



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA SUPERVISOR CON MODELOS BASADOS EN
SIMILITUD PARA LA DETECCIÓN Y AISLAMIENTO DE FALLAS EN TURBINA A GAS NATURAL**

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL
ELECTRICISTA**

SEBASTIÁN FELIPE FUENTEALBA TORO

**PROFESOR GUÍA
MARCOS ORCHARD CONCHA**

**MIEMBROS DE LA COMISIÓN:
DORIS SÁEZ HUEICHAPAN
HECTOR AGUSTO ALEGRÍA**

**SANTIAGO DE CHILE
JULIO 2012**

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL ELECTRICISTA
POR: SEBASTIÁN FUENTEALBA T.
FECHA: 03/07/2012
PROF. GUÍA: Sr. MARCOS ORCHARD CONCHA

“DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA SUPERVISOR CON MODELOS BASADOS EN
SIMILITUD PARA LA DETECCIÓN Y AISLAMIENTO DE FALLAS EN TