procesos dado que la proyección del estado sobre el espacio de visualización define un “puntero” que describe en todo momento la condición del proceso en función de las regiones que atraviese en dicho espacio.

Como se argumentó en la sección 1.4, bajo ciertas condiciones impuestas al sistema de medidas, el vector de medidas o características puede considerarse como una función del estado del proceso que preserva información relativa a la condición, sugiriéndose como tema central de esta Tesis la búsqueda de una transformación posterior que permita reducir la dimensionalidad hasta conseguir un espacio imagen visualizable, preservando al mismo tiempo la información relativa a la condición.

En el capítulo 3 se analizaron diversas técnicas de reducción de la dimensionalidad entre las que destaca el SOM por sus propiedades de preservación de la topología, aproximación de la función de densidad y reducción de la dimensionalidad, que permiten describir mediante un mapa 2D (ó 3D) las características más importantes de la geometría de los datos a analizar, constituyendo, de hecho, un verdadero análisis de componentes principales no lineal. Especialmente, la propiedad de *preservación de la topología*, mediante la cual neuronas próximas en el espacio de entrada se corresponden con neuronas próximas en la retícula (espacio de visualización) evoca el concepto de *continuidad uniforme* – aunque, por supuesto, no en un sentido estricto– a lo largo de todo el

dominio en el que se define el mapa. Esta propiedad permite, en general, que la topología de las regiones definidas en el espacio de entrada se conserve, induciendo regiones en el espacio de visualización cuya geometría y posiciones relativas se asemejan a las existentes en el espacio de entrada.

En el capítulo 4 se propone una arquitectura fundamentada en núcleos de base radial centrados en las neuronas del SOM (Ec. 4.3), que generaliza el mapa del SOM al caso continuo. Si la preservación de la topología puede verse como una propiedad de “continuidad global” –hablando en un sentido amplio–, la arquitectura propuesta goza, además, de *continuidad local*. El SOM puede verse entonces como una verdadera “lámina elástica”<sup>1</sup> que se amolda a los datos doblándose, estirándose y encojiéndose para producir un mapa en el que las condiciones del proceso están asociadas a regiones concretas del espacio de visualización.

Dado un vector de características del proceso,  $\mathbf{f}(t)$ , su proyección o trayectoria  $\mathcal{T}(t)$  sobre el espacio de visualización será

$$\mathcal{T}(t) = \mathcal{S}(\mathbf{f}(t)|\mathbf{m}_i, g_i) = \frac{\sum_i \Phi(\|\mathbf{f}(t) - \mathbf{m}_i\|)g_i}{\sum_i \Phi(\|\mathbf{f}(t) - \mathbf{m}_i\|)} \quad (5.1)$$

## 5.3 Mapas de Características

En el SOM resulta posible formar un mapa de características en el espacio de visualización. Aprovechando la disposición uniforme de las unidades en la retícula del espacio de visualización, es posible generar  $n$  mapas discretos –uno por cada característica– formándose el mapa  $k$ -ésimo por asignación de la coordenada  $k$ -ésima de  $\mathbf{m}_i$  a la celdilla definida por la posición  $g_i$  en la retícula. En la práctica suele asignarse un nivel de gris o pseudocolor a cada uno de los “píxeles” formados por las unidades, formando mapas o planos de características.

Estos mapas de características son de gran interés en la supervisión del estado del proceso dado que conectan un aparato matemático como el visto hasta ahora, en apariencia abstracto, con variables físicas del proceso o características dotadas de significado. Debe tenerse en cuenta aquí que no es extraño encontrar características con un significado físico

<sup>1</sup>Kohonen emplea el término “elastic net” en [21]. La palabra “net” (red, retícula) es más apropiada en el SOM discreto, en el que sería menos apropiado hablar de lámina o sábana, que evocan la continuidad del mapa.

tangible ya que, a menudo, las características de una máquina o proceso son escogidas en base a criterios fundamentados en el conocimiento físico del mismo.

Los mapas de características, junto con la trayectoria proyectada superpuesta a ellos, permiten al operario determinar de manera muy visual la evolución de una característica determinada en el proceso. Si la característica es, por ejemplo, una temperatura en los cojinetes y se forma el mapa utilizando una gama de pseudocolor que denote con tonos cálidos (rojos) valores elevados y mediante tonos fríos (azules) valores bajos, podrá advertirse, de un simple vistazo, la temperatura de los cojinetes en cada momento sin más que ver si el puntero se posa sobre zonas rojas o azules.

La generación de mapas continuos es directa, aplicando (4.3):

$$\mathcal{F}(u) = \mathcal{S}(u|g_i, \mathbf{m}_i) \quad (5.2)$$

El vector  $\mathcal{F}(u) = [f_1(u), f_2(u), \dots, f_n(u)]^T$  para cada punto  $u$  del espacio de visualización define los  $n$  planos. Si, el espacio de visualización es bidimensional, pueden formarse mapas representando en 2D los puntos  $[u_1, u_2, f_k(u)]^T$  mediante una posición o pixel  $u = (u_1, u_2)$ , y el valor de la  $k$ -ésima componente,  $f_k(u)$ , en pseudocolor.

## 5.4 Mapas de Activación

El SOM almacena la información adquirida durante la fase de entrenamiento organizándola espacialmente. En particular, [62], cuando cada señal en el espacio de características es representada por una excitación *local* en una posición específica del *campo neuronal*<sup>2</sup> se tiene un *mapa* de las señales sobre dicho campo. Por su capacidad de preservar la topología, las regiones estimulares<sup>3</sup> del espacio de entrada inducen mediante el

---

<sup>2</sup>Amari utiliza el término *campo neuronal* para describir un continuo de neuronas similar al córtex cerebral. El SOM, no es sino una versión discreta de ese continuo, aunque la aplicación de técnicas de interpolación como la de Nadaraya-Watson, propuesta en esta Tesis, se ajusta más a los conceptos descritos por Amari en [62] basados en sistemas autoorganizados de carácter continuo.

<sup>3</sup>Utilizamos aquí la palabra “estímulo” para poner de manifiesto las connotaciones biológicas que conlleva la forma en la que el SOM almacena la información. En el caso de procesos, un estímulo puede representar una condición o estado del mismo. En

SOM otras tantas regiones en el espacio de visualización topológicamente estructuradas de forma similar a las primeras. Paralelamente, por su capacidad de aproximar la función de densidad de los estímulos de entrada, el SOM destina más neuronas en aquellas zonas con mayor densidad de estímulos, con lo que las regiones inducidas en el espacio de visualización serán mayores cuanto más intenso sea el estímulo al que representan. El resultado es el siguiente:

- a) Una subdivisión difusa del espacio de visualización en regiones que representan otros tantos tipos o grupos de estímulos similares.
- b) Las regiones son más amplias cuando los estímulos son más intensos, es decir, cuando existe una mayor densidad de ejemplos en la región del espacio de entrada a la que representan.
- c) Regiones próximas entre sí representan estímulos similares, mientras que regiones lejanas representan estímulos dispares.

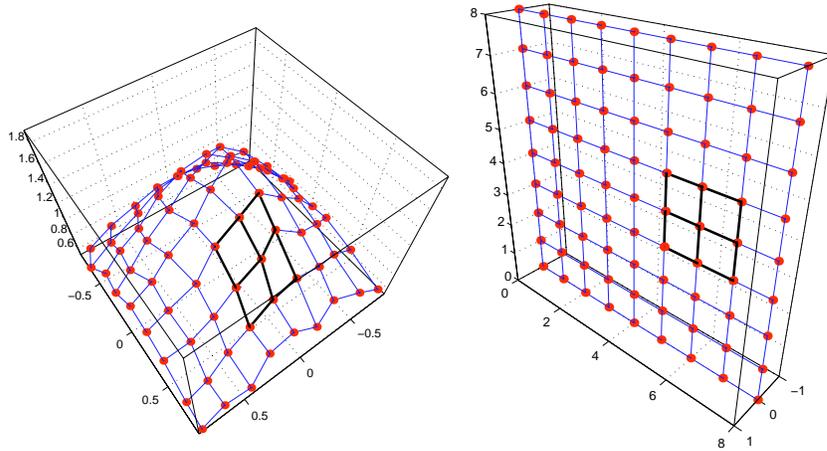
Estas regiones, sin embargo, no se hacen patentes de forma directa en el SOM. No obstante, pueden ser determinadas fácilmente a partir de los estímulos de entrenamiento evaluando los *mapas de activación*, formados por los índices de activación de cada neurona ante la totalidad de los estímulos considerados. Para determinar la región de activación correspondiente a un subconjunto  $S$  de los estímulos de entrenamiento,  $S = \{\mathbf{x}_k\}_{k \in I_S}$ , formado por  $N_S$  elementos, donde  $I_S$  es el conjunto de índices que agrupa los estímulos cuya región de activación se pretende determinar, es preciso considerar el mecanismo de interacción lateral inherente al SOM.

Un principio básico (cfr. sección 3.8) que interviene en el entrenamiento del SOM es la *interacción lateral*. Este principio es, precisamente, lo que distingue al SOM del algoritmo VQ y el principal responsable de sus propiedades de autoorganización. Durante la fase de adaptación, en el entrenamiento, para cada estímulo  $\mathbf{x}_k$  el SOM adapta no sólo la

---

[22] y en [55], Kohonen describe cómo el cerebro de los animales superiores organiza la información en *mapas cerebrales*. Es un hecho constatado que el cerebro destina regiones específicas para tratar estímulos específicos. Así, por ejemplo, el cerebro destina regiones diferentes del córtex para el reconocimiento de palabras, audición, visión, etc.

unidad  $\mathbf{m}_c$  más próxima a éste, sino que también adapta, aunque en menor medida, las unidades definidas como vecinas en la topología de la retícula.



(a) Efecto de un estímulo representado en el espacio de entrada.

(b) Efecto del mismo estímulo representado en el espacio de visualización. La situación es muy similar a la que se da en el córtex cerebral.

Figura 5.1: Principio de cooperatividad del SOM. Ante un estímulo procedente de un dato del espacio de características, reaccionan (se activan) la neurona vencedora y sus vecinas según la topología definida por la retícula, marcadas en trazo grueso.

Una unidad  $\mathbf{m}_i$  se adaptará al estímulo  $\mathbf{x}_k$  con un factor  $h_{ci}$  que está directamente relacionado con el grado de proximidad a la vencedora,  $\mathbf{m}_c$ , en la retícula. En definitiva, en el SOM junto al carácter *competitivo* rige un principio de carácter *cooperativo*: lo que le sucede a una neurona afecta directamente a sus vecinas.

La determinación de los índices de activación de cada neurona debe poner de manifiesto este carácter cooperativo. Parece razonable, por tanto, definir la activación acumulada de cada unidad tras la introducción del estímulo  $S = \{\mathbf{x}_k\}_{k \in I_S}$  como:

$$A_i = \frac{\sum_{k \in I_S} h_{c(k)i}}{\sum_j h_{ij}} \quad (5.3)$$

donde,

$$c(k) = \arg \min_i \{d(\mathbf{x}_k, \mathbf{m}_i)\}, \quad k \in I_S \quad (5.4)$$

En la expresión anterior, cada neurona incrementa su activación por cada estímulo en función de su proximidad reticular a la neurona vencedora. El término del denominador es una medida de la influencia total ejercida mediante interacción lateral por todas las neuronas. Este factor permite considerar el efecto borde derivado de la discriminación que sufren las neuronas periféricas al tener una vecindad más pequeña. Al igual que en casos anteriores, la generación de mapas de activación continuos es directa aplicando (4.3):

$$\mathcal{A}(u) = \mathcal{S}(u|g_i, A_i) \quad (5.5)$$

Los mapas de activación son de enorme utilidad en la supervisión y monitorización de procesos. El trazado en pseudocolor de los mapas de activación sobre el espacio de visualización hace patentes las regiones activadas por un estímulo o conjunto de estímulos determinados. Esto permite correlar estados conocidos del proceso utilizados durante el entrenamiento del SOM con la condición actual del mismo. Así, por ejemplo, tras el entrenamiento de un SOM utilizando ejemplos de un motor de inducción (cfr. sección 6.1) sometido a cuatro estados diferentes –situación normal, asimetría en la alimentación, asimetría mecánica en el rotor, o simultáneamente asimetría en la alimentación y asimetría mecánica en el rotor–, es posible determinar, por ejemplo, la región de activación asociada a la asimetría en la alimentación tomando el conjunto de ejemplos correspondiente a dicha situación y utilizando (5.3)–(5.5). Una incursión de la trayectoria  $\mathcal{T}(t)$  en la región de activación revelaría que el proceso está aportando estímulos muy similares a los de dicha situación, al tiempo que aportaría una medida del grado de coincidencia.

## 5.5 Mapas de Distancias Interneuronales

Las distancias interneuronales representan un buen indicador de la densidad de datos de entrada y ayudan a reconocer agrupaciones (*clusters*) de

datos en el espacio de entrada y regiones con baja probabilidad de ocurrencia. Al igual que en la determinación de activaciones, parece plausible recurrir al principio de *cooperatividad* para determinar las distancias. Un valor representativo de la distancia interneuronal puede obtenerse promediando la distancia entre dos neuronas  $i, j$ , por su grado de vecindad  $h_{ij}$ , lo que conduce a la siguiente expresión:

$$D_i = \frac{\sum_j h_{ij} d(\mathbf{m}_i, \mathbf{m}_j)}{\sum_j h_{ij}} \quad (5.6)$$

Una representación continua de la distancia interneuronal puede ser obtenida utilizando, al igual que en las expresiones anteriores, el estimador de Nadaraya-Watson:

$$\mathcal{D}(u) = \mathcal{S}(u|g_i, D_i) \quad (5.7)$$

donde  $u$  es un punto del espacio de visualización (habitualmente bidimensional).

## 5.6 Error de Modelado.

El SOM (cf. sección 3.8), al igual que los algoritmos de cuantificación vectorial, induce una partición del espacio de entrada en tantas regiones de Voronoi como unidades empleadas. Cada punto del espacio de entrada es representado por la unidad del SOM más próxima a él, de acuerdo con cierta medida de distancia definida en dicho espacio. En otras palabras, el SOM describe el espacio de entrada mediante una colección discreta de puntos, a saber, los vectores de codificación  $\mathbf{m}_i$ , produciéndose una *cuantificación* del mismo. En la sección 3.8 se describió el *error de cuantificación* como el error cometido al aproximar un vector  $\mathbf{x}$  del espacio de entrada por su neurona más próxima,  $\mathbf{m}_c$ :

$$q(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_c\|, \quad c = \arg \min_i \{\|\mathbf{x} - \mathbf{m}_i\|\} \quad (5.8)$$

En el SOM continuo basado en el estimador de Nadaraya-Watson propuesto en esta Tesis, no se produce un fenómeno de cuantificación propiamente dicho, al tratarse de un mapa continuo, siendo más apropiado hablar de *error de aproximación* o *error de modelado*.

El principal problema que se plantea aquí es que no existe explícitamente una “entidad” –una superficie continua– en el espacio de entrada respecto a la cual referir los errores de aproximación.

En [11], Wilson e Irwin utilizan el concepto de *autoconsistencia* para resolver un problema muy similar. La autoconsistencia describe la capacidad de un modelo para reconstruir, con un error mínimo, el propio dato que le llega. Para ello Wilson e Irwin, hallan la transformación inversa del dato proyectado sobre el espacio de visualización, con lo cual obtienen realmente un punto de la “superficie principal”.

**Error de modelado global:** En realidad, la operación descrita por Wilson e Irwin equivale a proyectar, de acuerdo con el modelo, sobre el propio espacio de entrada en lugar de hacerlo sobre el espacio de visualización. El modelo KR-SOM descrito en el capítulo precedente permite de forma sencilla y directa determinar la proyección del dato actual sobre el propio espacio de entrada:

$$\hat{\mathbf{x}} = \mathcal{S}(\mathbf{x}, |\mathbf{m}_i, \mathbf{m}_i) \quad (5.9)$$

El conocimiento de  $\hat{\mathbf{x}}$  permite describir la autoconsistencia del modelo mediante la diferencia entre el dato actual  $\mathbf{x}$  y su proyección  $\hat{\mathbf{x}}$ :

$$\mathcal{E}(\mathbf{x}) = \|\mathbf{x} - \hat{\mathbf{x}}\| = \|\mathbf{x} - \mathcal{S}(\mathbf{x}, |\mathbf{m}_i, \mathbf{m}_i)\| \quad (5.10)$$

Este concepto es análogo al de *residuo* o *redundancia analítica* descritos en el apartado 1.2.2. De hecho, el SOM constituye un modelo del proceso creado mediante la información procedente de los datos de entrenamiento. El error de modelado (5.10) es, en realidad, un residuo, cuyo valor tiende a cero cuando el proceso se ajusta al modelo aprendido. De acuerdo con la filosofía de los métodos de redundancia analítica, este error constituye un parámetro de enorme importancia en la monitorización de la condición de procesos ya que un incremento del mismo refleja una desviación del proceso respecto al modelo de comportamiento derivado de los ejemplos, lo que permite detectar la presencia de anomalías.

**Error de modelado vectorial:** La expresión (5.10) corresponde a un error del modelo o residuo de carácter *global*, aportando información

sobre la desviación total del proceso respecto a su comportamiento esperado. Evidentemente, es posible determinar la contribución a dicho error de cada una de las variables del modelo, sin más que considerar las componentes del vector diferencia  $\mathbf{x} - \hat{\mathbf{x}}$ , en lugar de hallar su módulo, como en (5.10):

$$\vec{\mathcal{E}}(\mathbf{x}) = \mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - \mathcal{S}(\mathbf{x}, |\mathbf{m}_i, \mathbf{m}_i) \quad (5.11)$$

Este error, que podemos denominar *error de modelado vectorial*, aporta dos ventajas esenciales:

- a) En primer lugar, permite determinar por separado la desviación de cada variable respecto del modelo. Ello aporta una información de enorme utilidad, ya que hace patentes las variables que más se desvían del modelo, facilitando la determinación de las causas de la anomalía.
- b) Las desviaciones contienen signo, lo que permite conocer si la desviación en una variable determinada es por defecto o por exceso.

En la figura 5.2 se muestra un ejemplo “de juguete” en el que se ilustran el error de modelado absoluto y el vectorial. Como puede apreciarse, el error de modelado global aporta una información en términos absolutos de la desviación del proceso respecto al modelo; no en vano representa la distancia entre el dato actual y la estimación más plausible del mismo compatible con el modelo. El error de modelado vectorial, sin embargo, permite descomponer el error en las dos componentes  $x$  e  $y$ , lo que aporta una información adicional. Puede comprobarse, por ejemplo, cómo para puntos como  $(x, y) = (0, -4)$  (cruz gruesa, en la figura) el error de modelado se atribuye más a la componente  $y$  que a la componente  $x$ , lo que parece un juicio razonable a la vista de la figura.

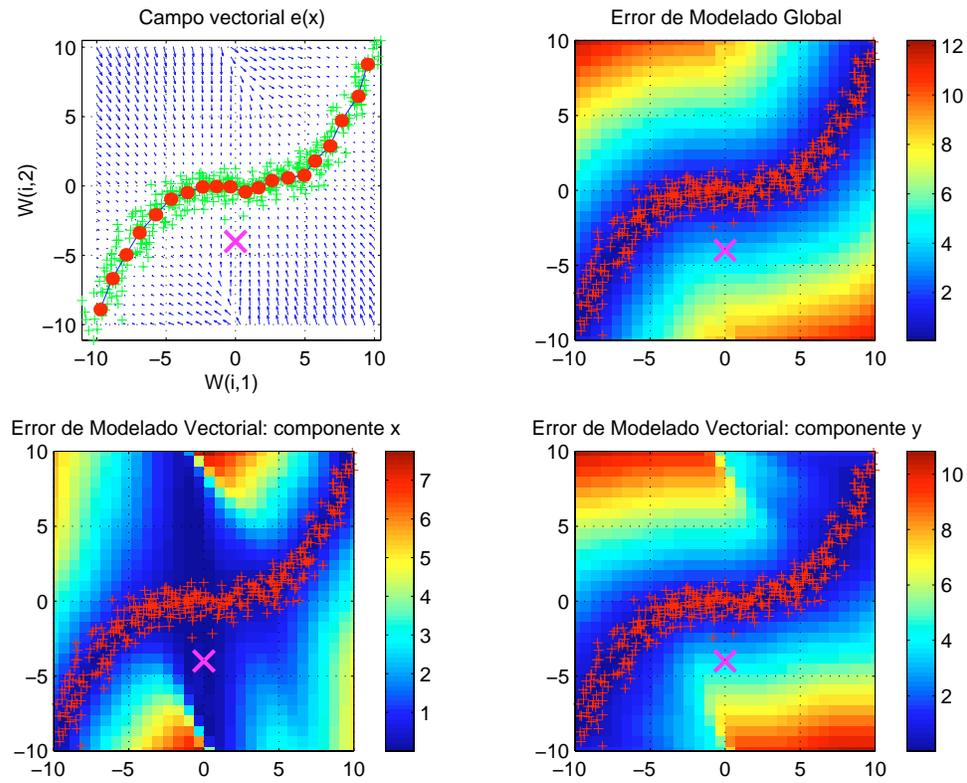


Figura 5.2: Error de modelado vectorial.

## 5.7 Visualización 3D mediante técnicas MDS

### 5.7.1 Introducción

Las técnicas de escalado multidimensional o técnicas MDS (cf. sección 3.4) constituyen un potente método para la visualización de los estados del proceso. El principal inconveniente de estos métodos reside en su enorme carga computacional, lo que hace impensable su aplicación cuando el volumen de los datos a analizar es elevado. Dado que el SOM tiende a aproximar la densidad de probabilidad de los datos de entrenamiento, los métodos MDS pueden aplicarse a los propios vectores de codificación,  $\mathbf{m}_i$ , los cuales constituyen una muestra representativa de los datos de entrada –i.e., están distribuidos de forma similar–, para visualizar de forma aproximada en 3D la geometría de los datos en el espacio de entrada [23].

Esta técnica ofrece un enorme potencial en la visualización y monitorización de procesos complejos, más allá de la mera visualización de los datos descrita en [23], gracias a la filosofía KR-SOM expuesta en la sección 4.2. A lo largo de esta sección se propone una serie de técnicas de visualización y monitorización del proceso enraizadas en el mismo principio de interpolación multidimensional empleado en el KR-SOM.

### 5.7.2 Visualización 3D de la geometría espacial de los datos en continuo

Como se indicó en 3.4, el principal objetivo de las técnicas MDS es la conservación de la matriz de distancias mutuas o, al menos, de alguna función de éstas. En un sentido amplio, esto equivale a preservar su “forma” o geometría<sup>4</sup>.

Según se describió anteriormente, la regresión de núcleo (4.3) permite realizar una interpolación entre puntos  $\{\mathbf{x}_i, y_i\}$  pertenecientes a espacios  $\mathbb{R}^n$  y  $\mathbb{R}^m$  de distinta dimensionalidad. La expresión (4.3), además, define de forma explícita una función continua  $\mathcal{S} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Cuando la dimensión  $n$  del espacio origen es inferior a la dimensión  $m$  del espacio imagen, el mapa  $\mathcal{S}$  describe una hipersuperficie de dimensión  $n$  sobre el espacio imagen  $\mathbb{R}^m$ , que aproxima la relación entre los puntos  $\mathbf{x}_i$  del espacio origen y los puntos  $y_i$  del espacio imagen.

<sup>4</sup>Nótese que esta condición es una generalización de la condición de invarianza de las distancias mutuas en un *solido rígido*.

Con estas consideraciones, cabe proponer el siguiente método para la visualización 3D de la geometría de los datos o características de un proceso dado:

- a) Se parte de una colección de vectores de codificación  $\{\mathbf{m}_i\}_{i=1,\dots,N}$  obtenidos tras el entrenamiento del SOM y sus correspondientes posiciones en la retícula  $\{\mathbf{g}_i\}_{i=1,\dots,N}$ .
- b) Aplicar un método MDS (p.ej. CCA) a los vectores de codificación  $\{\mathbf{m}_i\}_{i=1,\dots,N}$  dando lugar a un conjunto de vectores  $\{\mathbf{m}_i^*\}_{i=1,\dots,N}$  de  $\mathbb{R}^3$
- c) La proyección 3D de elementos del espacio de retícula se define aplicando regresión de núcleo con origen en el espacio de retícula e imagen en  $\mathbb{R}^3$ :

$$\mathbf{x}^* = \mathcal{S}_{\mathbf{g} \rightarrow 3D}(\mathbf{x}) = \mathcal{S}(\mathbf{x}|\mathbf{g}_i, \mathbf{m}_i^*) \quad \mathbf{x} = (x, y) \in \mathbb{R}^2$$

- d) La proyección 3D de elementos del espacio de características se define aplicando regresión de núcleo con origen en el espacio de medidas o características e imagen en  $\mathbb{R}^3$ :

$$\mathbf{x}^* = \mathcal{S}_{\mathbf{m} \rightarrow 3D}(\mathbf{x}) = \mathcal{S}(\mathbf{x}|\mathbf{m}_i, \mathbf{m}_i^*) \quad \mathbf{x} = (f_1, \dots, f_n)$$

El método descrito permite extender las posibilidades de representación hasta ahora descritas –proyección de elementos del espacio de características sobre el espacio de retícula en 2D– permitiendo visualizar en 3D y de forma continua el mismo tipo de representaciones (trayectoria proyectada, mapas de características, distancias interneuronales, mapas de activación, etc.), con la diferencia de que en lugar de un rectángulo en  $\mathbb{R}^2$  se tiene una superficie curva en  $\mathbb{R}^3$  cuya topología guarda una estrecha relación con la geometría de los datos del proceso. Esta representación tiene, a nuestro juicio, las siguientes ventajas:

- a) Al ser  $\mathcal{S}_{\mathbf{g} \rightarrow 3D}$  un mapa continuo, es posible trazar la superficie 3D del SOM mediante la transformación de puntos  $\mathbf{x}_i = (x_i, y_i)$  del espacio de la retícula según

$$\mathbf{x}_i^* = \mathcal{S}_{\mathbf{g} \rightarrow 3D}(\mathbf{x}_i) \tag{5.12}$$

donde los  $\mathbf{x}_i^*$  son los puntos de la superficie 3D. Resulta conveniente elegir puntos equiespaciados para proyectar (p.ej. formando una fina retícula cuadrículada). De esta manera los puntos proyectados aparecerán en la superficie 3D con una densidad directamente relacionada con la densidad de los ejemplos.

b) La trayectoria del proceso se obtiene directamente mediante:

$$\mathbf{f}^*(t) = \mathcal{S}_{\mathbf{m} \rightarrow 3D}(\mathbf{f}(t)) \quad (5.13)$$

La representación simultánea en 3D de la trayectoria y la superficie del SOM permite apreciar mayor cantidad de información sobre el estado del proceso. La información sobre el grado de proximidad entre los diversos estados del proceso se aporta visualmente por dos vías:

- Por una parte, por la propiedad de conservación de distancias de los métodos MDS, las distancias entre los  $\{\mathbf{m}_i^*\}$  son similares a las distancias entre los  $\{\mathbf{m}_i\}$ , es decir, la distancia entre dos puntos en la representación 3D se corresponde aproximadamente con la distancia real de los dos estados a los que representan<sup>5</sup>.
  - Simultáneamente, puede visualizarse sobre la superficie en 3D el mapa de distancias interneuronales en pseudocolor, lo que permite visualizar regiones con mayor densidad de datos, correspondientes a estados frecuentes del proceso.
- c) Es posible visualizar cualquiera de los mapas descritos en secciones anteriores asignando un nivel de pseudocolor a cada punto de la superficie 3D en función del valor de la característica a representar (activaciones, distancias interneuronales, características del proceso, etc.). Sencillamente, se asigna a cada punto  $\mathbf{x}^*$  de la superficie el mismo nivel de pseudocolor  $c$  que el correspondiente punto  $\mathbf{x}$  del espacio de la retícula.

---

<sup>5</sup>Esto es cierto para estados que sean bien predichos por el modelo. Para puntos en el espacio de entrada alejados de la superficie del SOM, la interpolación  $\mathcal{S}_{\mathbf{m} \rightarrow 3D}$  no conserva las distancias

## 5.8 Aplicación Práctica en Procesos Industriales

A lo largo del presente capítulo se han expuesto diversas técnicas que permiten extraer la información sobre el proceso contenida en las neuronas del SOM así como correlacionar dicha información con la situación actual del proceso, definida por la proyección sobre el espacio de visualización. En la presente sección se describirán algunos aspectos prácticos en cuanto a la aplicación de la metodología propuesta en la supervisión de procesos industriales.

La aplicación del KR-SOM en la supervisión de procesos puede resumirse en el flujograma de la figura 5.3. A lo largo de los siguientes apartados se describen con más detalle algunos de los elementos presentes en el flujograma.

### 5.8.1 Generación de Ejemplos

Como paso previo a la implementación de un sistema de diagnóstico y supervisión de procesos debe tenerse especial cuidado en la generación de los vectores que van a ser tratados por KR-SOM. El diseño puede desglosarse en tres fases:

**Adquisición de datos:** En esta fase deben seleccionarse el número, tipo, ubicación, calidad, etc. de los sensores que van a aportar información sobre el proceso. Dada su gran dependencia de la aplicación, el conocimiento experto de la planta es esencial en esta fase.

**Preprocesado:** En esta etapa se utilizan técnicas sencillas para adecuar los datos para su posterior tratamiento, incluyendo tareas como el filtrado, normalización de escalas, etc.

**Extracción de características:** Esta fase tiene un gran componente de conocimiento *a priori* sobre el proceso y una gran dependencia de la aplicación. Exige, por ello, un proceso cuidadoso que, en ocasiones, requiere ajustar muchos parámetros. Así, en *variables dinámicas* –aquéllas en las que la forma de onda es importante, típicamente vibraciones, co-



rrientes, tensiones, pares electromagnéticos, flujo axial, etc.–. puede ser necesario determinar, entre otros, los siguientes parámetros:

- Periodo de muestreo.
- Tamaño de las ventanas de datos.
- Solapamiento entre ventanas consecutivas.
- Tipo de inventanado (hamming, hanning, etc.)
- Tipo de transformación (FFT, Cepstrum, técnicas HOSA, wavelets, etc.) más adecuada para el problema a tratar.
- Selección de los coeficientes más significativos.
- Extracción de otros tipos de características tales como kurtosis, nivel global de la señal, distorsión armónica total (THD), energía de la señal en bandas concretas, etc.

En *variables estáticas* –en las que sólo importa su valor en un momento determinado, tales como temperaturas, presiones, niveles globales de vibración, etc.– las técnicas de preprocesado (filtrado, normalización de escalas) suelen ser suficientes, pudiendo utilizarse, en ocasiones, un análisis PCA para eliminar redundancias debidas a dependencias lineales entre dos o más de éstas variables.

### 5.8.2 Base de datos de entrenamiento

Los métodos de supervisión basados en datos, como el que se propone en esta Tesis, requieren una *base de datos de entrenamiento* con los cuales se elabora un modelo de funcionamiento del proceso. Dicha base debe incorporar datos de diversas procedencias:

**Datos del Proceso:** El primer paso debe ir orientado a introducir en dicha base datos<sup>6</sup> disponibles del proceso funcionando éste en las distintas condiciones  $C_i$  que se pretenden monitorizar.

---

<sup>6</sup>Se entiende aquí por *datos* aquellos que resultan de aplicar las etapas de preprocesado y extracción de características a las medidas del proceso.

**Datos generados mediante Modelo:** A menudo, no se dispone de datos del proceso para algunas de las condiciones en las que éste puede funcionar. En tales casos, la disponibilidad de un *modelo físico del proceso* podría permitir generar datos del proceso funcionando en dichas condiciones.

**Datos del Proceso recopilados en línea:** Finalmente, la base de datos de entrenamiento debe ser alimentada con datos procedentes del proceso durante la fase de explotación según van apareciendo situaciones nuevas. El criterio más adecuado para seleccionar qué datos se incorporan a la base es el *error de modelado*  $\mathcal{E}(\mathbf{x})$ . Aquellos datos con un error de modelado superior a un umbral preestablecido pueden ser considerados procedentes de una situación del proceso no modelada y deben ir a formar parte de la base de datos entrenamiento.

### 5.8.3 Entrenamiento del Modelo

**Generación del modelo:** El entrenamiento del SOM con interpolación mediante regresión de núcleo tiene dos partes:

- a) Los parámetros de entrenamiento del SOM. En el entrenamiento del SOM deben considerarse diversos parámetros:
  - Número de épocas de entrenamiento.
  - Amplitud del campo receptivo  $h_{ci}(t)$ .
  - Distancia definida en la retícula. Casi siempre se consideran distancias de *tipo enlace* (distancia en enlaces del camino más corto entre las dos neuronas a considerar). Otras distancias habituales son las de tipo *caja* y tipo *hexagonal*.
  - Ritmo de aprendizaje  $\alpha(t)$  (no necesario en Batch SOM).
  - Número de neuronas de la retícula.
  - Topología de conexión de la retícula. Habitualmente la retícula es rectangular. No obstante, podrían ser interesantes otras topologías (toroidal, cúbica, en árbol, etc.) si se conocen aspectos de la geometría de los datos que las justifiquen.

En la sección 3.8 puede encontrarse una descripción más detallada sobre el significado e influencia de cada uno de estos parámetros en la calidad del modelo obtenido.

- b) El factor de ancho  $\sigma_i$  de las funciones de base radial. Una elección sencilla que da buenos resultados es la distancia promedio entre neuronas vecinas. Existen diversos trabajos proponiendo otras alternativas<sup>7</sup> aunque según las pruebas realizadas durante el desarrollo de esta Tesis (véase la sección 4.3.3), los resultados obtenidos en la interpolación no parecen justificar el esfuerzo.

#### 5.8.4 Interpretación de los Resultados

El KR-SOM tiene como principal objetivo determinar una proyección del estado del proceso sobre un espacio de visualización 2D ó 3D (campo neuronal) sobre el que se pueden definir una serie de mapas que facilitan su interpretación. A lo largo de este capítulo se han descrito diversas técnicas que permiten extraer la información contenida en el KR-SOM, la cual se hace patente mediante regiones en los mapas. A continuación se resumirá la utilidad y las principales características de los distintos tipos de mapas:

**Mapas de Activación:** Los mapas de activación describen la región del campo neuronal activada por un dato o conjunto de datos concreto. Esto permite determinar las regiones asociadas a las condiciones de las cuales hay disponibles datos de proceso. Al mismo tiempo, dado que cada punto del mapa tiene asociado un valor de activación, es posible determinar el grado de coincidencia de un estado concreto del proceso con la condición que dio lugar al mapa de activación.

**Mapas de Distancias:** Los mapas de distancias describen gráficamente en el campo neuronal las agrupaciones (clusters) más importantes entre los datos de proceso. Estas agrupaciones corresponden a las condiciones

---

<sup>7</sup>Existen en la literatura otras formas de determinar los factores de ancho para cada neurona, válidos también para el caso general de núcleos elipsoidales (factor de ancho expresado mediante matrices de covarianzas), basados en principios estadísticos (método EM, *Expectation Maximization*, cfr. [61]).

más significativas del proceso entre los datos utilizados para el entrenamiento. Dada la propiedad de preservación de la topología del SOM, la proximidad entre dos regiones en el mapa revela proximidad en el espacio de los datos, y por tanto que las condiciones son similares. Ello se traduce en la posibilidad de determinar visualmente, por su organización en el mapa, posibles estructuraciones jerárquicas entre las condiciones.

**Mapas de Características:** Los mapas de características resultan de ubicar en el campo neuronal, en pseudocolor, los valores de cada característica para cada uno de los puntos del mapa. Dado que las características, en general, tienen un significado físico claro (e.g. armónico de 50 Hz, nivel de vibración global, etc.), los mapas permiten al usuario identificar las distintas regiones del espacio de visualización con valores altos o bajos de cada característica. Comparándolas, por similitud, con los mapas de activación, permiten deducir visualmente (o automáticamente, con técnicas propias de visión artificial) las *influencias* de cada variable en las distintas condiciones. Una comparación entre los propios mapas de características, permite también apreciar las *correlaciones* existentes entre las distintas variables.

**Error de Modelado:** El error de modelado tiene una gran relación con los conceptos de *residuo* ó *error de redundancia analítica*, utilizados en los métodos basados en modelos. El error de modelado, en condiciones de funcionamiento del proceso compatibles con los datos de entrenamiento, adopta valores típicamente bajos. Un incremento de éste por encima de un umbral determinado revela alteraciones en el proceso respecto a las condiciones consideradas en los vectores de entrenamiento. Esto permite avisar de un posible fallo aunque sin especificar cuál. El *error de modelado vectorial* (vector diferencia entre el dato y su estimación), permite determinar cuál de las variables contribuye más a la desviación, lo que aporta una pista muy valiosa para la determinación de la causa del fallo. A continuación se describen algunas situaciones que pueden darse junto con su significado y las actuaciones que deben tomarse:

**Error de modelado alto:** Significa que el comportamiento del proceso difiere del comportamiento modelado. La causa más probable es la aparición de una nueva condición del proceso no contemplada en los datos de entrenamiento.

*Actuación:* La política más adecuada es incorporar los nuevos datos al conjunto de datos de entrenamiento y reentrenar el modelo. Paralelamente, debe investigarse la etiología de la nueva condición para identificar la región que le corresponde en el nuevo mapa y proceder a su etiquetado. Puede obtenerse una “pista” de la causa observando en el error de modelado vectorial, qué variable o variables contribuyen más al error de modelado.

**Error de modelado persistentemente alto:** Esta situación es típica de un cambio en la dinámica del proceso respecto a la que tenía el conjunto de datos de entrenamiento de los que se obtuvo el modelo.

*Actuación:* Si se sospecha que el cambio en la dinámica es *circunstancial*, se aconseja añadir los nuevos datos al conjunto de datos de entrenamiento y reentrenar el modelo. Si el cambio parece *definitivo*, es más aconsejable eliminar los datos antiguos del conjunto de entrenamiento y reentrenar.

**Error de modelado bajo y resultados contradictorios:** Esta situación revela una incapacidad del modelo para representar la condición del proceso, sugiriendo como causa más probable la *ausencia de información* en las medidas o características obtenidas del proceso, lo que permitiría que distintas condiciones del mismo puedan dar lugar a las mismas medidas o características.

*Actuación:* Se aconseja verificar los sensores ante posibles fallos en éstos. En ausencia de fallos en los sensores, el sistema puede necesitar incorporar alguna nueva medida, relacionada con la nueva condición. También puede ocurrir que la información esté presente en las medidas pero la etapa extracción de características oculte información, lo que obligaría a replantear dicha etapa.

En la tabla 5.1 se resumen los distintos tipos de representación de la información que pueden obtenerse mediante el KR-SOM junto con la información que aportan sobre la condición del proceso.

**CAPÍTULO 5. METODOLOGÍA PARA MONITORIZACIÓN DE LA  
CONDICIÓN BASADA EN SOM INTERPOLADO**

128

<b>Tipo de Representación</b>	<b>Información que aporta</b>
Mapas de Activación	Regiones $R_i$ asociadas a condiciones $C_i$ . Nivel de severidad de fallos.
Mapas de Distancias	Agrupaciones (condiciones) más significativas. Similaridades y afinidades entre condiciones. Estructura geométrica de los datos.
Mapas de Características	Aportan sentido físico a los mapas. Correlaciones entre variables. Influencia de variables sobre condiciones. Monitorización de la evolución de las variables.
Error de Modelado (módulo)	Cambios en la dinámica del proceso. Aparición de condiciones nuevas. Desviaciones respecto del modelo.
Error de Modelado (vectorial)	Variable que más se desvía del modelo (más sospechosa de la desviación).

Tabla 5.1: Representaciones del KR-SOM e información que aportan.

# Capítulo 6

## Resultados Experimentales

### 6.1 Caso 1: Supervisión del estado de un motor de inducción. Prototipo de Laboratorio.

#### 6.1.1 Descripción del Banco de Pruebas

El primer banco de pruebas para la aplicación de la metodología de supervisión y monitorización de procesos basada en KR-SOM consiste en un motor de inducción con las siguientes características:

Vel. nominal	1500 rpm
Num. pares de polos	$p = 2$
Conexión	Y
Tensión de alimentación	380 V
Potencia	4 KW

En orden a verificar experimentalmente la metodología propuesta se sometió al motor a dos tipos de fallos.

**Fallo 1:** El primer fallo consiste en provocar una asimetría en la alimentación mediante supresión de una fase. En tales condiciones, el motor asíncrono es capaz de girar, aunque incrementa sus niveles de vibración y decrece su rendimiento.

**Fallo 2:** El segundo fallo provocado consistió en la colocación de una pequeña masa asimétrica en el rotor que genera un desequilibrio mecánico y, por tanto, vibraciones que pueden ser captadas por los acelerómetros.

Las posibles combinaciones de ambos fallos dan lugar a cuatro modos de funcionamiento:

- a) Sin fallos
- b) Asimetría eléctrica (fallo 1).
- c) Asimetría mecánica (fallo 2).
- d) Asimetrías eléctrica y mecánica simultáneamente (fallos 1 y 2).

En el banco de pruebas se instalaron tres acelerómetros: dos en direcciones horizontal y vertical, denominados  $a_X$  y  $a_Y$ , y uno situado sobre los cojinetes en dirección perpendicular al plano  $x - y$ , denominado  $a_Z$ . En las figuras 6.1 y 6.2 se muestran, respectivamente, una foto y el esquema completo del banco de pruebas:

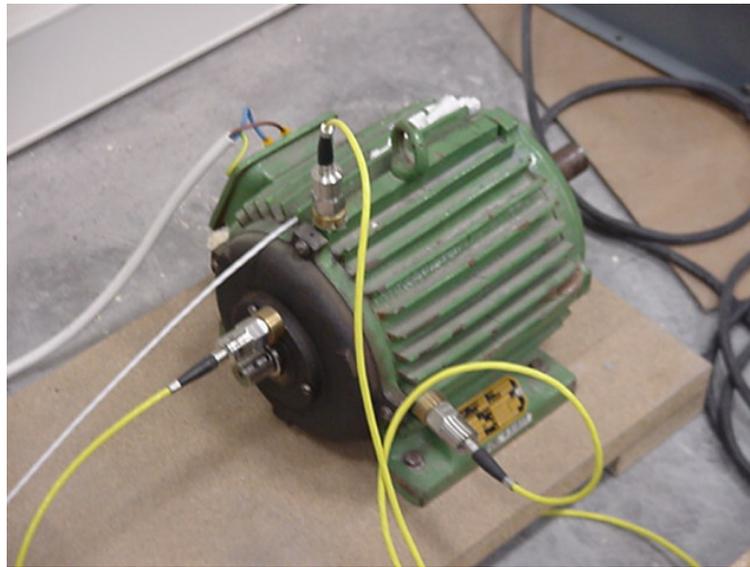


Figura 6.1: Foto del banco de pruebas

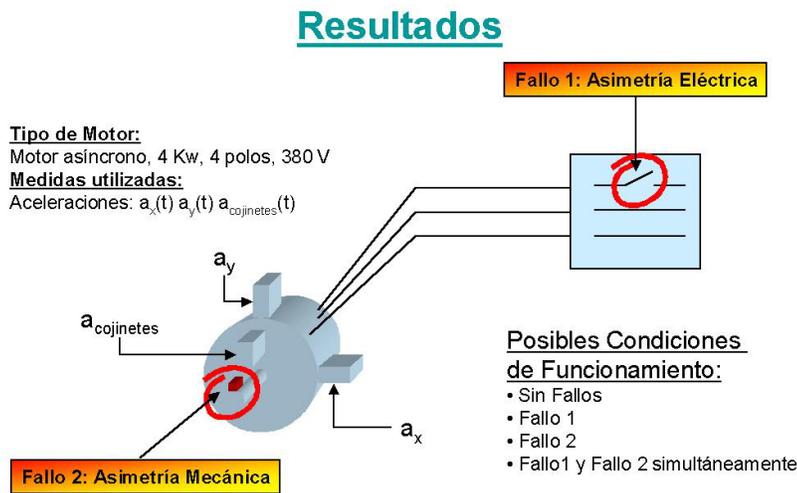


Figura 6.2: Esquema del experimento.

### 6.1.2 Extracción de Características

Cuando un motor de inducción es alimentado utilizando sólo dos fases, su nivel de vibración se incrementa a ciertas frecuencias características (en especial a 100 Hz). Análogamente, la presencia de una masa asimétrica, produce modificaciones en los niveles de vibración a frecuencias 1 y 2 veces la frecuencia de giro mecánica del motor. Si ambos fenómenos se dan simultáneamente, en general, interactúan y producen sonido y vibraciones a ciertos armónicos característicos. Por lo general, la interacción no



Figura 6.3: Diagrama de bloques de la etapa de extracción de características en el Caso 1.

suele ajustarse al principio de superposición, es decir, la interacción suele ser no lineal. En cualquier caso, no obstante, ambos fenómenos, ya sea por separado o de forma simultánea, son bien descritos por frecuencias en el rango comprendido entre cero y unos pocos cientos de hercios.

Para la extracción de características se muestrearon los datos de cada uno de los canales (aceleraciones  $a_X(t)$ ,  $a_Y(t)$  y  $a_Z(t)$ ) a 40 KHz. Los datos temporales fueron segmentados en ventanas de 4096 elementos con un solapamiento del 95 % para conseguir una buena definición en el tiempo, empleándose una ventana de Hanning para reducir el derrame espectral. A cada ventana se le realizó una FFT. En la figura 6.3 se muestra el diagrama de bloques correspondiente a la extracción de características realizada.

Las características obtenidas mediante la técnica anterior, pueden representarse en un diagrama tiempo-frecuencia denominado *espectrograma*, que refleja la evolución del espectro de potencias a lo largo del tiempo. En la figura 6.4 se muestra el espectrograma correspondiente al acelerómetro  $a_X$  para las cuatro condiciones a las que se vio sometida la máquina: normal, asimetría eléctrica (“fallo 1”), asimetría mecánica (“fa-

llo 2<sup>o</sup>) y asimetrías eléctrica y mecánica simultáneamente. Como puede apreciarse, los cambios más significativos se producen en la zona de bajas frecuencias. Como vectores de características se eligieron las energías de los 11 primeros armónicos<sup>1</sup> de la FFT de cada ventana, para cada uno de los tres canales. Estos armónicos comprenden el rango de frecuencias de 9.76 a 107.4 Hz, dando lugar a vectores de 33 características:

$$\mathbf{f}(k) = \underbrace{[X_1(k), \dots, X_{11}(k)]}_{a_X}, \underbrace{[Y_1(k), \dots, Y_{11}(k)]}_{a_Y}, \underbrace{[Z_1(k), \dots, Z_{11}(k)]}_{a_Z}]^T \quad (6.1)$$

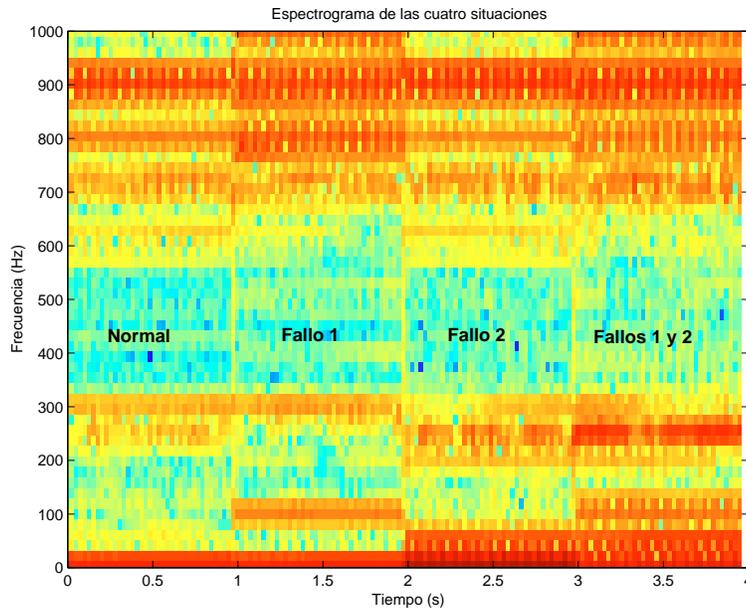


Figura 6.4: Espectrograma (diagrama tiempo-frecuencia) del acelerómetro  $a_X$  a lo largo de las cuatro condiciones a las que se sometió al motor: estado *normal*, estado de *asimetría eléctrica*, estado de *asimetría mecánica* y estado de *asimetrías eléctrica y mecánica combinadas*.

### 6.1.3 Entrenamiento del Modelo

El entrenamiento del SOM fue realizado utilizando vecindad gaussiana

$$h_{ci} = e^{-\frac{d_{g(c,i)}^2}{\sigma^2}} \quad (6.2)$$

<sup>1</sup>No se tomó la componente de continua, al tratarse de aceleraciones.

siendo  $d_g$  una distancia de tipo enlace<sup>2</sup> definida sobre una retícula de  $15 \times 15$  con topología rectangular, como la que se muestra en la figura 6.5.

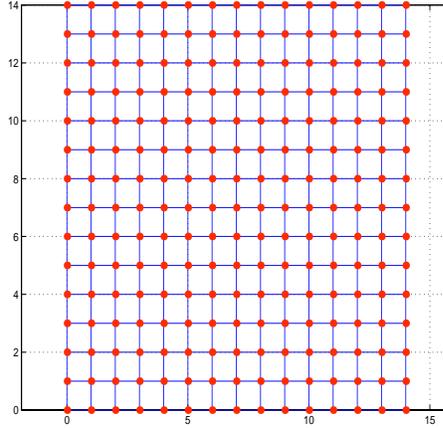


Figura 6.5: Retícula  $15 \times 15$  utilizada en el experimento del Caso 1.

Los valores de  $\sigma$  (parámetro directamente relacionado con la amplitud del campo receptivo), fueron decrementados monótonamente durante el entrenamiento desde aproximadamente un tercio del tamaño total de la red ( $\sigma = 5$ ), lo que permite una perfecta autoorganización, hasta un valor final de  $\sigma = 0.5$  que permite afinar las posiciones de las neuronas ya autoorganizadas. Los datos de entrenamiento consistieron en ejemplos de vectores de características  $\mathbf{f}(k)$  tomados de 4 registros de 1 segundo de duración correspondientes a las cuatro situaciones de fallo descritas anteriormente. Todos los ejemplos disponibles fueron utilizados 20 veces (épocas) durante el entrenamiento.

### 6.1.4 Generación de mapas

Tras la convergencia del SOM se obtiene un conjunto de  $15 \times 15$  vectores de codificación  $\mathbf{m}_i$  que permiten trazar ya los mapas. Aplicando las expresiones (5.2), (5.5) y (5.7), es posible trazar los mapas continuos de

<sup>2</sup>La distancia de tipo enlace entre dos nodos de una retícula se define como el número de enlaces que es necesario atravesar para ir de un nodo al otro por el camino más corto.

características, activaciones y distancias interneuronales, respectivamente. Para trazarlos en una figura se han evaluado dichas expresiones para valores de  $u$  correspondientes a una sub-retícula fina de  $50 \times 50$  de la retícula original.

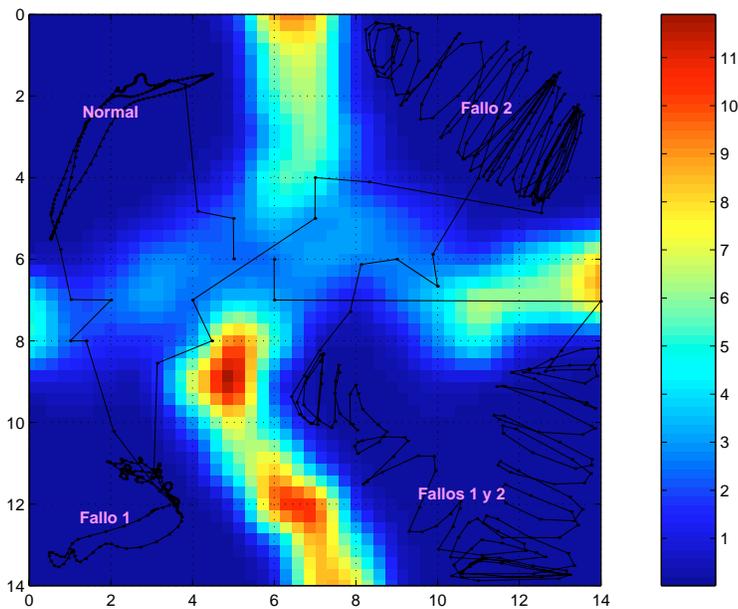


Figura 6.6: Mapa de Distancias continuo del SOM entrenado con las cuatro situaciones del Caso 1.

**Mapa de Distancias:** En la figura 6.6 se muestra el mapa de distancias continuo al que dan lugar los ejemplos analizados del motor, junto con la trayectoria descrita por éste al pasar por las cuatro situaciones. La matriz de distancias aporta información de manera visual sobre las agrupaciones de datos en el espacio de entrada, que en este caso corresponden a distintos estados o condiciones del proceso. Pueden comprobarse, perfectamente delimitadas, las cuatro zonas que distingue el SOM. Las zonas marcadas en azul representan neuronas muy próximas entre sí, correspondiendo cada zona a cada una de las condiciones de funcionamiento del motor consideradas en los ejemplos.

**Mapas de Características:** Los mapas de características aportan información de enorme utilidad sobre el proceso. Su principal ventaja

radica en el hecho de que las características, casi siempre (dependiendo de la etapa de extracción de características), tienen un significado físico, lo que permite al experto en el proceso o al operario extraer conclusiones sobre el proceso de manera casi directa. En la figura 6.7, se muestra el mapa de distancias interneuronales junto con los mapas de los niveles de vibración<sup>3</sup> a 29.3 Hz, 48.8 Hz y 97.6 Hz en el acelerómetro  $a_X$  del motor.

Un simple vistazo a los mapas aporta una gran cantidad de información con significado físico sobre el proceso. En efecto, puede comprobarse fácilmente que el mapa del armónico de 29.3 Hz, el más próximo en la FFT al de 25 Hz, separa claramente dos zonas:

- a) La zona azul, correspondiente a bajos niveles de vibración en la banda centrada en 29.3 Hz. Esta zona abarca, como se ve en la figura, el funcionamiento normal del motor y el fallo 1 (asimetría eléctrica), es decir los estados de funcionamiento en los que no existe asimetría mecánica.
- b) La zona roja, relativa a elevados niveles de vibración en dicha banda, que comprende los estados que presentan asimetría mecánica.

Esta separación tiene un evidente sentido físico. Las asimetrías mecánicas, producen armónicos de vibración esencialmente a múltiplos  $1x$  y  $2x$  de la velocidad de giro del motor ( $1x = 1500 \text{ rpm} = 25 \text{ Hz}$ ), por lo que, como era de esperar, la banda de 29.3 Hz es altamente significativa frente a este tipo de asimetrías.

El mapa del armónico de 97.6 Hz –próximo a los 100 Hz–, por el contrario, es altamente significativo en cuanto a las asimetrías eléctricas (“fallo 1”). Efectivamente, la ausencia de una fase, induce una vibración de origen electromagnético sobre el motor a dos veces la frecuencia de alimentación, es decir, a 100 Hz. En ausencia de asimetría eléctrica, este armónico se reduce. La situación puede apreciarse nítidamente en el mapa inferior derecho de la figura 6.7.

El mapa revela, no obstante, información más sutil sobre el proceso. Como puede observarse en el mapa de 97.6 Hz, las zonas “fallo 2” y “fallos 1 y 2” revelan un ligero corrimiento hacia tonos rojos respecto a las zonas

---

<sup>3</sup>Las frecuencias corresponden a los armónicos tercero, quinto y décimo de una FFT de 4096 elementos sobre secuencias muestreadas a 40000 Hz, siendo éstas múltiplos de  $40000/4096=9.7656 \text{ Hz}$

“fallo 1” y “normal”. En efecto, la presencia de una asimetría mecánica (“fallo 2”) incrementa ligeramente los niveles de vibración respecto a una misma situación eléctrica ya que el armónico de 100 Hz corresponde al múltiplo 4x de la velocidad de giro que, aunque menos que los 1x y 2x, también se ve afectado por la asimetría mecánica.

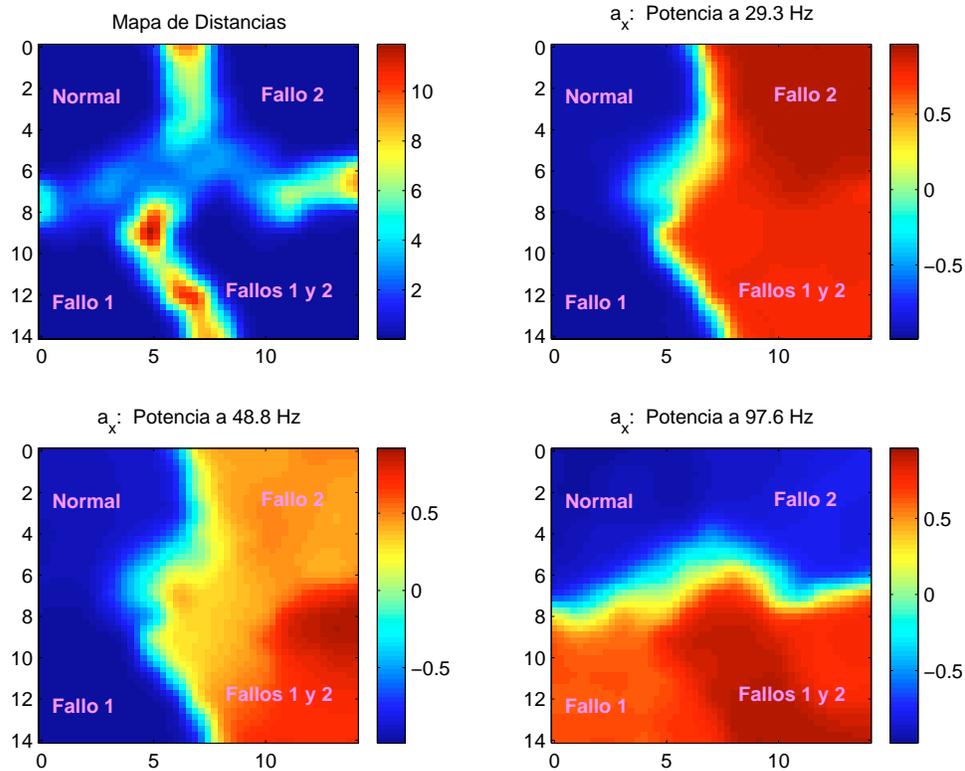


Figura 6.7: De izquierda a derecha y de arriba a abajo, mapa continuo de distancias y mapas continuos de niveles de vibración a 29.3 Hz, 48.8 Hz, y 97.6 Hz.

**Mapas de Activaciones:** Los mapas de activaciones constituyen una potente herramienta en el análisis del proceso. Como se dijo en el apartado 5.4 este tipo de mapas permite correlar visualmente los datos generados por el proceso en cada momento con estados previos conocidos pertenecientes al conjunto de datos de entrenamiento. En la figura 6.8 se muestran los mapas de activación continuos generados por los estímulos

correspondientes a cada una de las cuatro situaciones del motor. Como cabe esperar, las regiones estimuladas se corresponden con las cuatro regiones delimitadas por el mapa de distancias y evidenciadas por los mapas de características.

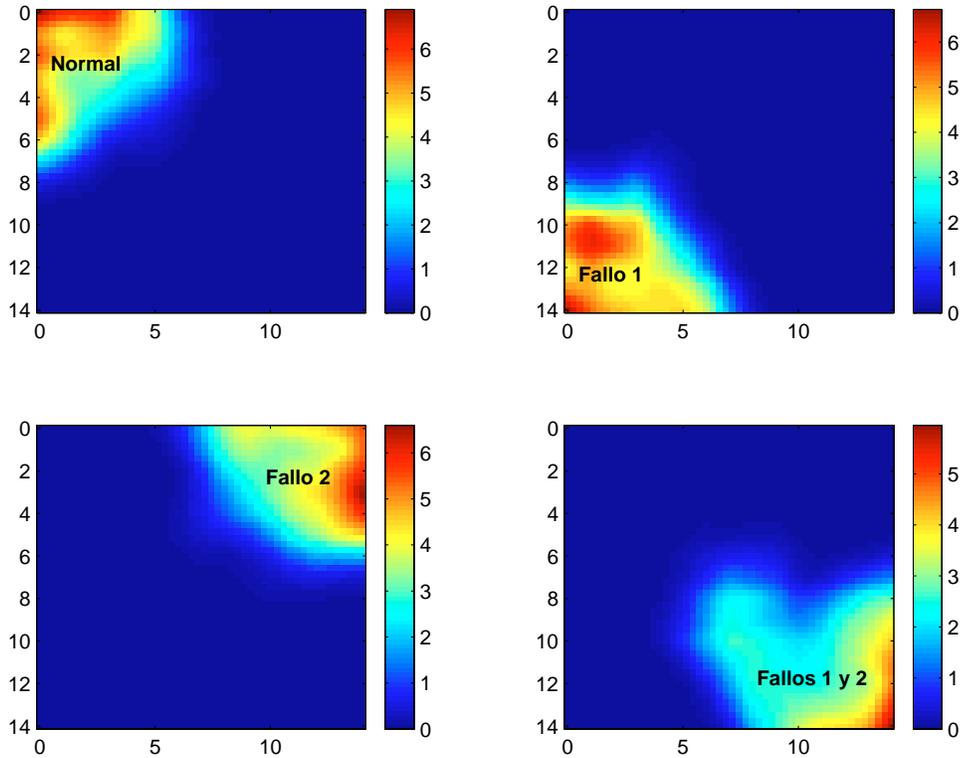


Figura 6.8: Mapas de Activación del Caso 1. De izquierda a derecha y de arriba a abajo, mapas de activación correspondientes a las situaciones “normal”, “fallo 1”, “fallo 2” y “fallos 1 y 2”.

**Error de Modelado:** El error de modelado permite detectar *on-line* situaciones no previstas en los ejemplos de entrenamiento. Este parámetro, como se indicó en el apartado 5.6, representa el nivel de discrepancia entre el modelo que el KR-SOM construye sobre el proceso y la situación real. Cuando los datos que genera el proceso se adaptan al modelo aprendido por el SOM, el error es pequeño; cuando, por el contrario, el proceso atraviesa una situación distinta a las consideradas en los datos de entre-

namiento, el modelo es incapaz de generar aproximaciones correctas a los datos de proceso y, en consecuencia, el error aumenta.

En la figura 6.9 se muestra la aplicación del error de modelado para detectar una situación posiblemente anómala. Se registraron datos del motor durante los minutos 1 y 11 de funcionamiento. Al cabo de 11 minutos de funcionamiento, el motor se calienta ligeramente, modificando las propiedades de los materiales –dilataciones, cambios de viscosidad en el aceite, etc.–, lo que ocasiona modos de vibración que difieren ligeramente de los observados durante el entrenamiento. Evidentemente, esto no tiene por qué representar necesariamente un fallo o anomalía en el proceso aunque, en cualquier caso, sí constituye una desviación respecto del modelo construido sobre la base de los ejemplos. En dicha figura se muestra el error de modelado correspondiente a los datos de entrenamiento (cuatro primeros tramos) así como el correspondiente a los datos de test (dos últimos tramos, correspondientes a los minutos 1 y 11 de funcionamiento). En la figura puede apreciarse cómo el error de modelado global hace patente esta desviación. Por otra parte, el error de modelado vectorial revela las variables que más contribuyen al error de modelado, en este caso el armónico de 58.6 Hz del acelerómetro  $a_Y$  y el de 68.4 Hz en el acelerómetro  $a_Z$ , apreciándose también una ligera disminución –tonos azules– en la energía del armónico de 68.4 Hz en el acelerómetro  $a_X$ .

En la figura 6.10 pueden apreciarse las proyecciones de las trayectorias correspondientes a 1 minuto y a 11 minutos trazadas sobre los mapas de distancias y de las características correspondientes a los niveles de vibración a 48.8 y 97.6 Hz en el acelerómetro  $a_X$ . En estas últimas, se puede apreciar que ambas describen trayectorias distintas, aunque permaneciendo en la región etiquetada como “normal”.

### **Proyección 3D de la trayectoria y la hipersuperficie del SOM:**

Finalmente, se muestra la proyección 3D de la trayectoria junto con la proyección 3D de la hipersuperficie que define el SOM en el espacio de características, obtenidas mediante la combinación de técnicas MDS y regresión de núcleo propuesta en la sección 5.7. En pseudocolor, se ha trazado el mapa de distancias interneuronales. En el dibujo pueden apreciarse cuatro zonas teñidas de azul oscuro, las más frecuentadas por la trayectoria, que corresponden a distancias interneuronales pequeñas, es

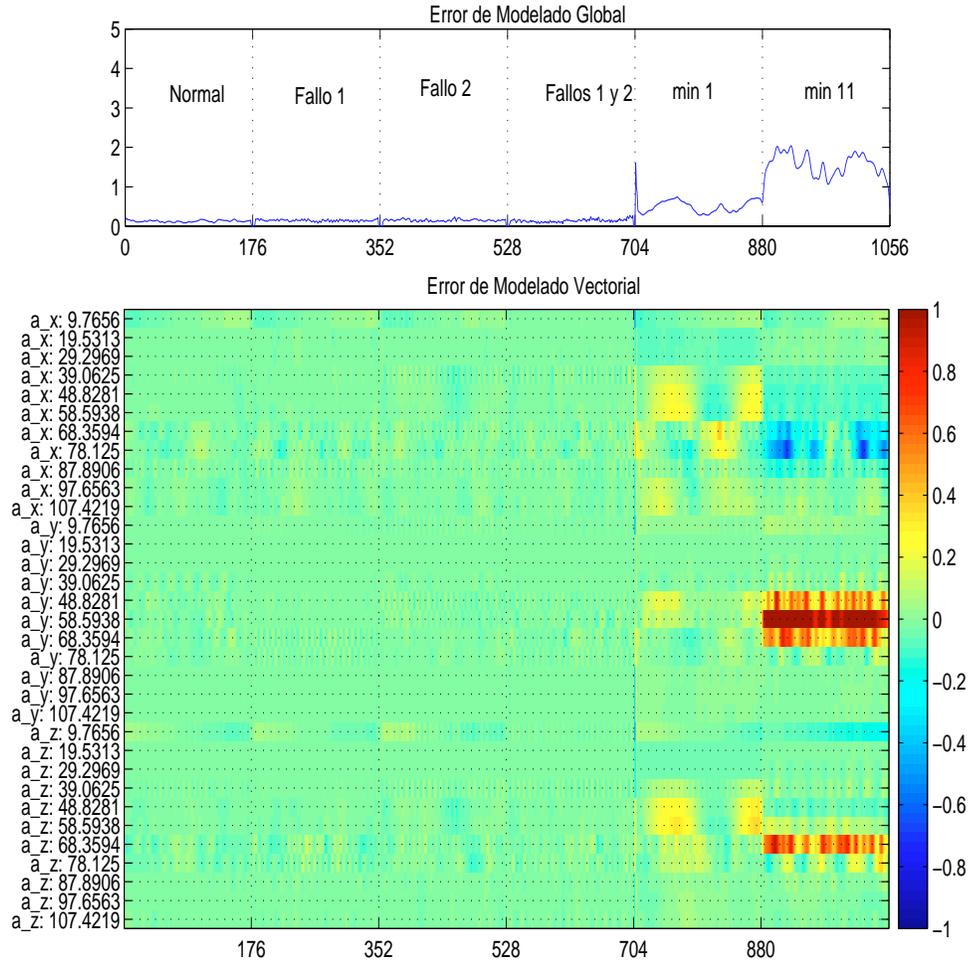


Figura 6.9: Error de Modelado. Caso 1. Evolución de los errores de modelado global y vectorial. Los 4 primeros tramos corresponden a datos de entrenamiento con las 4 situaciones posibles. Los dos tramos restantes son datos de test, sin fallos provocados, tomados al cabo de 1 y 11 minutos funcionando.

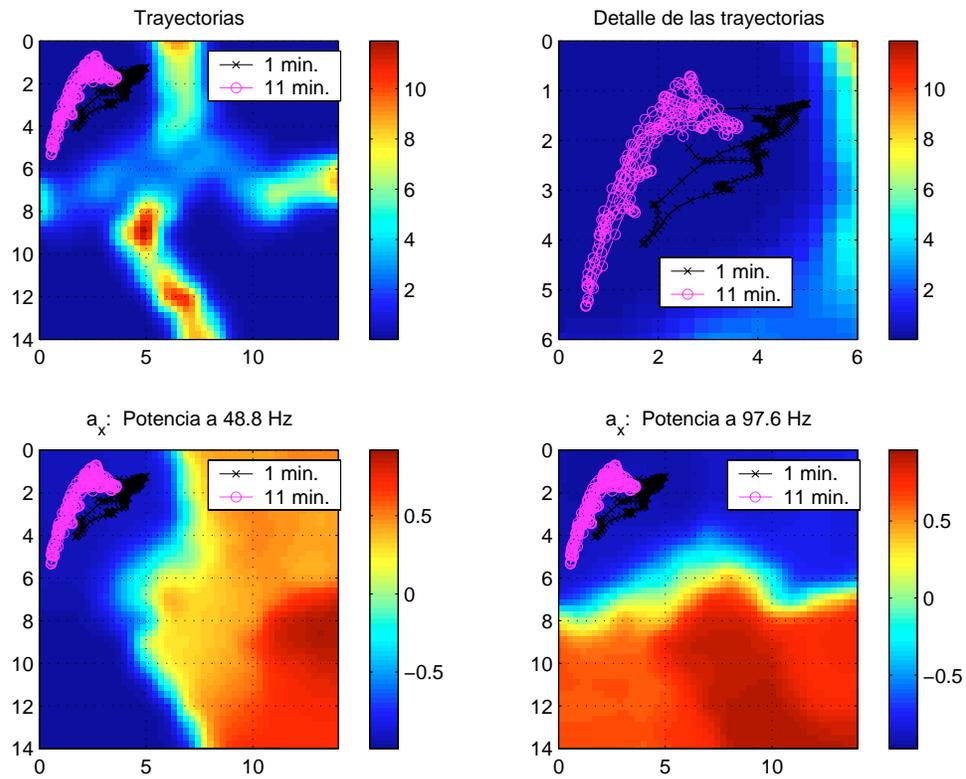


Figura 6.10: Trayectorias proyectadas sobre el mapa de distancias y sobre los mapas de los armónicos 48.8 Hz y 97.6 Hz, junto con un detalle de la trayectoria.

decir, clusters de neuronas que representan las cuatro situaciones del motor. La trayectoria discurre entre las cuatro zonas, abandonándolas esporádicamente en los tránsitos entre dos situaciones, al igual que ocurría en los mapas 2D. Esta representación, como puede comprobarse, no resulta tan práctica como las representaciones 2D aunque, a nuestro juicio, puede llegar a tener un enorme potencial en combinación con técnicas de representación 3D y realidad virtual.

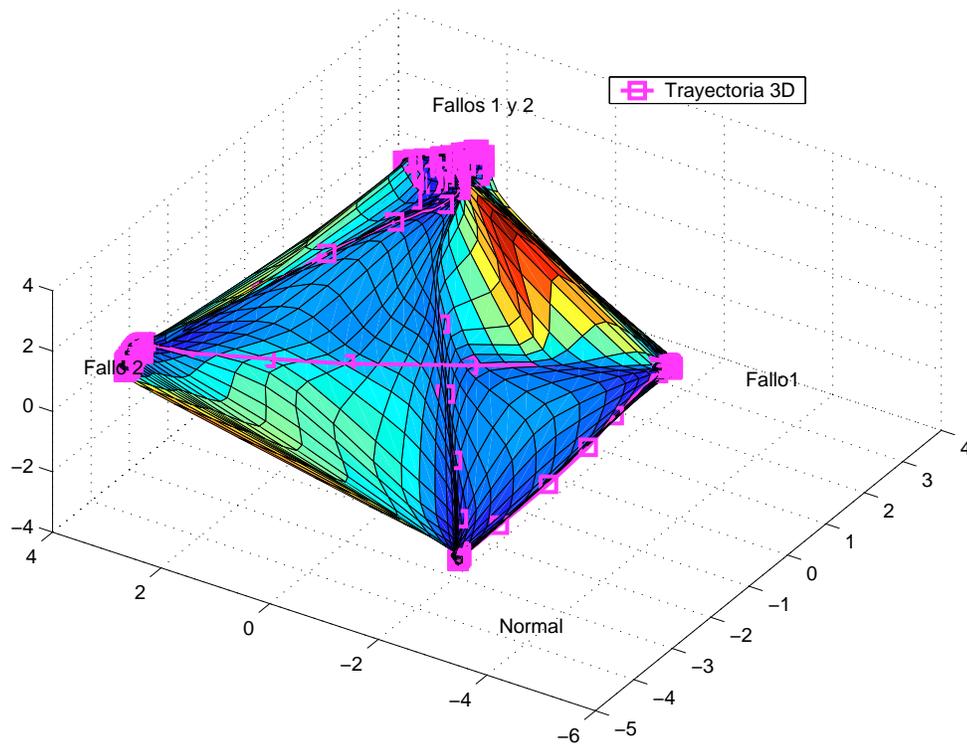


Figura 6.11:

## 6.2 Caso 2: Supervisión de un motor de CC de 7000 Kw en un Tren de Laminación Semicontinuo

El método propuesto fue aplicado sobre datos reales procedentes de uno de los motores de corriente continua de 7000 kW situado en la parte acabadora del tren de laminación de bandas en caliente que Aceralia tiene en su factoría de Avilés, Asturias. Se disponía de 4 señales del motor: corriente de inducido  $i_{ind}$ , corriente de excitación  $i_{exc}$ , tensión de inducido  $V_{ind}$  y frecuencia de giro,  $f_g$ . La evolución de estas variables depende del tipo de chapa que se esté laminando, de los sistemas de regulación implementados en el accionador, de la velocidad a la que se lamina, de la fuerza con la que actúen los rodillos en la laminación, así como de otras variables de proceso, entre otros factores.

En la figura 6.12 se muestra la evolución de las medidas del proceso durante el laminado de tres bobinas. Como puede apreciarse, durante todo el proceso la tensión de inducido se mantiene casi constante, en torno a 600 V, trabajando el motor en debilitamiento de campo, a potencia constante. Durante cada laminación, la corriente de inducido se eleva para aportar el par necesario y la corriente de excitación es utilizada para variar la velocidad del motor mediante debilitamiento de campo, reduciéndola para incrementar la velocidad y viceversa. La velocidad, como puede apreciarse, en la figura, sigue una dinámica opuesta a la corriente de excitación.

Esta descripción, relativamente simple, ilustra el comportamiento del proceso a grandes rasgos. Sin embargo, la dinámica subyacente del proceso puede ser muy compleja cuando se examina en detalle. El motor cuenta, por ejemplo, con sistemas de potencia basados en doble puente trifásico de tiristores, tanto para el circuito de excitación como para el circuito de inducido. Estos tiristores generan una serie de armónicos en las corrientes cuya energía varía con arreglo a ciertas leyes, según la condición del proceso, lo que hace pensar que dichos armónicos son sensibles a condiciones del proceso relacionadas con el subsistema de potencia. El motor, además, está acoplado a un sistema mecánico más o menos complejo compuesto de elementos de transmisión, rodamientos, reductores, cilindros de laminación, la propia chapa, etc. que generan armónicos a frecuencias que, en general, dependen de la velocidad de giro. Estos

armónicos aportarían información sobre aspectos relacionados con el subsistema mecánico tales como las fuerzas de los cilindros sobre la chapa, defectos en el sistema de transmisión, etc.

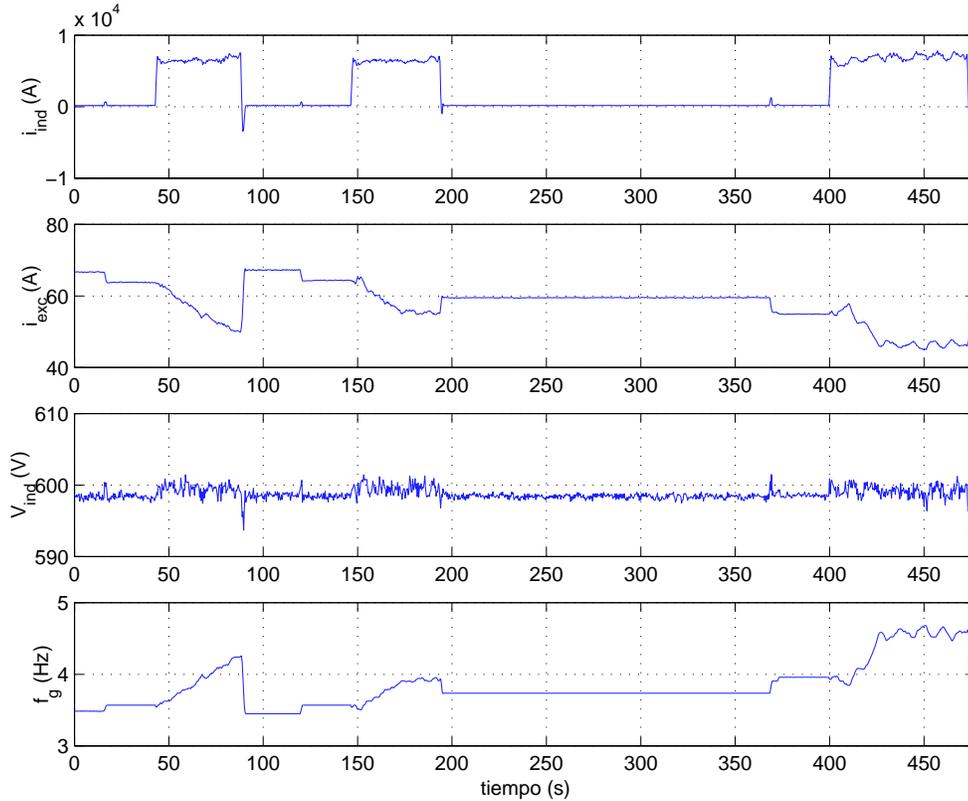


Figura 6.12: Evolución de la corriente de inducido  $i_{ind}$ , la corriente de excitación,  $i_{exc}$ , la tensión de inducido  $V_{ind}$  y la frecuencia de giro  $f_g$  en un motor de CC de 7000 kW de un tren de laminación, a lo largo de tres laminaciones.

Aunque no es el objeto de esta Tesis la descripción exhaustiva de cada uno de estos factores y su influencia sobre el estado del proceso<sup>4</sup>,

<sup>4</sup>Ello no implica que un análisis exhaustivo de la física del proceso, o incluso un modelado del mismo, sea innecesario, muy al contrario serían de enorme utilidad ya para el diseño de la extracción de características ya en combinación con la metodología propuesta, por ejemplo, utilizando los residuos del modelo como características para la proyección.

la integración de estas variables permite realizar una supervisión más completa del proceso, en el sentido de considerar un abanico más amplio de condiciones que el que proporcionan las variables fundamentales, al tiempo que sirven para ilustrar la utilidad de la metodología propuesta en la supervisión de procesos difíciles de modelar. En los términos planteados en la sección 1.4, la utilización de variables sensibles a condiciones de otras etiologías se traduce en un mayor número de regiones del espacio de estados representadas sin ambigüedad tras los procesos de medida y extracción de características. Esto permite, con técnicas de proyección adecuadas, representar un abanico de condiciones más amplio en el espacio de visualización.

### 6.2.1 Extracción de Características.

Como se indicó anteriormente, aunque se han medido tan sólo cuatro variables del proceso, es posible obtener un número muy superior de características que pueden aportar gran cantidad de información sobre la condición del proceso. En la figura 6.13 se muestra el espectrograma de  $i_{ind}(t)$ , en el que puede apreciarse la evolución del contenido espectral de la señal a lo largo del tiempo. En particular, destacan, los armónicos de 300 y 600 Hz, originados por los puentes de tiristores del equipo de potencia que alimenta al motor de continua, cuya frecuencia se mantiene constante, variando su amplitud en función de las condiciones de funcionamiento. Este armónico se encuentra también presente en las variables  $i_{exc}$  y en  $V_{ind}$ .

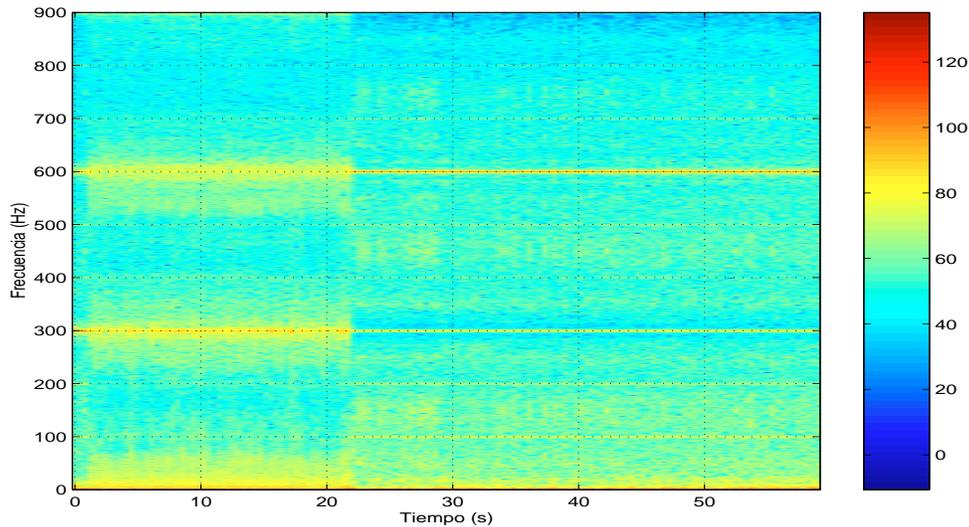


Figura 6.13: Espectrograma de  $i_{ind}(t)$ .

En la figura 6.14 se muestran en detalle las zonas de bajas frecuencias de los espectrogramas de  $i_{ind}$ ,  $i_{exc}$  y  $V_{ind}$  en las que se aprecia la presencia de dos armónicos de frecuencia variable. Comparando con la figura 6.12 se aprecia que sus frecuencias corresponden exactamente a 1 y 2 veces la frecuencia de giro  $f_g(t)$  del motor (en adelante, armónicos 1x y 2x). Estos armónicos tienen casi siempre un origen mecánico, y pueden aportar información acerca de desalineaciones, defectos mecánicos en los rodillos

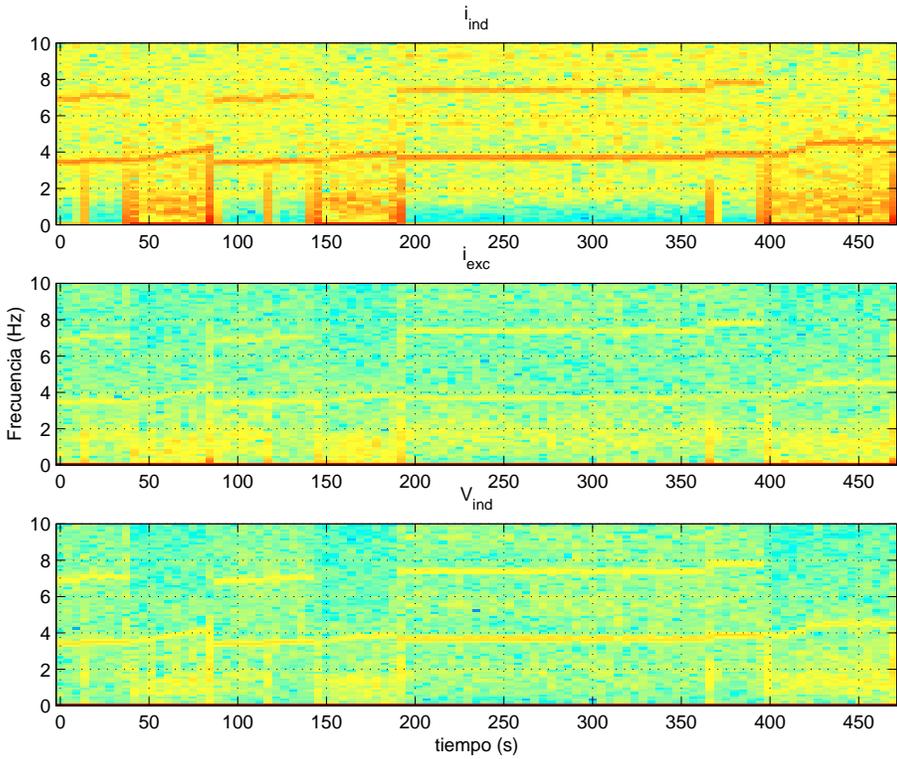


Figura 6.14: Detalles de los espectrogramas de  $i_{ind}(t)$ ,  $i_{exc}(t)$  y  $V_{ind}(t)$ .

o engranajes de la transmisión, etc. Para la supervisión del motor se eligió un vector formado por las siguientes 16 características:

$$\mathbf{f}(k) = [i_{ind}, i_{ind}^{1x}, i_{ind}^{2x}, i_{ind}^{300}, i_{ind}^{600}, i_{exc}, i_{exc}^{1x}, i_{exc}^{2x}, i_{exc}^{300}, i_{exc}^{600}, V_{ind}, V_{ind}^{1x}, V_{ind}^{2x}, V_{ind}^{300}, V_{ind}^{600}, f_g]^T$$

es decir, las magnitudes  $i_{ind}$ ,  $i_{exc}$  y  $V_{ind}$ , junto con sus respectivos armónicos 1x, 2x, 300 Hz y 600 Hz, más la frecuencia de giro  $f_g$  del motor.

Para obtener las características citadas se procedió a subdividir cada una de las variables en ventanas de 1024 datos. Dado que sólo se toman 5 armónicos de cada señal, no es necesario hallar la FFT completa de cada ventana, siendo necesario hallar sólo 5 puntos de ésta. Los armónicos 1x y 2x, sin embargo, requieren un tratamiento especial. Por ser éstos de frecuencia variable, su frecuencia no es constante a lo largo de la ventana, con lo que su energía se dispersa a lo largo de varias frecuencias. Una extracción de características adecuada, válida tanto para los armónicos fijos (300 y 600) como para los de frecuencia variable (1x y 2x) es la correlación con una exponencial compleja de frecuencia variable y conocida:

$$\theta(k) = \sum_{i=1}^k f(i)\Delta t_i = \sum_{i=1}^k f(i)T_m = \quad (6.3)$$

$$c = \frac{1}{N} \sum_{k=1}^N w(k) \cdot s(k) e^{-j \cdot 2\pi \theta(k)} \quad (6.4)$$

donde,  $N$  es el número de elementos de la ventana,  $T_m$  es el periodo de muestreo,  $\theta(k)$  es el ángulo recorrido hasta la muestra  $k$ -ésima por la exponencial compleja de frecuencia  $f(k)$  conocida, y  $s(k)$  es la señal a analizar. Finalmente,  $w(k)$  es una ventana de carácter opcional que permite reducir el fenómeno de dispersión espectral de Gibbs. Esta transformación permite hacer un seguimiento o *tracking* de las energías de los armónicos de frecuencias variables pero conocidas. Para los armónicos de 300 y 600 Hz, basta hacer

$$\begin{aligned} f_{300}(k) &= 300 \\ f_{600}(k) &= 600 \end{aligned}$$

mientras que en los armónicos 1x y 2x se tomará

$$\begin{aligned}f_{1x}(k) &= f_g(t_k) \\f_{2x}(k) &= 2 \cdot f_g(t_k)\end{aligned}$$

donde  $t_k$  es el instante de la muestra  $k$ . Una vez determinado el vector de características, se optó por acondicionar los datos mediante filtrado digital para eliminar ruido y *outliers*. Posteriormente, es recomendable realizar una normalización de los datos para equiparar los niveles de variación de cada una de las características. En este caso se optó por una normalización a  $[-1, +1]$ .

**Conjuntos de datos del experimento.** Se aplicaron las técnicas descritas a dos conjuntos de datos:

- a) Un primer conjunto de *datos de entrenamiento*, utilizado para generar el modelo.
- b) y un segundo conjunto de *datos de test* que servirá para ilustrar la aplicación de la metodología propuesta ante situaciones nuevas no consideradas en el entrenamiento.

En las figuras 6.15, 6.16 y 6.17, se muestra la evolución a lo largo del tiempo de las características obtenidas para el conjunto de datos de entrenamiento. En las figuras 6.18, 6.19 y 6.20, se muestra la evolución a lo largo del tiempo de las características obtenidas para el conjunto de datos de test.

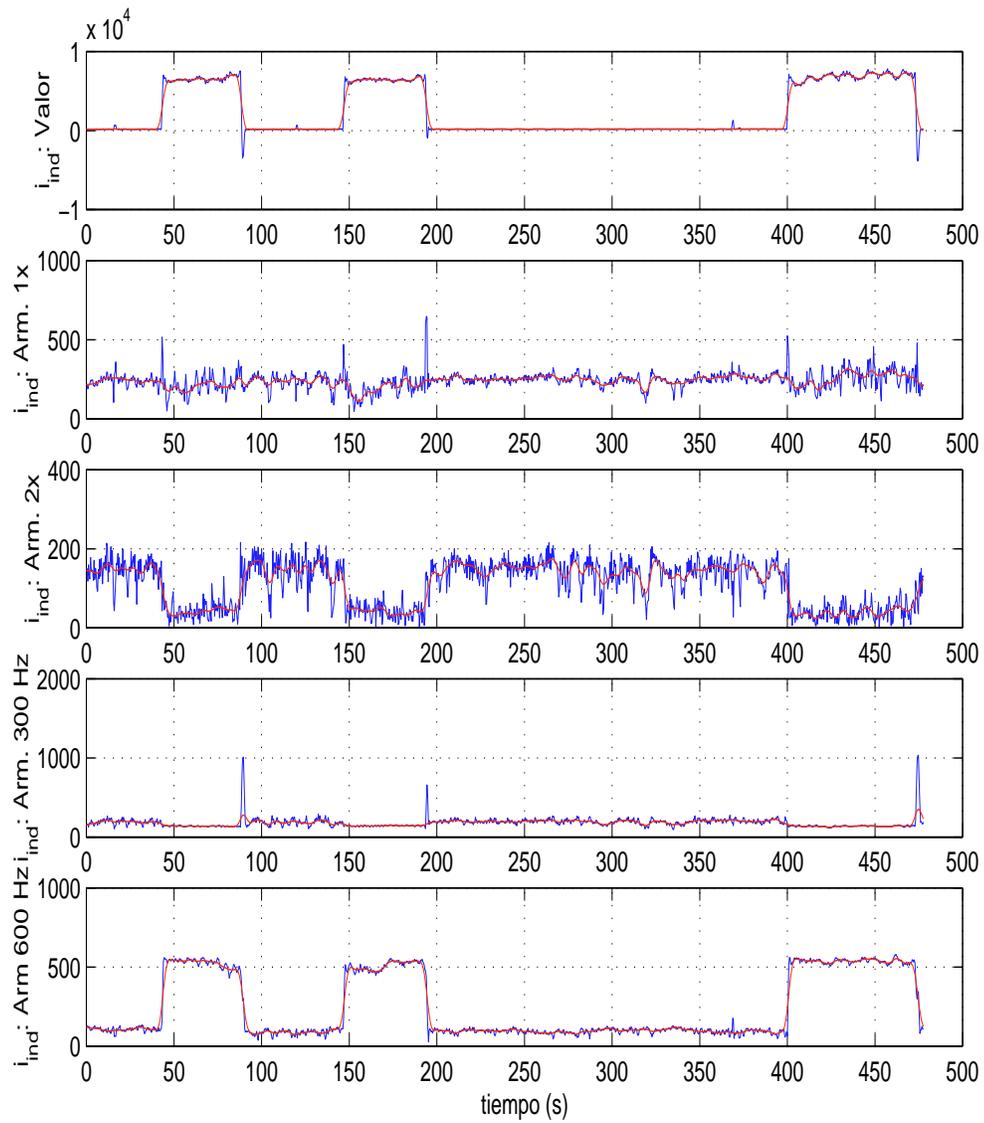


Figura 6.15: Evolución temporal de las características de  $i_{ind}$  (datos de entrenamiento).

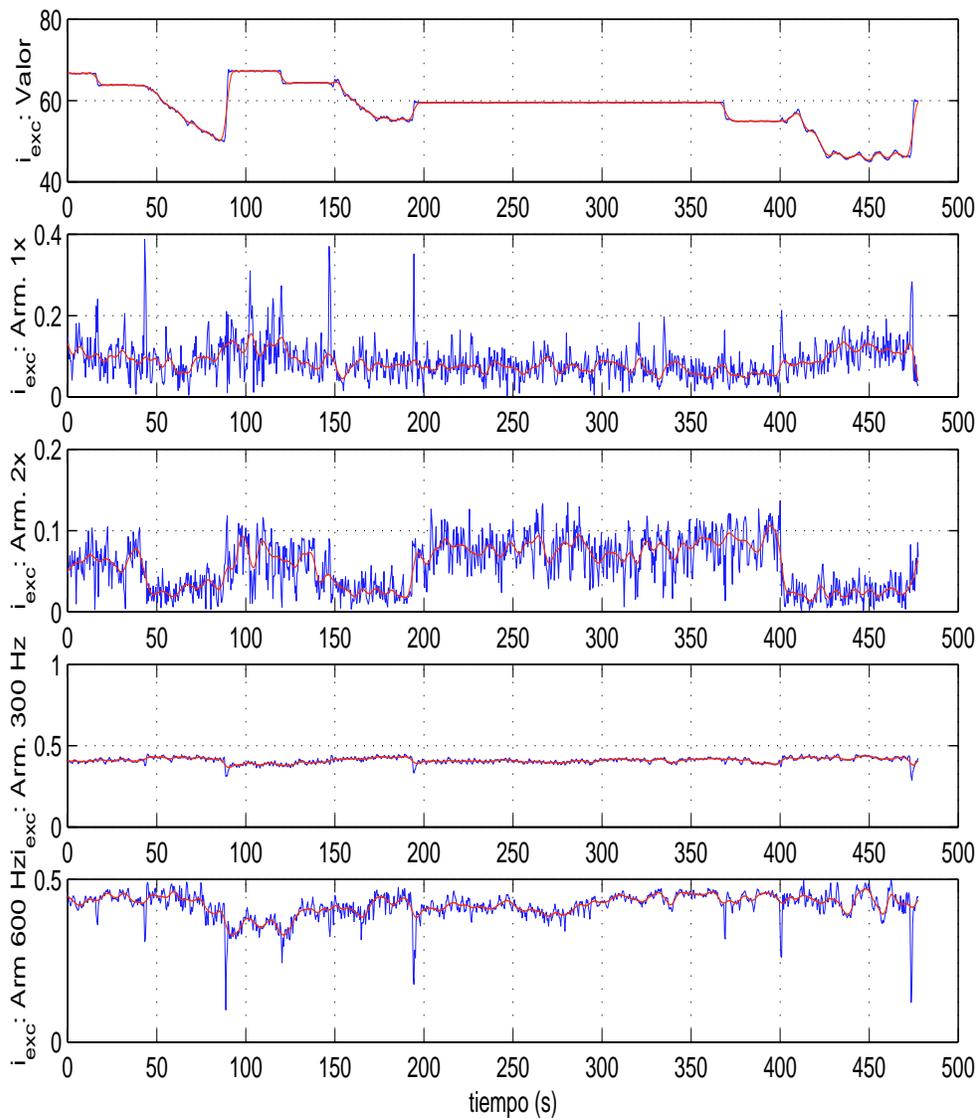


Figura 6.16: Evolución temporal de las características de  $i_{exc}$  (datos de entrenamiento).

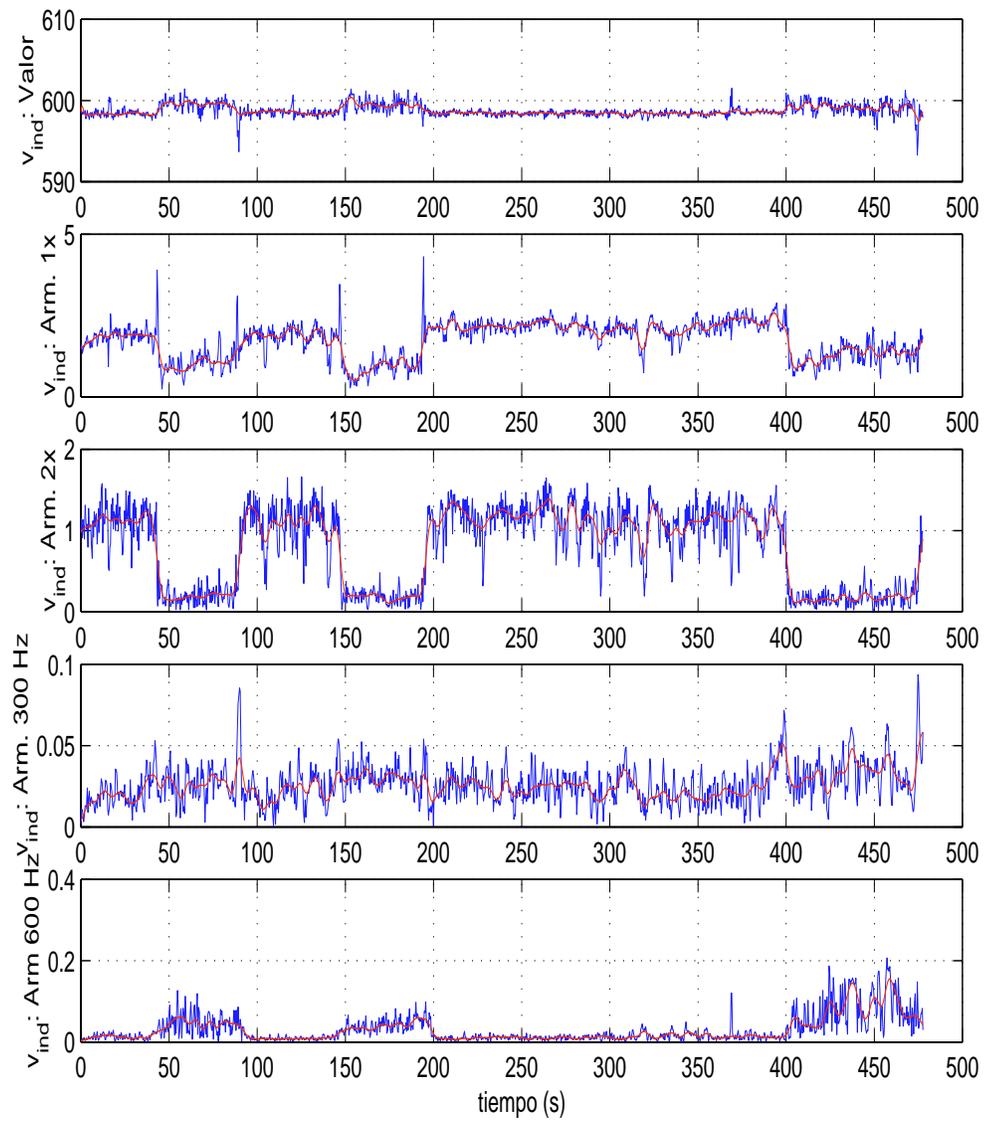


Figura 6.17: Evolución temporal de las características de  $V_{ind}$  (datos de entrenamiento).

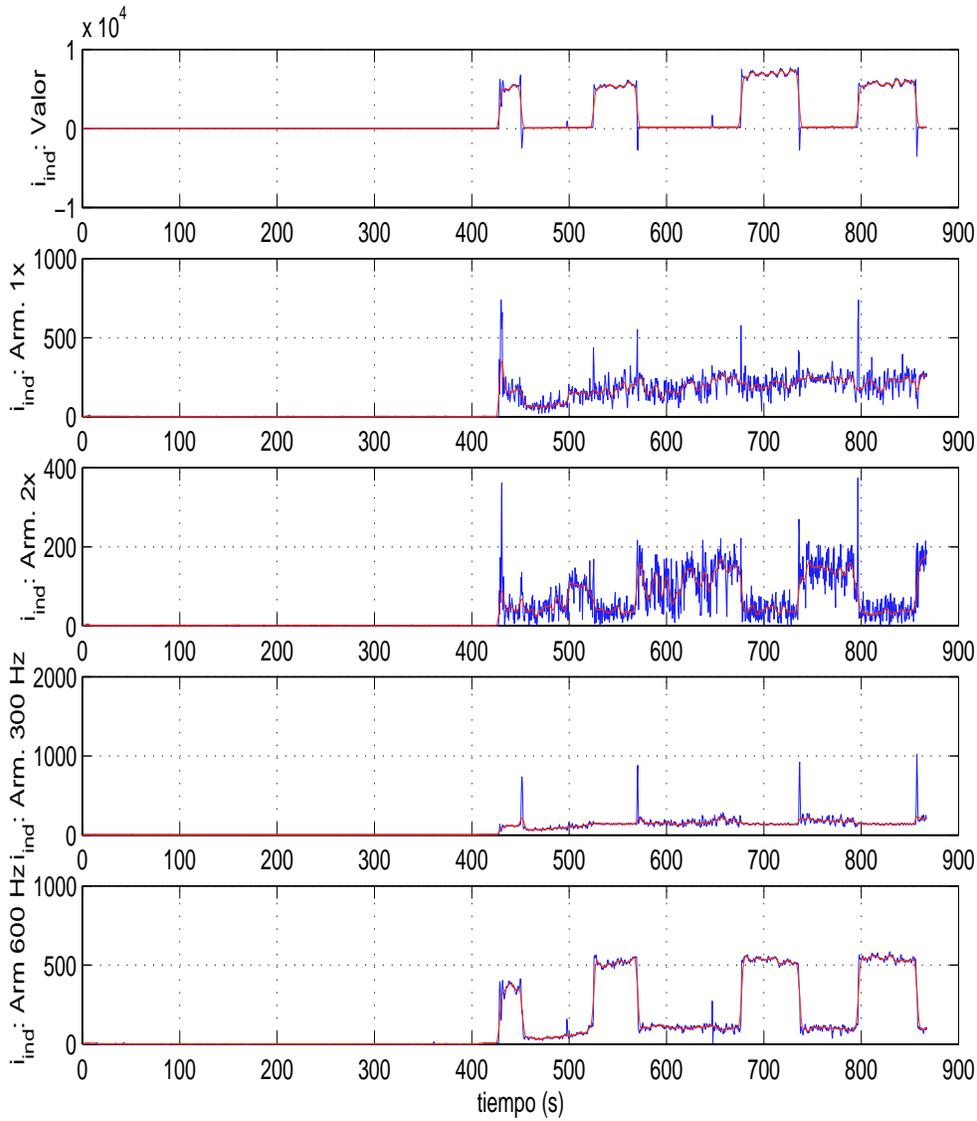


Figura 6.18: Evolución temporal de las características de  $i_{ind}$  (datos de test).

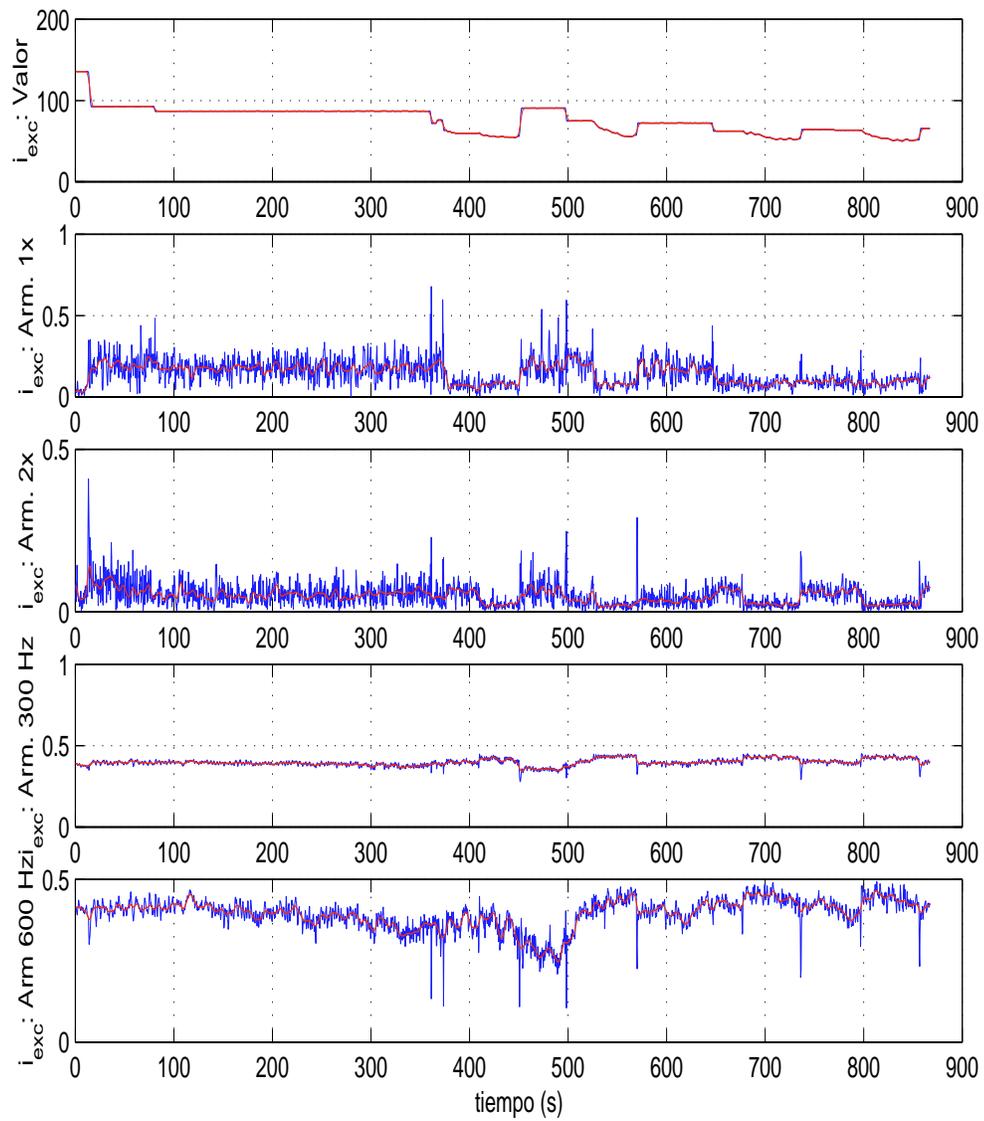


Figura 6.19: Evolución temporal de las características de  $i_{exc}$  (datos de test).

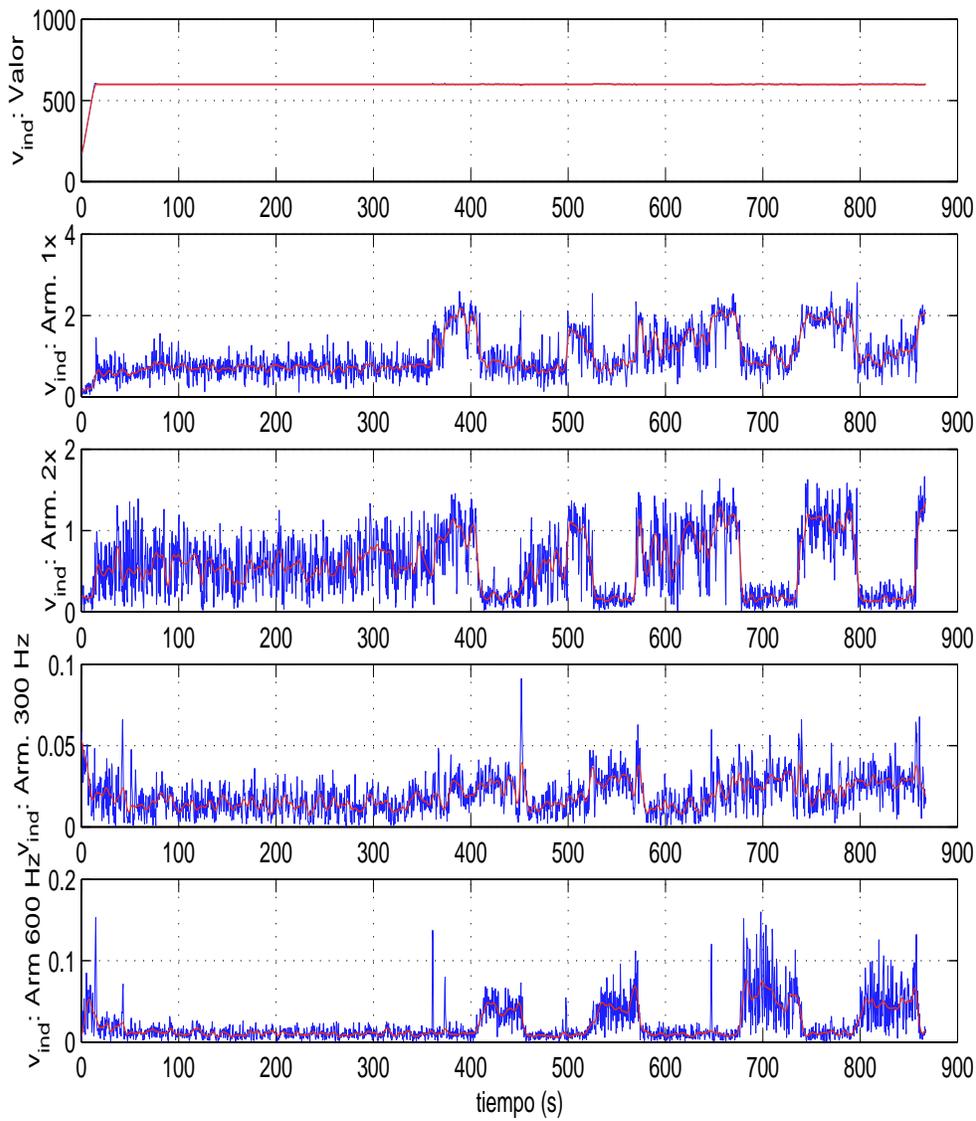


Figura 6.20: Evolución temporal de las características de  $V_{ind}$  (datos de test).

### 6.2.2 Entrenamiento del Modelo

El entrenamiento del SOM, al igual que en el Caso 1, fue realizado utilizando vecindad gaussiana:

$$h_{ci} = e^{-\frac{d_g(c,i)^2}{\sigma^2}} \quad (6.5)$$

siendo  $d_g$  una distancia de tipo enlace definida sobre la retícula. Dada complejidad del proceso, cabe suponer que la geometría de los datos en el espacio de características es compleja requiriendo una definición más precisa. Por ello, el tamaño elegido para la retícula fue de  $20 \times 20$  neuronas, con topología rectangular.

Los valores del campo receptivo del SOM,  $\sigma(t)$ , al igual que el factor de aprendizaje  $\alpha(t)$  fueron decrementados monótonamente –con una ley de decrecimiento logarítmica– durante el entrenamiento, desde aproximadamente un tercio del tamaño total de la red ( $\sigma = 6.33$ ) hasta un valor final de  $\sigma = 0.40$ . El número de épocas en el entrenamiento fue de 50.

Para la interpolación se emplearon núcleos esféricos y sus factores de ancho  $\{\sigma_i\}_{i=1,\dots,K}$  se eligieron todos iguales y aproximadamente iguales a la distancia promedio entre neuronas vecinas.

### 6.2.3 Interpretación de los resultados

**Mapa de Distancias.** Tras el entrenamiento del modelo se obtuvo el mapa de distancias que se muestra en la figura adjunta. Con el objeto de facilitar la visualización de las regiones, se muestra también el mismo mapa representando las distancias en escala logarítmica y con líneas de nivel.

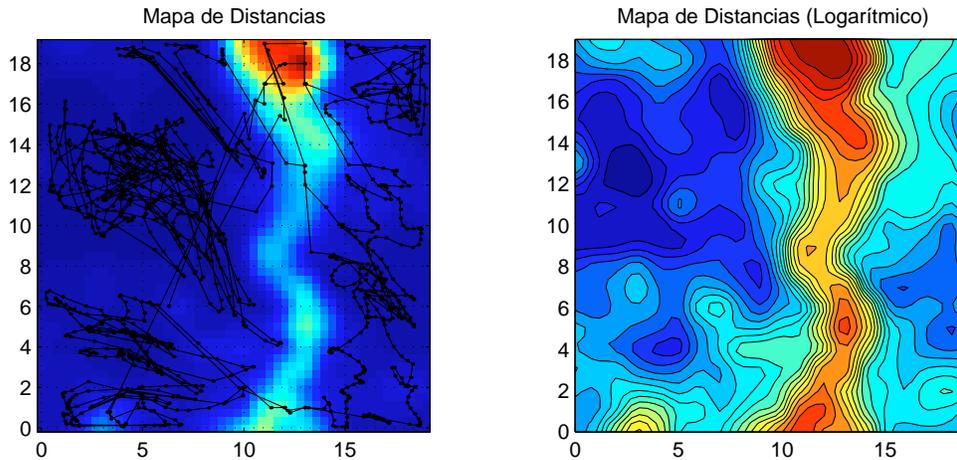


Figura 6.21: Mapas de distancias interneuronales. A la izquierda, mapa de distancias junto con la trayectoria desarrollada por los datos de entrenamiento. A la derecha, el mismo mapa en escala logarítmica y con líneas de nivel.

A primera vista, es fácil deducir la presencia de dos grandes regiones –una, más grande, a la izquierda y otra a la derecha–, claramente delimitadas por una franja de distancias interneuronales elevadas –franja rojiza– que representan dos macro-condiciones, correspondientes al motor laminando y al motor en vacío. Dado que el tamaño de las regiones es aproximadamente proporcional al número de ejemplos –ver sección 3.8–, puede deducirse que la región izquierda se corresponde con el motor en vacío, ya que la mayoría de los ejemplos en el conjunto de entrenamiento corresponden a este estado, mientras que la región de la derecha, corresponde al motor durante los periodos de laminación.

**Mapas de Características.** En el mapa de distancias anterior es posible diferenciar otras subregiones dentro de las primeras, que delimitan condiciones con diferencias más sutiles. En este sentido, los mapas de características aportan una información de enorme utilidad en la interpretación de las distintas regiones del mapa, ya que las propias características a menudo tienen un claro significado físico que permite aplicar el –posiblemente enorme– conocimiento apriorístico del técnico experto en el proceso. En la figura adjunta se muestran los mapas de las 16 características extraídas del proceso, obtenidos mediante el KRSOM:

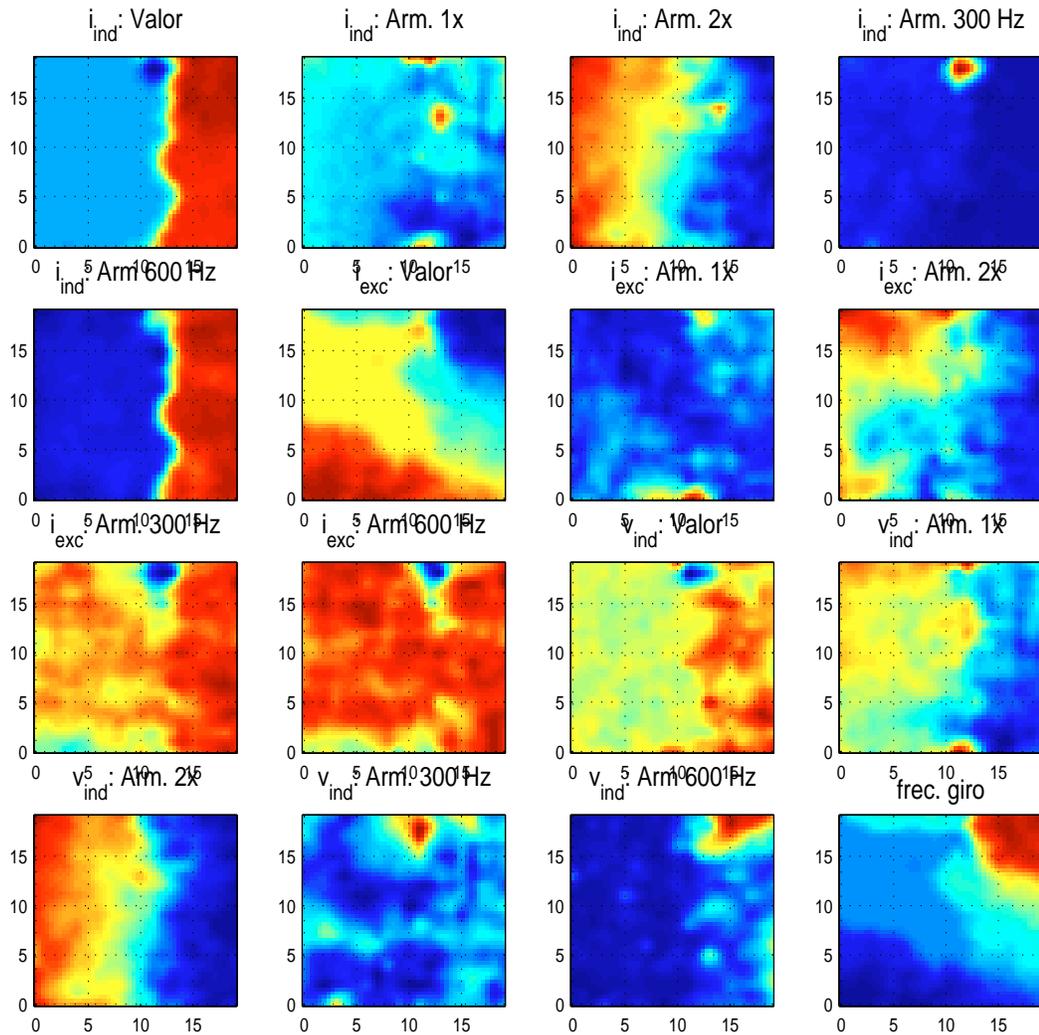


Figura 6.22: Mapas de las 16 características del proceso.

Analizando el mapa de la característica 1, correspondiente al valor de  $i_{ind}$ , puede apreciarse que las dos zonas principales delimitadas en el mapa de distancias corresponden a valores bajos (izquierda) y altos (derecha) de la corriente de inducido, lo que nos permite corroborar la conclusión anterior de que ambas zonas se corresponden con las situaciones de funcionamiento en vacío y laminación, respectivamente.

Un examen más minucioso de los mapas permite, por ejemplo, des-

cribir las diferencias existentes entre las subregiones que se aprecian en la región de laminación. Observando la característica 16, correspondiente a la frecuencia de giro  $f_g$ , la región superior derecha corresponde a velocidades de giro elevadas, mientras que las regiones inferiores corresponden a velocidades más bajas. Podemos concluir que las diferentes subregiones corresponden a laminaciones a diferentes velocidades. Así, como se reflejará más adelante en los mapas de activación, la tercera de las laminaciones en el conjunto de entrenamiento se proyectará sobre la región superior derecha, mientras que las dos primeras laminaciones se proyectan sobre las regiones inferior (al inicio de la laminación) y central (durante el régimen permanente de la laminación) en la zona derecha.

**Caza de Correlaciones.** Otra forma de extraer información sobre el proceso a partir de los mapas de características es la *caza de correlaciones*. Dos variables correladas tienden a generar mapas de características con geometrías similares. Un ejemplo muy claro de correlación entre variables lo vemos entre los valores de  $i_{exc}$  y  $f_g$ , cuyos mapas son casi idénticos, aunque uno es el “negativo” del otro, lo cual significa que están fuertemente correlados aunque dicha correlación es negativa. En efecto en un motor de CC, a tensión constante una disminución en la corriente de excitación se corresponde con un incremento en la velocidad y viceversa.

Otras dos variables fuertemente correlacionadas son la corriente de inducido  $i_{ind}$  y su armónico de 600 Hz,  $i_{ind}^{600}$ . Este último corresponde al doble puente trifásico de tiristores existente en la etapa de potencia, que originan señales con un desfase de  $30^\circ$  (12 veces cada 20 ms), cuya amplitud es proporcional a la corriente de inducido,  $i_{ind}$ , lo que explicaría la correlación entre ambas.

Puede apreciarse una correlación negativa entre los mapas de la corriente de inducido,  $i_{ind}$ , y su armónico de 300 Hz,  $i_{ind}^{300}$ , especialmente en torno a la región de coordenadas (11,17) –arriba al centro–. En el motor analizado, los dos puentes de tiristores funcionan de forma distinta. Mientras que uno de los puentes puede motorizar y regenerar, el otro sólo es capaz de motorizar. Durante la laminación, ambos motorizan y tienen un comportamiento idéntico. Sin embargo, durante los periodos de regeneración –zona en torno a (11,17)–, que ocurren especialmente al final de cada laminación, su comportamiento es totalmente distinto.

Dado que cada uno de ellos tiene un desfase de  $60^\circ$  y están intercalados,  $30^\circ$ , un funcionamiento distinto, implica la aparición de un armónico de 300 Hz que no existe cuando se da una perfecta simetría –en cuyo caso sólo se da el armónico de 600 Hz–. Durante el funcionamiento en vacío, el motor trabaja en el límite entre motorización y regeneración, lo que origina la aparición –aunque más tenue– del armónico de 300 Hz, tal y como se aprecia en los mapas de características.

Otra fuerte correlación negativa puede encontrarse entre  $i_{ind}$  y  $V_{ind}^{2x}$  –y, aunque en menor medida, con  $i_{ind}^{2x}$  e  $i_{exc}^{2x}$ –. El origen de esta correlación no es tan evidente. Los armónicos 1x y 2x –múltiplos de la velocidad de giro–, son origen mecánico. Durante su funcionamiento normal, el subsistema mecánico acoplado al motor, debido a desajustes mecánicos tales como defectos en la alineación, masas asimétricas, etc., produce una demanda de par con una periodicidad correspondiente a la frecuencia de giro. Cuando el defecto mecánico es meramente un desajuste la forma de onda es aproximadamente senoidal. Los desajustes suelen tener un carácter permanente respecto al rango de tiempos en el que se desarrolla este caso, lo que explica la relativa invarianza del armónico 1x frente a los distintos estados del proceso. Otros tipos de defectos mecánicos, como en el caso de zonas muertas en los engranajes, golpeteos, etc. dan lugar a formas de onda no senoidales lo que origina armónicos múltiplos de la fundamental. En este caso, durante la laminación el sistema mecánico “engrana” y no se producen golpes en los piñones, dado que están siempre trabajando en tracción. Cuando el motor trabaja en vacío, se producen holguras en los piñones lo que origina un “golpeteo” periódico. Esto se traduce en demandas de par no senoidales con la consiguiente aparición del armónico 2x observada en la zona izquierda (vacío) de los mapas de características de  $V_{ind}^{2x}$ ,  $i_{ind}^{2x}$  y  $i_{exc}^{2x}$ .

**Determinación de las Regiones a partir de datos.** Otra forma de asociar las condiciones de funcionamiento del proceso con las regiones en el mapa son los *mapas de activación*. Mientras los mapas de características permiten sacar partido al conocimiento apriorístico de un experto en el proceso para la determinación de regiones, los mapas de activación permiten determinar regiones en el mapa excitadas por conjuntos concretos de datos.

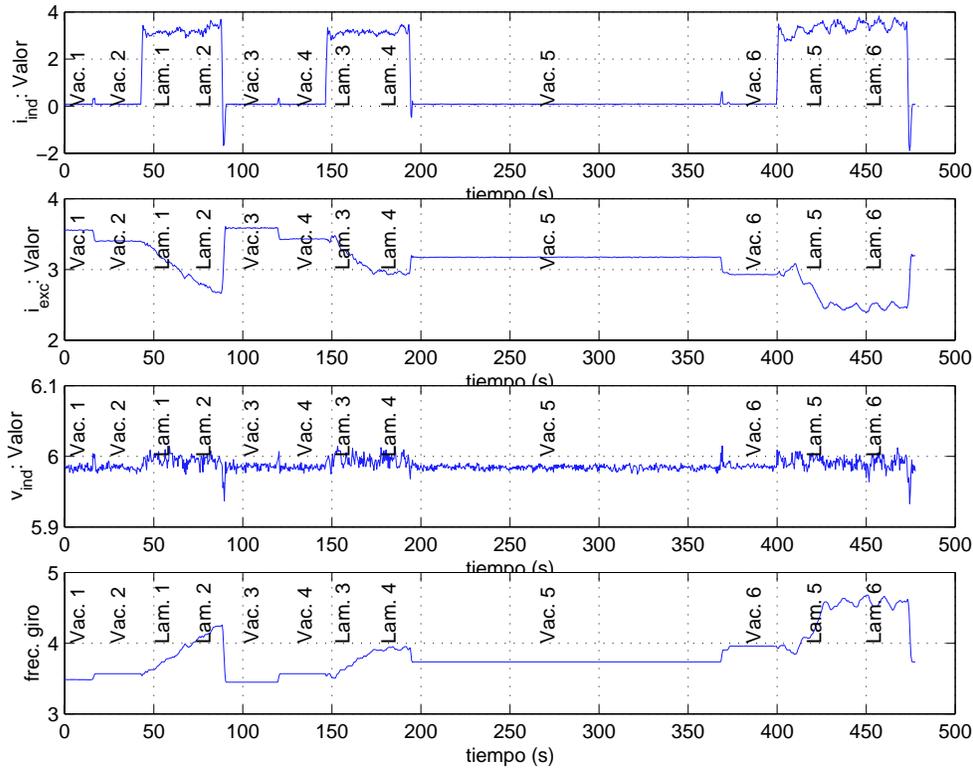


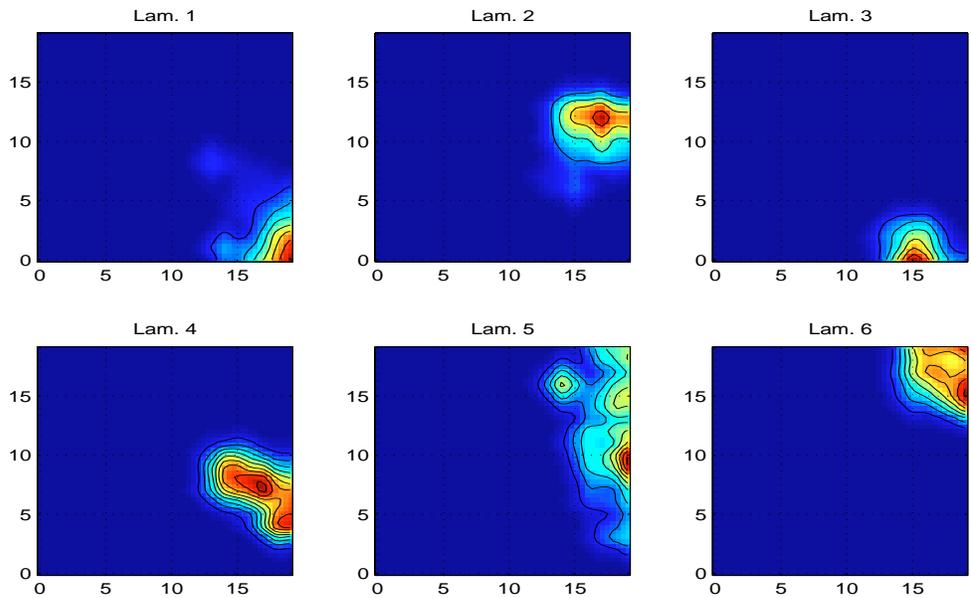
Figura 6.23: Conjuntos de datos correspondientes a diversas condiciones de funcionamiento del motor.

En la figura 6.24(a) se muestran los mapas de activación correspondientes a diversos periodos de laminación, detallados en la figura 6.23. Comparando los mapas de activación con los tramos correspondientes en la figura 6.12 y con los mapas de características pueden apreciarse las diferencias sutiles existentes entre las diferentes regiones. Lógicamente, se verifica todo lo deducido a partir del mapa de distancias y de los mapas de características. Las diferencias entre las regiones obedecen a diferencias más concretas en algunas de las variables, lo que puede comprobarse tras un examen minucioso de todos los gráficos. Del mismo modo, en la figura 6.24(b) se muestran los mapas de activación correspondientes a diversos periodos de funcionamiento en vacío. Puede comprobarse cómo las regiones están en la parte izquierda.

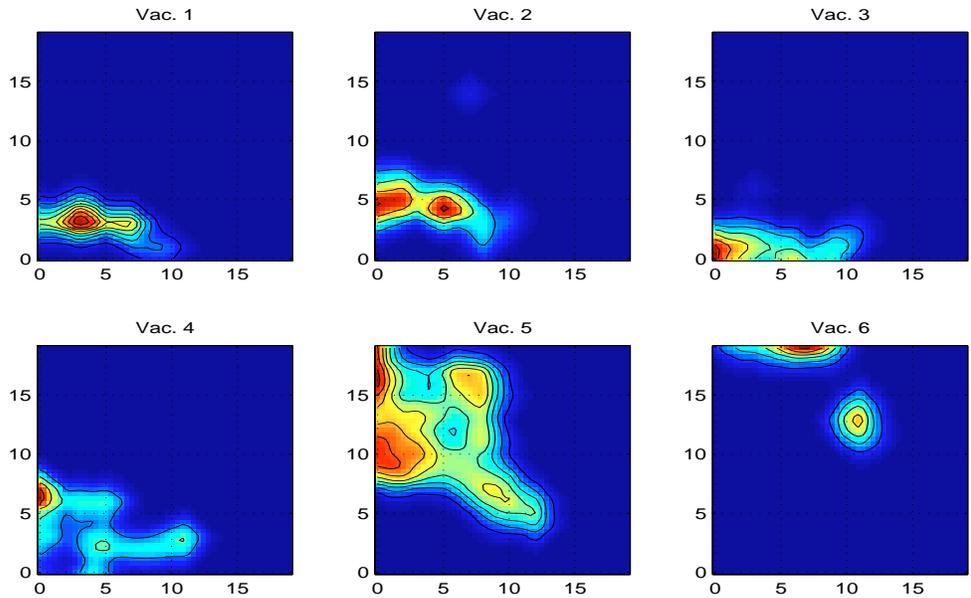
La información que aportan los mapas de activación permite etiquetar las regiones del mapa como se muestra en la figura 6.25. Desde un punto

de vista práctico, el trazado de etiquetas en el espacio de visualización es de gran utilidad, ya que permite al usuario conocer en todo momento la condición del proceso de un golpe de vista: basta con comprobar en qué región se encuentra el puntero en cada momento.

**Mapas de activación frente a mapas de características.** La combinación de los mapas de activación con los mapas de características constituye una potente herramienta para el análisis de los datos, como se puede comprobar en la figura 6.26. En dicha figura, se muestra una combinación de los mapas de características (color de fondo) y los mapas de activación (curvas de nivel) correspondientes a distintos estados del motor. Puede apreciarse claramente en el mapa de  $i_{ind}$  cómo las regiones correspondientes a periodos de laminación se comprenden regiones del mapa con altos valores de dicha variable agrupándose a la derecha, mientras que las de periodos de funcionamiento en vacío se ubican a la izquierda donde los valores de  $i_{ind}$  son menores. Asimismo, los mapas de corriente de excitación,  $i_{exc}$  y la frecuencia de giro  $f_g$  distinguen claramente el funcionamiento entre las zonas “Vac. 1” y “Vac. 5” así como entre “Lam. 1” y “Lam. 6”. Las zonas “Vac. 1” y “Lam. 1” discurren a menor velocidad y con mayor corriente de excitación. Finalmente, puede apreciarse cómo la energía del armónico 2x de la corriente de inducido es muy superior en las dos regiones de funcionamiento en vacío.



(a) Laminando.



(b) Funcionando en vacío.

Figura 6.24: Mapas de Activación del Caso 2 para diferentes condiciones de funcionamiento del proceso.

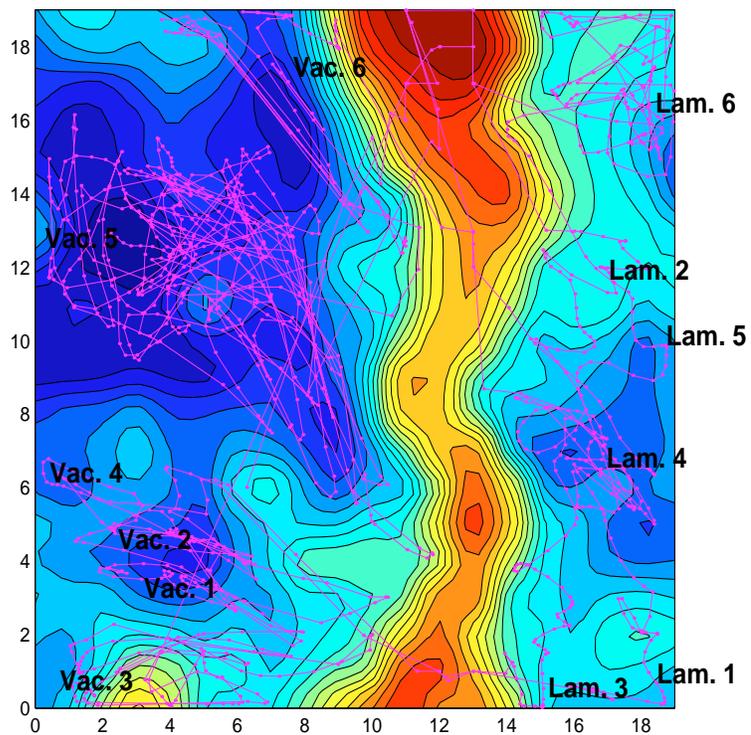


Figura 6.25: Etiquetado de regiones en el espacio de visualización. Como fondo figura el mapa de distancias interneuronales.

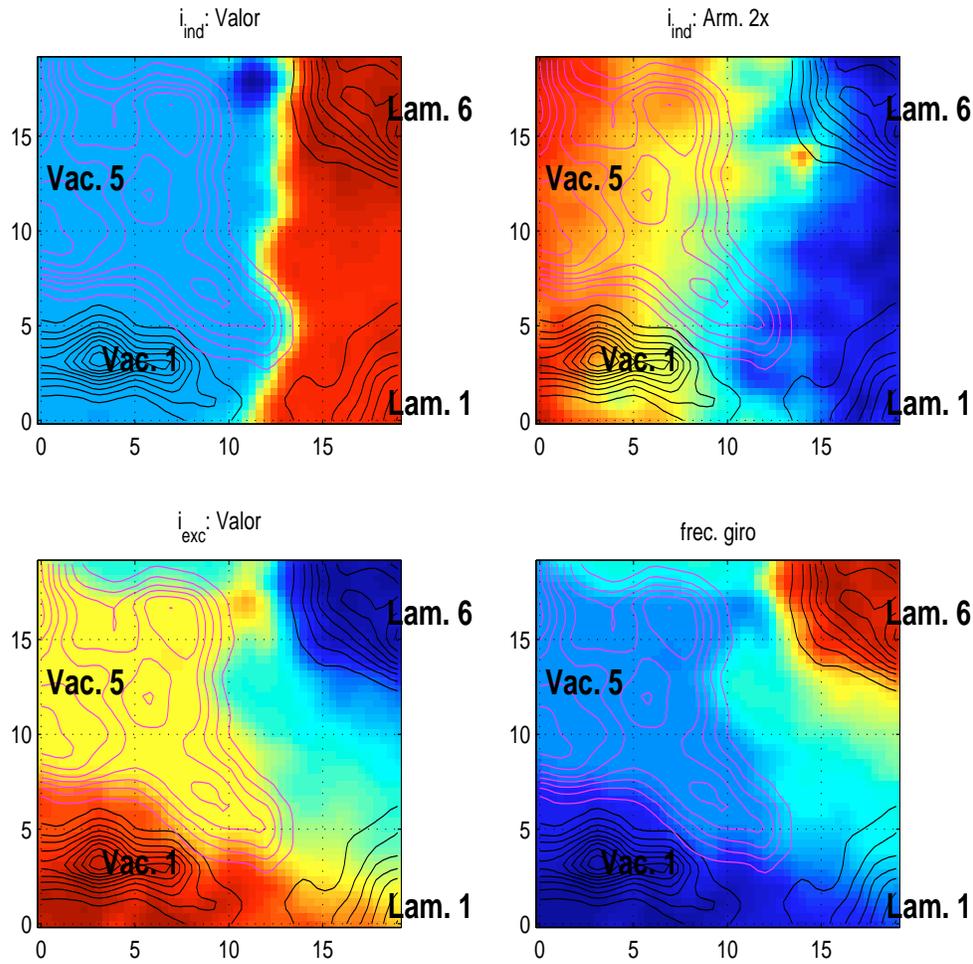


Figura 6.26: Las regiones de activación situadas sobre mapas de características permiten extraer información útil sobre las características del proceso en cada condición de funcionamiento.

**Detección de anomalías. Errores de Modelado.** Una vez generado el modelo del proceso, el proceso puede atravesar condiciones diferentes a las consideradas en el conjunto de datos de entrenamiento. Esta situación puede darse por varios motivos, tales como indisponibilidad de datos correspondientes a ciertas condiciones del proceso o cuando, aun disponiendo de ellos, éstos no recubren adecuadamente todos los modos de funcionamiento posibles dentro de dichas condiciones.

Los errores de modelado global y vectorial (ver sección 5.6) permiten detectar desviaciones del proceso respecto a los datos utilizados en el entrenamiento. En las figuras 6.27 y 6.28 se muestra la evolución del error de modelado para los datos de entrenamiento y de test, respectivamente. Los mismos gráficos del error de modelado se repiten con más detalle en la figura 6.29, en la que se muestra una etiqueta con cada variable en el error de modelado vectorial. En la primera de ellas, correspondiente al conjunto de datos de entrenamiento, puede verse que los errores de modelado –tanto el global como el vectorial– se mantienen en valores muy bajos, lo que indica que el modelo es autoconsistente, es decir, se ajusta correctamente a los datos. La figura 6.28, correspondiente a los datos de test, muestra varias laminaciones realizadas justo después de una parada de mantenimiento en el tren. Al principio del registro, el motor gira en vacío pero las consignas que llegan del computador de proceso son distintas de las habituales y los valores de las variables difieren totalmente de los que adopta el proceso en condiciones normales lo que provoca un elevado error de modelado global: se trata, de hecho, de una condición nueva, no contemplada en el entrenamiento.

Un vistazo al gráfico del error de modelado vectorial en la figura 6.29 permite apreciar inmediatamente qué variables están fuera de los límites. El error de modelado vectorial muestra los siguientes factores:

- Los armónicos mecánicos de la corriente de inducido,  $i_{ind}^{1x}$ ,  $i_{ind}^{2x}$ , y el armónico de 300 Hz,  $i_{ind}^{1x}$ , se mantienen en valores inferiores a lo esperado.
- La corriente de excitación  $i_{exc}$  tiene valores muy elevados.
- Armónicos  $V_{ind}^{1x}$  y  $V_{ind}^{2x}$  en valores bajos, es decir, un bajo rizado de origen mecánico en la tensión de inducido.
- La frecuencia de giro  $f_g$  tiene valores inferiores a lo esperado.

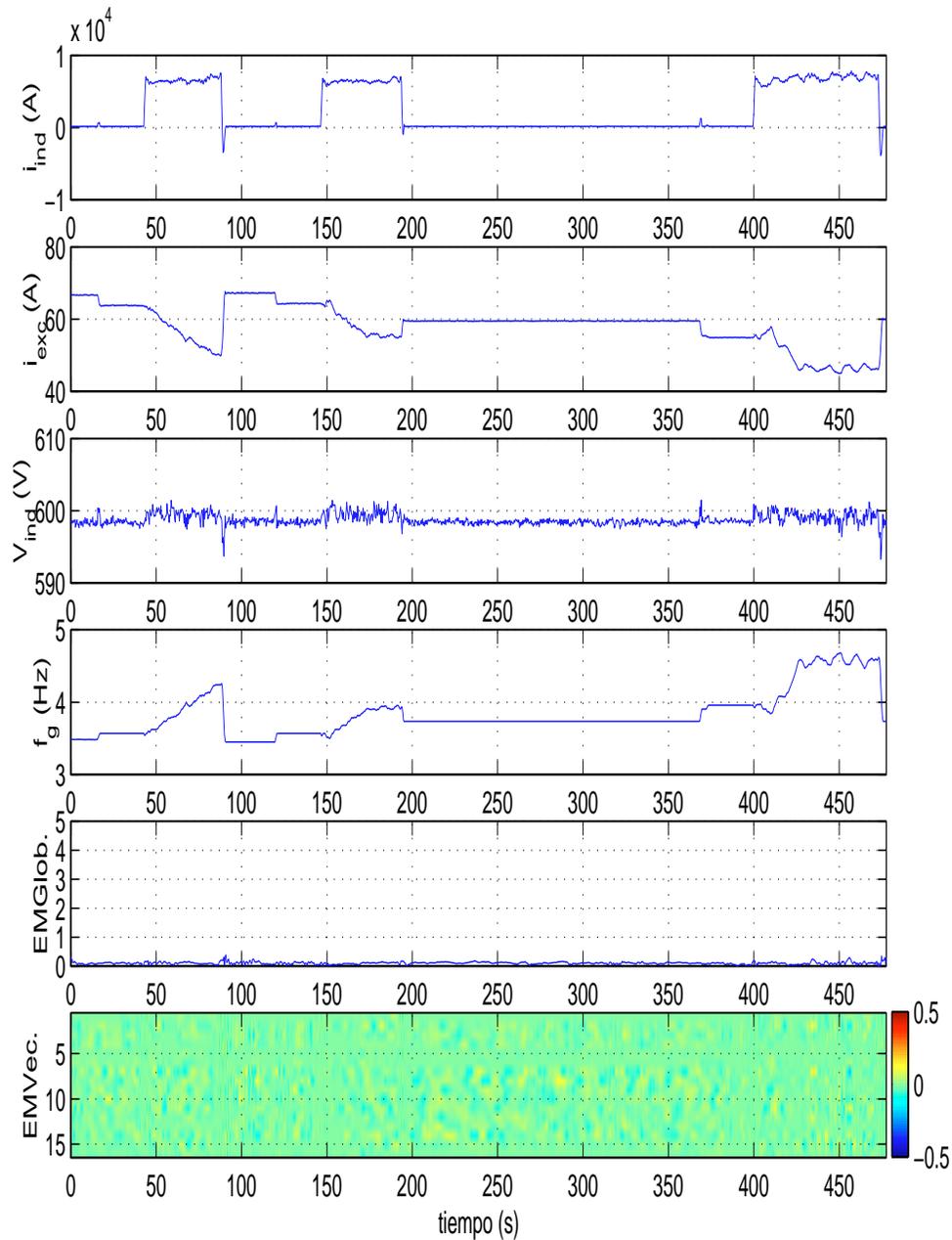


Figura 6.27: Errores de modelado global y vectorial para los datos de entrenamiento.

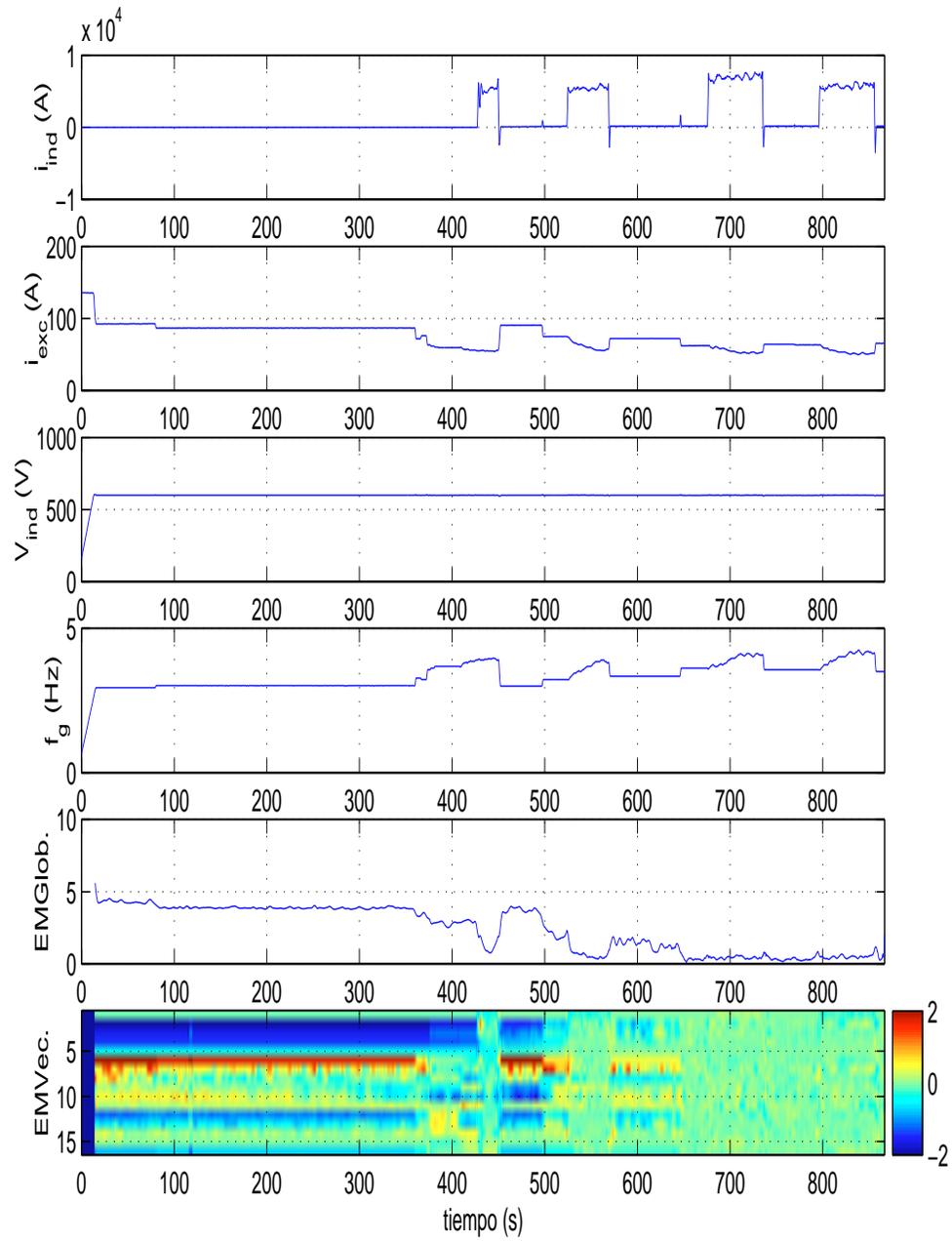
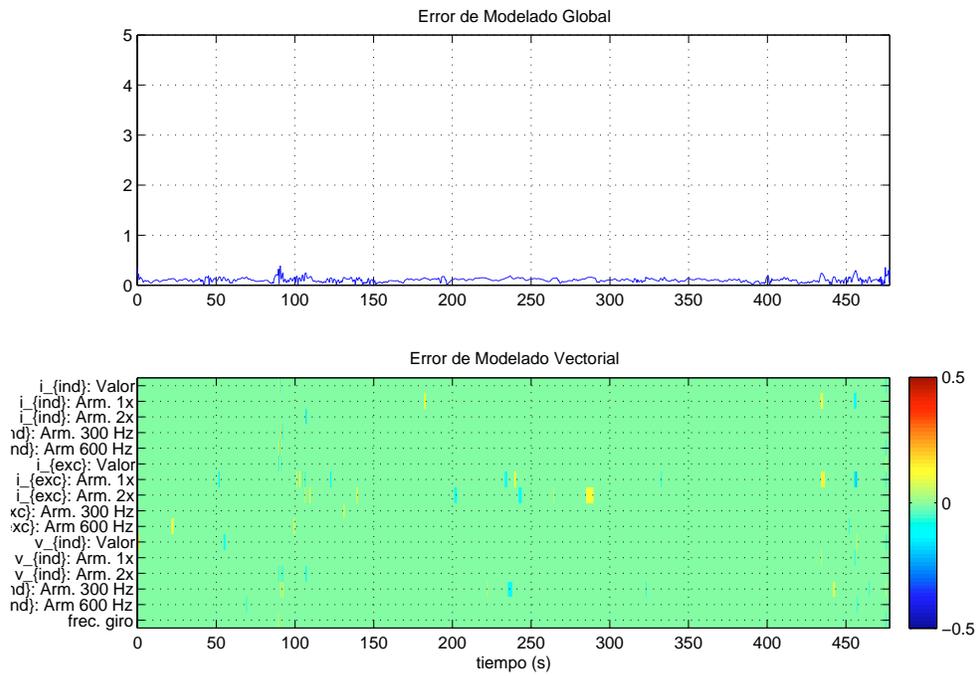
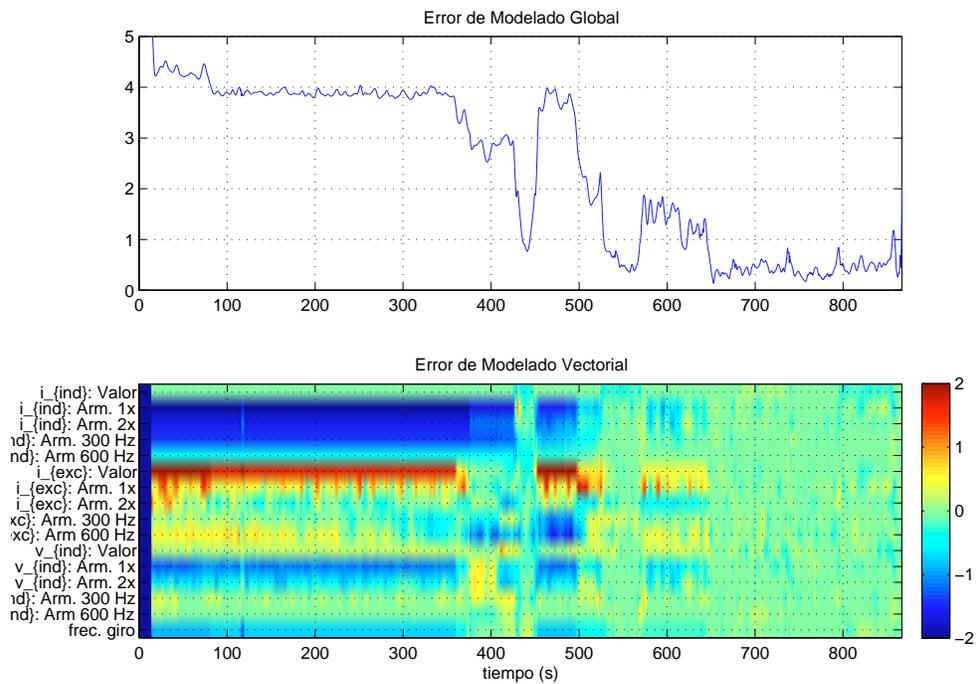


Figura 6.28: Errores de modelado global y vectorial para los datos de test.



(a) Datos de entrenamiento.



(b) Datos de test.

Figura 6.29: Detalle de los errores de modelado global y vectorial para los conjuntos de entrenamiento y de test.

En la figura 6.30, se muestra la trayectoria descrita por las proyecciones del estado del proceso sobre el mapa de distancias interneuronales. Como cabe esperar, la trayectoria discurre por diversas zonas, tanto de laminación como de vacío. En la trayectoria se aprecian diversas agrupaciones que corresponden a los distintos estados de funcionamiento que atraviesa el motor.

A diferencia de las proyecciones correspondientes a los datos de entrenamiento, existen aquí algunas agrupaciones ubicadas en regiones de elevadas distancias interneuronales –obsérvese la región en torno al punto de coordenadas (10, 1)–. En general, las regiones con elevadas distancias interneuronales corresponden a zonas con escasa activación y, en definitiva, poco probables según el modelo. En efecto, un examen más minucioso revela que estas agrupaciones corresponden a las primeras regiones de funcionamiento en vacío, como se revela en la figura 6.31 cuyos errores de modelado –ver fig. 6.28– apuntan a una disparidad frente al modelo.

En la figura 6.32 se muestran las regiones de activación (contornos) correspondientes a cada una de las cuatro laminaciones realizadas durante el conjunto de datos de test trazadas sobre el mapa de distancias. Como se puede comprobar en la figura 6.28, los errores de modelado durante estas laminaciones no son muy elevados, por lo que las posiciones en el mapa pueden considerarse representativas del estado de funcionamiento. Puede apreciarse en los mapas de activación que las diversas laminaciones ocupan distintas posiciones en la región derecha lo que, a la vista de los mapas de características, refleja diferencias en algunas características, especialmente en cuanto a la frecuencia de giro, aunque otras variables como  $i_{exc}^{300}$  e  $i_{exc}^{600}$  diferencian la primera laminación –región en torno a (14, 5), en la que ambas variables tienen valores ligeramente inferiores– del resto.

Finalmente, siguiendo el esquema propuesto en la metodología, los datos correspondientes a regiones de elevados errores de modelado –en este caso los tres primeros tramos de funcionamiento en vacío– deberían incorporarse a la base de datos de entrenamiento y procederse al reentrenamiento del modelo y etiquetado de las nuevas regiones. Con ello, tras la siguiente parada el error de modelado no dará cuenta de una situación desconocida y existirá una región etiquetada en el mapa describiendo dicha condición.

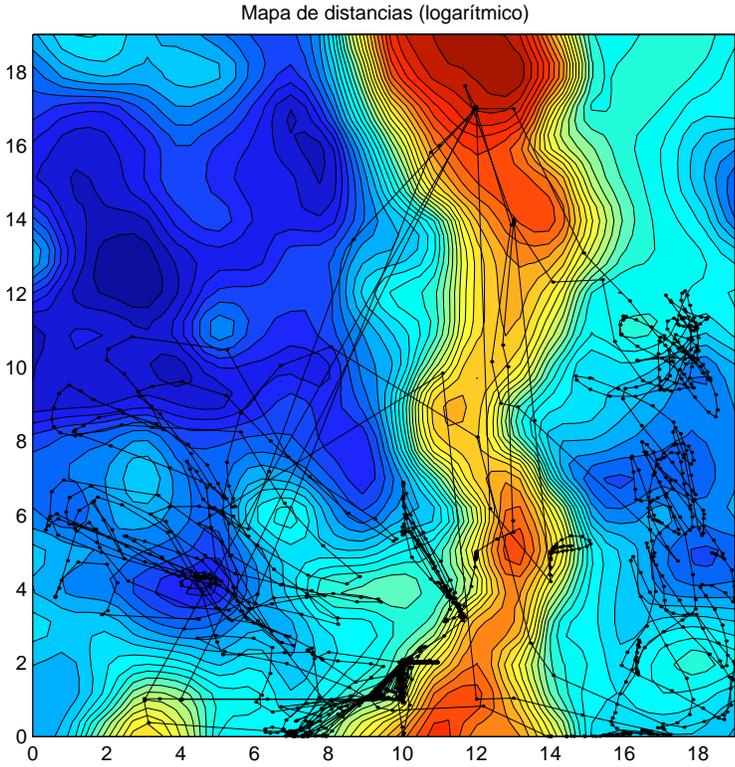


Figura 6.30: Mapa de distancias junto con las proyecciones de los datos de test.

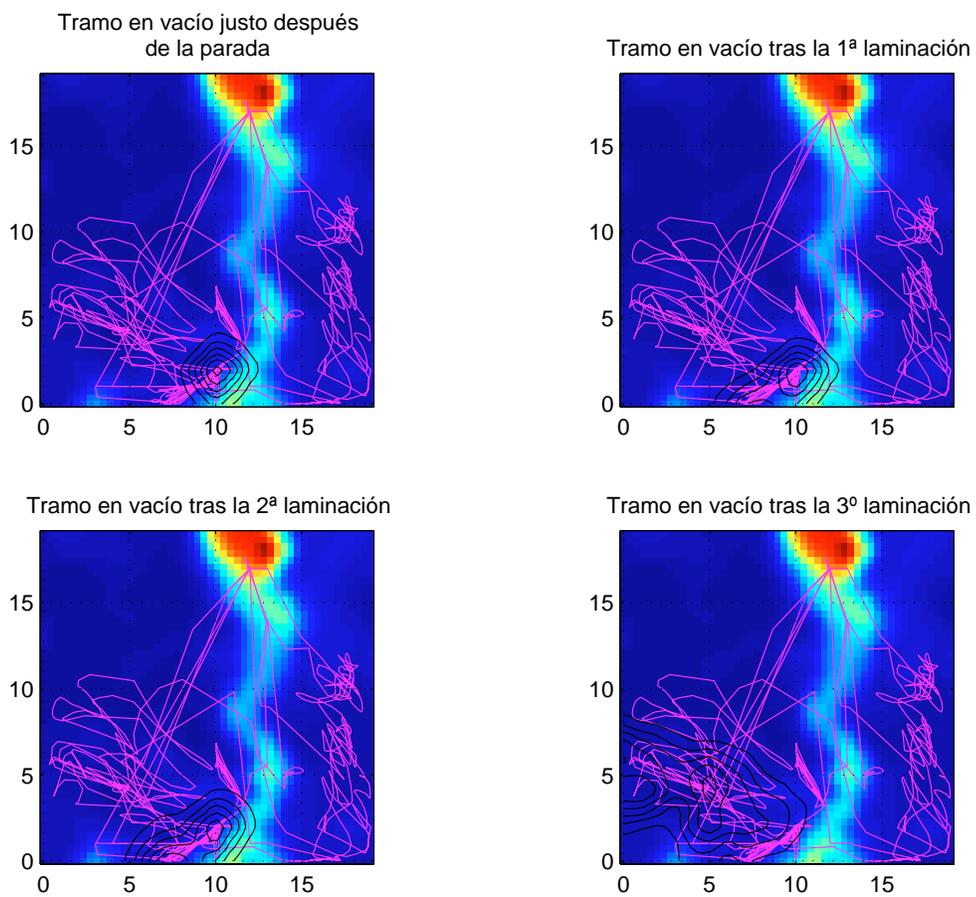


Figura 6.31: Mapas de activaciones de diversos tramos en vacío.

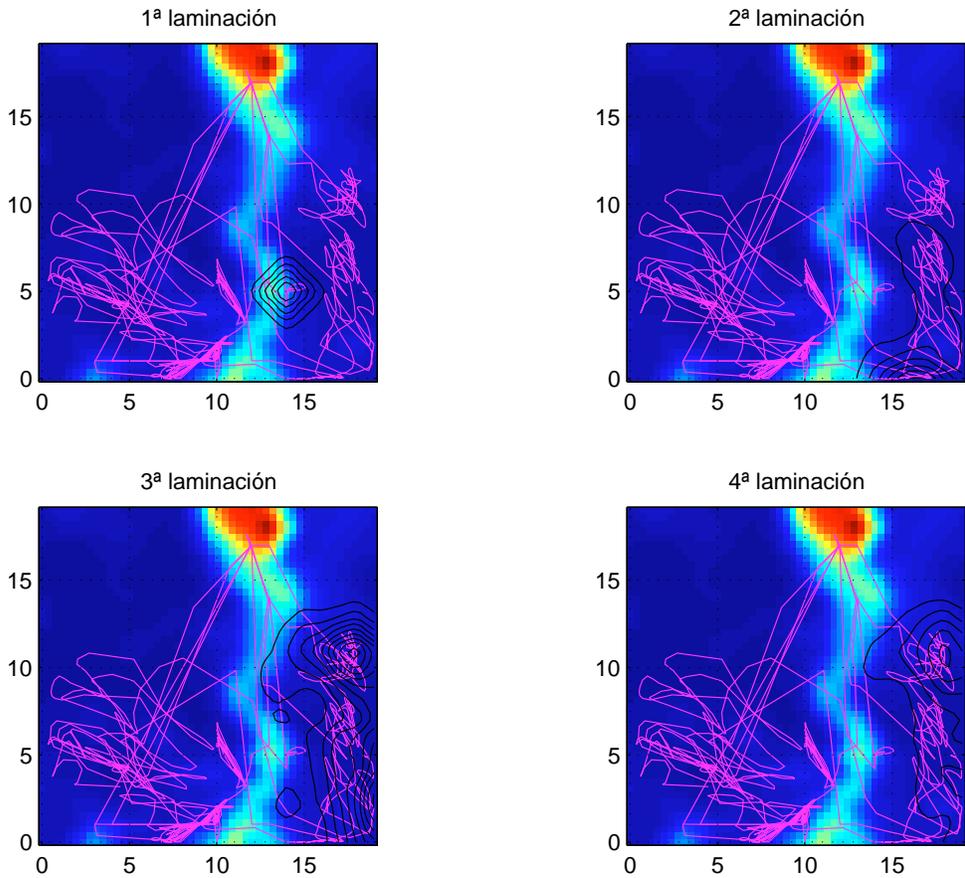


Figura 6.32: Mapa de activaciones de diversos tramos laminando.



# Capítulo 7

## Discusión General y Líneas Futuras

### 7.1 Conclusiones finales

El contenido de esta Tesis se enmarca dentro de uno de los paradigmas en el campo de la monitorización y diagnóstico procesos: el de los métodos basados en datos (denominados también *model-free*, en oposición a los métodos *model-based*). Este tipo de métodos se brinda como una solución idónea en procesos altamente complejos, en los que la modelización se hace imposible o excesivamente costosa. Su principal ventaja reside en la capacidad de extraer la información presente en los datos del proceso. Aunque existen innumerables técnicas de minado de datos (*data mining*) orientadas a extraer el conocimiento latente en masas ingentes de datos, el enfoque adoptado en esta Tesis orientó el esfuerzo de investigación hacia los métodos de reducción de la dimensionalidad.

Se pudo comprobar cómo bajo ciertas condiciones relativamente débiles, exigibles al sistema de medidas y a las técnicas de extracción de características, el problema de la monitorización de la condición puede resolverse mediante técnicas capaces de detectar las estructuras geométricas fundamentales de los datos del proceso y sintetizarlas en unos pocos parámetros. En todas ellas subyace el concepto esencial de *componentes principales* y todas ellas son, en un modo u otro, generalizaciones no lineales del método PCA.

Desde el inicio de las investigaciones de esta Tesis, el SOM se reveló

como una herramienta extremadamente eficiente: era capaz de captar la geometría inherente a los datos y, además, mediante un algoritmo sorprendentemente simple y robusto. Se encontraron algunas deficiencias, citadas en esta Tesis, como el hecho de no constituir una aproximación exacta de la función de densidad de los datos, así como el carácter discreto de la proyección, siendo éste último el principal inconveniente.

El estudio de las redes de base radial (RBF's) pronto condujo a las redes GRNN (*Generalized Regression Neural Network*), una versión regularizada de las redes de base radial, fundamentada en el concepto de regresión de núcleo. El carácter sencillo e intuitivamente sugerente – evoca conceptos como el de centro de masas – de este método, así como su robustez (especialmente en problemas con pocos datos y alta dimensionalidad) animaron a centrar el esfuerzo en este tipo de métodos, lo que pronto llevó a descubrir la sólida base estadística que las respalda.

Se trató de conectar las propiedades estadísticas del SOM con las de la regresión de núcleo con lo que se pretendió justificar un modelo continuo para el SOM cuyas propiedades constituyeran –al menos asintóticamente– una extensión de las del SOM original. La consecución de un mapa continuo, con las propiedades del SOM, y de entrenamiento sencillo, robusto, e intuitivo dio pie al desarrollo de una metodología dentro del paradigma de los métodos de reducción de dimensionalidad para la monitorización en continuo de procesos.

Las connotaciones biológicas del SOM no pasaron desapercibidas. El SOM organizaba la información adquirida de manera topológica, reservando regiones para grupos de estímulos (situaciones o condiciones del proceso) que guardaban alguna relación entre sí, al igual que el cerebro de los animales superiores. Los mapas de activación sugeridos en la metodología propuesta, constituyen un método que permite “teñir” o colorear la región que el SOM destina para cada estímulo, al igual que los mapas cerebrales de los estímulos obtenidos mediante PET <sup>1</sup>. Su utilidad es enorme ya que permite describir –de manera continua, utilizando la técnica propuesta– la morfología de la región asociada a un estímulo, consistente en una condición conocida del proceso.

El resultado final de la metodología propuesta, es una representación de la evolución del proceso que permite visualizar, de manera muy intuitiva, gran cantidad de información sobre éste de forma gráfica, tal

---

<sup>1</sup>*Positron Emission Tomography*

como la evolución de las características del proceso con significado físico (temperaturas, intensidades, etc.), a la vez que se determina su condición en función de las regiones que atraviesa la proyección de la trayectoria de estado. Al mismo tiempo, otro tipo de información, más abstracta pero también intuitiva, se deriva de los mapas de distancias, que revelan las regiones más importantes del mapa, o de los propios mapas de activaciones.

Finalmente, se introduce el error de modelado, que permite describir la divergencia entre la evolución del proceso y el modelo generado a partir de los datos, lo que permite detectar la presencia de condiciones de funcionamiento no consideradas en el conjunto de datos de entrenamiento. Se describe asimismo, el error de modelado vectorial, que considera el error de modelado independientemente para cada una de las características lo que permite visualizar las variables más susceptibles de haber sufrido desviaciones, así como el signo de las desviaciones y su severidad, aportando a nuestro juicio una información de gran utilidad para el diagnóstico de la anomalía.

## 7.2 Aportaciones

Las principales aportaciones que, a juicio del autor, ofrece este trabajo son las siguientes:

- Se sugiere la descripción de procesos complejos bajo un enfoque enraizado en la Teoría de Sistemas, considerándolos como sistemas dinámicos no lineales y se plantea el problema de la monitorización de la condición en términos de una proyección no lineal o *mapa* de la trayectoria del estado del proceso sobre un espacio de baja dimensionalidad (2D ó 3D) o *espacio de visualización*, preservando la información relativa a la condición del proceso.
- Se propone una arquitectura del SOM, basada en el estimador de Nadaraya-Watson, que constituye una generalización del SOM al caso continuo. Este estimador es extremadamente simple, tanto en la ejecución como en el ajuste de los parámetros, tiene una sólida base estadística y, según Haykin, constituye una forma de regularización de una red RBF.

- Se propone una generalización de la metodología propuesta por Kohonen en [22] para la monitorización de la condición de procesos. Dicha metodología emplea la arquitectura propuesta para la generación de mapas continuos de características, distancias y activaciones, así como la proyección continua de la trayectoria de estado del proceso sobre el espacio de visualización.
- El carácter continuo de las proyecciones permite visualizar tendencias, ciclos límite, derivas, etc. en su estado más incipiente. Esta posibilidad no la permite el SOM original de Kohonen dado el carácter discreto –retícula de  $N_i \times N_j$  nodos– del espacio sobre el que se proyectan los datos.
- Se propone el uso de *mapas de activación* –no considerados en la metodología de Kohonen– para la determinación de las regiones en el espacio de visualización correspondientes a conjuntos determinados de datos del proceso, lo que permite la descripción de regiones en el espacio de visualización asociadas a condiciones de funcionamiento concretas.
- Se proponen *técnicas de visualización 3D* basadas en la aplicación de técnicas MDS y el estimador de Nadaraya-Watson sobre los vectores de codificación del SOM en el espacio de características. Estas técnicas, aunque pueden demandar rutinas de programación gráfica muy específicas –sombreados, ocultación de superficies, etc.–, permiten aportar visualmente información sobre las distancias y posiciones relativas entre las condiciones del proceso y, en general, sobre la geometría de los datos de proceso.
- Se describe un procedimiento para detectar condiciones de funcionamiento no consideradas en el entrenamiento a través del *error de modelado*, concepto éste similar al de *residuo* o *error de redundancia analítica*. Este error, evaluado independientemente para cada característica –*error de modelado vectorial*–, aporta además una información selectiva, al poner de manifiesto aquéllas variables más susceptibles de causar la desviación, lo que, además de la detección, facilita el diagnóstico de la anomalía.

## 7.3 Futuras Líneas de Trabajo

El resultado de la investigación de esta Tesis abre un sinfín de líneas de trabajo. Se tratará de enumerar a continuación algunas de las que se consideran más importantes:

**Integración de Modelos:** Una interesante línea de investigación no explorada en esta Tesis es la fusión de la metodología propuesta con métodos basados en modelo. Los residuos generados en este tipo de métodos, ya sea mediante las ecuaciones de paridad, observadores u otras técnicas, constituyen variables con un alto grado de empaquetamiento y una alta especificidad en cuanto a la condición de un proceso dado y se ajustan a las condiciones sugeridas en las secciones 1.4 y 2.1. Esta fusión permitiría incorporar información *apriorística* derivada de modelos parciales o totales del proceso que se complementarían con la aportada por los datos del proceso.

**Mapas de Certeza:** Otra posibilidad de incorporar información *apriorística* que abre otra excitante línea de investigación es la generación de mapas de certeza mediante *sistemas de inferencia borrosos* (FIS, *Fuzzy Inference Systems*). Las características de cada vector de codificación del SOM pueden someterse a un banco de reglas borrosas, determinadas a partir de un conocimiento experto previo, que permitan inferir un nivel de certeza de una proposición dada (relativa, por ejemplo, a un tipo de fallo concreto), para cada vector de codificación. El mismo procedimiento descrito en el capítulo 5 para visualizar los mapas continuos de características, distancias y activaciones puede utilizarse para visualizar en un mapa los niveles de certeza o verosimilitud de dicha proposición de manera continua.

**SOM de topología Variante:** El uso de topologías distintas a la rectangular (p.ej. toroidal, en árbol, geometría cúbica o hipercúbica, etc.) puede permitir al SOM adaptarse a geometrías específicas de los datos del proceso [22]. Aunque es difícil contar con un conocimiento *apriorístico* del tipo de geometría de los datos de un proceso, cabe la posibilidad de reconfigurar la topología del SOM, en tiempo de entrenamiento, mediante algún algoritmo heurístico, eliminando y/o añadiendo conexiones entre neuronas en función de algún criterio. La misma topología resultante del SOM aportaría,

además, información sobre la geometría e interrelaciones entre las distintas condiciones del proceso (p.ej. una posible organización jerárquica de los tipos de fallo o modos de funcionamiento).

**Núcleos Hiperelipsoidales:** El estimador de Nadaraya-Watson, aplicado en esta Tesis para la obtención de mapas continuos se basa en núcleos gaussianos de simetría esférica. La geometría de los núcleos empleados tiene, sin duda, una importante repercusión en la calidad de los mapas generados. Caben diversas generalizaciones para el estimador propuesto, tales como emplear factores de ancho –*width factor*,  $\sigma$  en la ecuación (4.4)– distintos para cada neurona o, de forma más general, núcleos hiperelipsoidales –Ec. (4.5), donde el factor de ancho es una matriz de covarianzas–. Hasta donde el autor ha podido comprobar, la cuestión relativa a la influencia de los parámetros de éstas en la calidad de la solución final, así como la determinación de sus valores óptimos es un tema aún no resuelto.

**Kernel PCA:** La idea central en los métodos PCA de núcleo, descritos en la sección 3.3, consistente en aplicar métodos clásicos en un espacio de características de mayor dimensionalidad inducido por una transformación no lineal definida mediante un núcleo abre un vasto campo de investigación. Según Scholkopf [49] la cuestión relativa al núcleo que debe aplicarse para cada problema concreto es aún un problema abierto. Una elección adecuada del mapa de características, puede convertir un problema altamente complejo en otro fácilmente resoluble mediante métodos clásicos, incluso lineales. Esto es aplicable tanto a la clasificación como a la generación de mapas, la estimación o la predicción, todos ellos campos estrechamente ligados a la detección y el diagnóstico de fallos.

# Apéndice A

## Nomenclatura

A continuación se muestra una lista de algunos de los símbolos utilizados en esta Tesis junto con su significado más habitual:

$X$	Espacio de estados
$E$	Espacio de entradas
$S$	Espacio de salidas (medidas)
$x(t)$	Vector de estados del proceso
$y(t)$	Vector de salidas (medidas) del proceso
$u(t)$	Vector de entradas del proceso
$\mathcal{C}_i$	$i$ -ésima condición de funcionamiento del proceso
$C_i$	Conjunto del espacio de estados en el que el proceso se considera en la condición $\mathcal{C}_i$
$C'_i$	Conjunto del espacio de medidas generado por los estados de $C_i$
$C''_i$	Conjunto del espacio de visualización generado por los estados de $C_i$
$f(.,.,.)$	Función de estado
$g(.,.,.)$	Función de observación
$X(n), Y(n), \dots$	Coefficientes de la DFT de $\{x_k\}, \{y_k\}, \dots$
$S_{xx}(f)$	Espectro de potencia de la señal $\{x_k\}$
$c_x(n)$	Coefficientes cepstrales de $\{x_k\}$
$\text{cum}(x_1, \dots, x_n)$	Cumulante de $x_1, \dots, x_n$
$r_{xx}(k)$	Secuencia de autocorrelación de la señal $\{x_k\}$

$\mathbf{R}_{xx}$	Matriz de correlaciones de la variable $x$
$\lambda_i$	$i$ -ésimo valor propio
$\mathbf{u}_i$	$i$ -ésimo vector propio
$\varepsilon$	término residual en el método PCA
$\psi(\mathbf{x})$	En general, mapa no lineal de $\mathbf{x}$
$K(\mathbf{x}, \mathbf{y})$	Núcleo en el método Kernel PCA
$\mathbf{K}$	Matriz de núcleo formada por $\{K(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1,\dots,N}$
$\mathbf{x}$	Punto del espacio de entrada, dato.
$n$	Dimensión del espacio de entrada o de datos.
$p$	Dimensión del espacio de salida o de visualización.
$c$	Índice de la neurona ganadora.
$N$	Número de datos de un conjunto
$K$	Número de neuronas en el SOM
$\mathbf{m}_i$	$i$ -ésimo, vector de codificación del SOM
$\mathbf{m}_c$	Vector de codificación correspondiente a la neurona ganadora.
$\mathbf{g}_i$	$i$ -ésima coordenada en la retícula del SOM
$\mathbf{g}_c$	Coordenada en la retícula correspondiente a la neurona ganadora.
$h_{ci}$	Grado de vecindad en la retícula entre las neuronas $c$ e $i$
$\sigma(t)$	Amplitud del campo receptivo en el SOM en el instante $t$
$\alpha(t)$	Factor, ritmo de aprendizaje en el instante $t$
$\sigma$	Factor de ancho en funciones de base radial con núcleo esférico
$\mathbf{C}_i$	Factor de ancho (matriz de covarianzas) en funciones de base radial con núcleo hiperelipsoidal
$m(\mathbf{x})$	Factor de magnificación
$p(\mathbf{x})$	Función de densidad de probabilidad conjunta del vector $\mathbf{x}$
$p(y \mathbf{x})$	Densidad de probabilidad condicional de $y$ dado el vector $\mathbf{x}$
$q(\mathbf{x})$	Error de cuantificación del SOM para el vector $\mathbf{x}$
$\mathbf{f}(t)$	Vector de características en el instante $t$
$\mathbf{f}_g(t)$	Proyección de $\mathbf{f}(t)$ sobre la retícula
$\mathcal{S}(\mathbf{x} \mathbf{x}_i, \mathbf{y}_i)$	Estimador de Nadaraya-Watson de la interpolación del punto $\mathbf{x}$ dados puntos $\mathbf{x}_i$ e $\mathbf{y}_i$
$\Phi(\mathbf{x})$	Núcleo de base radial. Función RBF.
$\mathcal{E}(\mathbf{x})$	Error de modelado global del dato $\mathbf{x}$
$\vec{\mathcal{E}}(\mathbf{x})$	Error de modelado vectorial del dato $\mathbf{x}$
$\mathcal{S}_{\mathbf{g} \rightarrow 3D}(\cdot)$	Proyección 3D de elementos del espacio de retícula (visualización)
$\mathcal{S}_{\mathbf{m} \rightarrow 3D}(\cdot)$	Proyección 3D de elementos del espacio de entrada (espacio de datos)
$i_{ind}^{2x}, i_{ind}^{600}, \dots$	Armónicos dos veces la frecuencia de giro y de 600 Hz de la corriente de inducido (el resto de las variables siguen una nomenclatura análoga).

# Bibliografía

- [1] Peter J. Tavner and James Penman. *Condition Monitoring of Electrical Machines*. Research Studies Press Ltd., John Wiley and Sons Inc., 1987.
- [2] Erik Frisk. *Model-based fault diagnosis applied to an SI-Engine*. PhD thesis, Tekniska Hogskolan i Linkoping, Septiembre 1996.
- [3] Paul A. Sachs, Andy M. Paterson, and Michael H. M. Turner. Escort - an expert system for complex operations in real time. *Expert Systems*, 3(1):22–29, January 1986.
- [4] G. Fenu and T. Parisini. A note on nonparametric kernel smoothing for model-free fault symptom generation. *Automatica*, 35:1175–1179, 1999.
- [5] Timo Sorsa, Hieikki N. Koivo, and Hannu Kovisto. Neural networks in process fault diagnosis. *IEEE Transactions on systems man and cybernetics*, 21(4):815–849, july/august 1991.
- [6] C. Combastel, S. Gentil, and J.P. Rognon. Sensitivity of diagnostic residuals to modelling errors. In *Proc. of the IEEE Int. Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives IEEE, SDEMPED'99, Gijón*, pages 297–302, 1999.
- [7] E. Y. Chow and A. S. Willsky. Analytical redundancy and the design of robust failure detection filters. *IEEE Transactions on Automatic Control*, 7:603–614, 1984.
- [8] Mattias Nyberg. Parity functions as universal residual generators and tool for fault detectability analysis. Technical report, Vehicular Systems, ISY, Linkoping University, Septiembre 1997.

- 
- [9] Miguel A. Sanz-Bobi and Miguel A. Donaire. Diagnosis of electrical motors using artificial neural networks. In *Proc. of the IEEE Int. Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives IEEE, SDEMPED'99, Gijón*, pages 369–374, 1999.
- [10] T. Petsche, A. Marcantonio, C. Darken, S.J. Hanson, G.M. Kuhn, and I.Santoso. A neural network autoassociator for induction motor failure prediction. In D.S. Touretzky, M.C. Mozer, and M.E. Haselmo, editors, *Advances in Neural Information Processing Systems*, pages 924–930. Cambridge: MIT Press., 1996.
- [11] David J. H. Wilson and George W. Irwin. RBF principal manifolds for process monitoring. *IEEE Transactions on Neural Networks*, 10(6):1424–1434, November 1999.
- [12] M.F. Cabanas, M.G. Melero, J.M. Cano, F. Briz del Blanco, and G.A. Capolino. A new methodology for applying the FFT to induction motor on-line diagnosis. In *Proc. of the IEEE Int. Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives IEEE, SDEMPED'99, Gijón*, pages 537–543, 1999.
- [13] J. Penman and A. Stavrou. Broken rotor bars: their effect on the transient performance of induction machines. *IEE Proceedings on Electric Power Applications*, 143(6), november 1996.
- [14] R. Beguenane. Induction motors thermal monitoring by means of rotor resistance identification. *IEEE Transactions on Energy Conversion*, 14(3):566–570, September 1999.
- [15] K.R Cho, J.H. Lang, and S.D. Umans. Detection of broken rotor bars in induction motors using state and parameter estimation. *IEEE Transactions on Industry Applications*, 3:702–709, 1992.
- [16] S. Moreau, J.C. Trigeassou, G. Champenois, and J.P. Gaubert. Diagnosis of induction machines: A procedure for electrical fault detection and localization. In *IEEE International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives SDEMPED'99, Gijón*, pages 225–229, Gijón, 1999.

- 
- [17] Yumi Iwasaki. Real-world applications of qualitative reasoning. *IEEE Expert Intelligent Systems and their Applications*, pages 16–21, May 1997.
- [18] Tom Centrella. Ben Kuipers: Using qualitative reasoning. *IEEE Expert Intelligent Systems and their Applications*, pages 94–97, May 1997.
- [19] Daniel Louis Dvorak. *Monitoring and Diagnosis of Continuous Dynamic Systems Using Semiquantitative Simulation*. PhD thesis, University of Texas at Austin, 1992.
- [20] Marcos R. Vescovi, Marcelo M. Lamego, and Adam Farquhar. Modeling and simulation of a complex industrial process. *IEEE Expert Intelligent Systems and their Applications*, pages 42–46, May 1997.
- [21] Teuvo Kohonen, Erkki Oja, Olli Simula, Ari Visa, and Jari Kangas. Engineering applications of the self-organizing map. *Proceedings of the IEEE*, 84(10):1358–1384, october 1996.
- [22] Teuvo Kohonen. *Self-Organizing Maps*. Springer-Verlag, 1995.
- [23] Esa Alhoniemi, Jaakko Hollmén, Olli Simula, and Juha Vesanto. Process monitoring and modeling using the self-organizing map. *Integrated Computer Aided Engineering*, 6(1):3–14, 1999.
- [24] Josef Göppert and Wolfgang Rosenstiel. Topology-preserving interpolation in self-organizing maps. In *Proceedings of NeuroNimes 93, Nanterre, France*, October 1993.
- [25] Josef Göppert and Wolfgang Rosenstiel. The use of neural networks in the online analysis. Technical report.
- [26] Josef Göppert and Wolfgang Rosenstiel. Topological interpolation in som by affine transformations. In *Proc. of ESANN'95, Brussels, Belgium*, 1995.
- [27] Johan Fredrik Markus Svensen. *The Generative Topographic Mapping*. PhD thesis, Aston University, Abril 1998.

- [28] Ignacio Díaz Blanco, Alberto B. Díez González, Abel A. Cuadrado Vega, and José M. Enguita González. RBF approach for trajectory interpolation in self-organizing map based condition monitoring. In J.M. Fuertes, editor, *Proceedings of the 7th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA '99)*, pages 1003–1010, UPC, Barcelona, Catalonia, Spain, 1999.
- [29] Ignacio Díaz, Alberto B. Díez, Abel A. Cuadrado, Armandino Fombona, Faustino Obeso, and Juan A. González. Kernel regression interpolation in SOM-based process condition monitoring. In *DIAGNOS'2000*, La Habana, 2000.
- [30] Yong-Zai Lu. *INDUSTRIAL INTELLIGENT CONTROL: Fundamentals and Applications*. John Wiley and Sons, 1996.
- [31] Andrew Craig McCormick. *Cyclostationary and Higher-Order Statistical Signal Processing Algorithms for Machine Condition Monitoring*. PhD thesis, Department of Electronic and Electrical Engineering. University of Strathclyde, Mayo 1998.
- [32] Therman Eddie Green and John R. Snell. Thermographic inspection of hydraulic systems. *P/PM Technology*, october 1996.
- [33] Manés Fernández Cabanas. *Detección de Fallos Incipientes en Motores de Inducción en Funcionamiento mediante la aplicación de los Análisis Espectrales de Corrientes y Vibraciones*. PhD thesis, Departamento de Ingeniería Eléctrica, Electrónica de Computadores y Sistemas. Universidad de Oviedo, octubre 1995.
- [34] Mohammed S. Arefeen Hamid A. Toliyat and Alexander G. Parlos. A method for dynamic simulation of air-gap eccentricity in induction machines. *IEEE Transactions on Industry Applications*, pages 910–917, 1996.
- [35] Manuel García Melero. *Detección de fallos incipientes en el aislamiento del devanado estático de motores de inducción trifásicos mediante técnicas a máquina parada y en funcionamiento*. PhD thesis, Departamento de Ingeniería Eléctrica, Electrónica de Computadores y Sistemas. Universidad de Oviedo, 1998.

- 
- [36] R. A. Roberts and C. T. Mullis. *Digital signal processing*. Addison-Wesley, 1987.
- [37] John G. Proakis and Dimitris G. Manolakis. *Digital Signal Processing. Principles, Algorithms and Applications*. Prentice Hall, 1998.
- [38] Alan V. Oppenheim and Ronald W. Schaffer. *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [39] Karl J. Aström y B. Wittenmark. *Adaptive Control*. Addison-Wesley, 1989.
- [40] Lennart Ljung and Svante Gunnarson. Adaptation and tracking in system identification – a survey. *Automatica*, Vol. 26, No. 1, pages 7–21, 1990.
- [41] Alexej Barkov and Natalja Barkova. Condition assessment and life prediction of rolling element bearings-part 1. *Sound and Vibration*, pages 10–17, june 1995.
- [42] A. Murray and J. Penman. Extracting useful higher order features for condition monitoring using artificial neural networks. *IEEE Transactions on Signal Processing*, 45(11):2821–2828, November 1997.
- [43] Jerry M. Mendel. Tutorial on higher order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. *Proceedings of the IEEE*, pages 278–304, 1991.
- [44] Simon Haykin. *Adaptive Filter Theory*. Prentice–Hall International Editions, 1991.
- [45] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. Academic Press, 1999.
- [46] Carles M. Cuadras. *Métodos de Análisis Multivariante*. PPU, 1991.
- [47] Simon Haykin. *Neural Networks, a Comprehensive Foundation*. Prentice-Hall, Inc., 1999.
- [48] Bernhard Scholkopf. *Support Vector Learning*. PhD thesis, Technischen Universität Berlin, Septiembre 1997.

- 
- [49] Bernhard Scholkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.
- [50] Varios Autores. *Advances in Kernel Methods*. MIT Press, 1999.
- [51] Samuel Kaski. Data exploration using self-organizing maps. Technical report, Finnish Academy of Technology, 1997.
- [52] Pierre Demartines and Jeanny Herault. Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. *IEEE Transactions on Neural Networks*, 8(1):148–154, January 1997.
- [53] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [54] T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84:502–516, 1989.
- [55] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, september 1990.
- [56] H. U. Bauer, R. Der, and M. Herrmann. Controlling the magnification factor of self-organizing feature maps. *Neural Computation*, 8(4):757–71, 1996.
- [57] Paul Yee and Simon Haykin. A dynamic regularized radial basis function network for nonlinear nonstationary time series prediction. *IEEE Transactions on Signal Processing*, 47(9):2503–2521, September 1999.
- [58] Anthony Zaknich. Introduction to the modified probabilistic neural network for general signal processing applications. *IEEE Transactions on Signal Processing*, 46(7):1980–1990, July 1998.
- [59] Duane DeSieno. Adding a conscience to competitive learning. In *Proc. ICNN'88, Int. Conf. on Neural Networks*, pages 117–124, Piscataway, NJ, 1988. IEEE Service Center.
- [60] H. Ritter. Parametrized self-organizing maps. 1993.

- 
- [61] M. Tipping and C. Bishop. Mixtures of probabilistic principal component analyzers, 1997.
- [62] Shun-Ichi Amari. Mathematical foundations of neurocomputing. *Proceedings of the IEEE*, 78(9):1443–1463, September 1990.
- [63] J.-S.R. Jang, C.-T. Sun, and E. Mizutani. *Neuro-Fuzzy and Soft Computing. A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, 1997.
- [64] Chin-Teng Lin and C.S. George Lee. *Neural Fuzzy Systems. A Neuro-Fuzzy Synergism to Intelligent Systems*. Prentice Hall, 1996.
- [65] James Berry. Detection of multiple cracked rotor bars on induction motors using both vibration and motor current analysis. *P/PM Technology*, pages 48–59, 1996.
- [66] Alexej Barkov and Natalja Barkova. Condition assessment and life prediction of rolling element bearings-part 2. *Sound and Vibration*, pages 27–31, september 1995.
- [67] Robert M. Jones. Enveloping for bearing analysis. *Sound and Vibration*, pages 10–15, february 1996.
- [68] Gary Einarson and Dave Kelly. Vibration analysis: A study in bearing frequencies and harmonics. *P/PM Technology*, pages 42–44, October 1997.
- [69] Mirza S. Hussain. Vibration analysis of slow speed machines. *P/PM Technology*, pages 60–63, February 1997.
- [70] Toshio Fukuda and Takanori Shibata. Theory and applications of neural networks for industrial control systems. *IEEE Transactions on industrial electronics*, 39(6):472–489, Diciembre 1992.
- [71] Bernard Widrow. 30 years of adaptive neural networks: Perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, 78(9):1415–1440, september 1990.
- [72] D. G. Melvin and J. Penman. Fusing human knowledge with neural networks in machine condition monitoring systems. *Proceedings*

- of the SPIE—The International Society for Optical Engineering*, 2492(pt. 1):276–83, 1995.
- [73] Jukka Laine. Analysis and monitoring of continuous casting mould with the self-organizing map. *ECSC workshop Application of artificial neural network systems in the steel industry*, pages 151–155, 1998.
- [74] Varios Autores. The past, present, and future of neural networks for signal processing. *IEEE Signal Processing Magazine*, pages 28–48, November 1997.
- [75] J. Frigiére, L. Bertrand, M.P. Chouvet, M. Lallier, and R. Perisse. The state of the art of neural network applications in the iron and steel industry outside europe. *ECSC workshop Application of artificial neural network systems in the steel industry*, pages 53–62, 1998.
- [76] K. G. Narendra, V.K. Sood, K. Khorasani, and R. Patel. Application of a radial basis function (rbf) neural network for fault diagnosis in a hvdc system. *IEEE Transactions on Power Systems*, 13(1), February 1998.
- [77] Zhenyuan Wang and Paul J. Griffin. A combined ann and expert system tool for transformer fault diagnosis. *IEEE Transactions on Power Delivery*, 13(4), October 1998.
- [78] W. Xu, D. Wang, Z. Zhou, and H.Chen. Fault diagnosis of power transformers: application of fuzzy set theory, expert systems and artificial neural networks. *IEE Proc. Sci. Meas. Technology*, 144(1), January 1997.
- [79] Ben Krose and Patrick van der Smagt. An introduction to neural networks. Technical report, University of Amsterdam, 1996.
- [80] Robert Schaback. On the efficiency of interpolation by radial basis functions. In *Surface Fitting and Multiresolution Methods*, 1997.
- [81] A.J. Marques Cardoso, A.M.S. Mendes, and S.M.A. Cruz. The park’s vector approach: New developments in on-line fault diagnosis of electrical machines, power electronics and adjustable speed

- drives. In *Proc. of the IEEE Int. Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives IEEE, SDEM-PED'99, Gijón*, pages 89–97, 1999.
- [82] Tomaso Poggio and Federico Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78(9):1481–1497, September 1990.
- [83] Paul J. Werbos. Neurocontrol and elastic fuzzy logic: Capabilities, concepts, and applications. *IEEE Transactions on Industrial Electronics*, 40(2):170–180, April 1993.
- [84] Kumpati S. Narendra. Neural networks for control: Theory and practice. *Proceedings of the IEEE*, 84(10):1385–1406, October 1996.
- [85] A.J. Morris and E.B. Martin. Neural networks: Panacea or pragmatic solution. In *ECSC Workshop. Application of Artificial Neural Networks Systems in the Steel Industry*, pages 9–42, Brussels, January 1998.
- [86] J. Frigiére, L. Bertrand, M-P. Chouvet, M. Lallier, and R. Perisse. The state of the art of neural network applications in the iron and steel industry outside europe. In *ECSC Workshop. Application of Artificial Neural Network Systems in the Steel Industry*, pages 53–62, Brussels, January 1998.
- [87] Juan Antonio González Rodríguez. *Predicción de la Calidad de Bandas Laminadas en Caliente mediante Modelos Inteligentes*. PhD thesis, Universidad de Oviedo, Marzo 1999.
- [88] Lennart Ljung and Torkel Glad. *Modelling of Dynamic Systems*. Prentice Hall, 1994.
- [89] Marti A. Hearst. Support vector machines. *IEEE Intelligent Systems*, pages 18–28, 1998.
- [90] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.

- [91] Jörg Walter and Helge Ritter. Rapid learning with parameterized self-organizing maps. *Neurocomputing*, 12:131–153, 1996.
- [92] Jörg Walter. PSOM network: Learning with few examples. In *Proc. Int. Conf. On Robotics and Automation (ICRA) IEEE*, 1998.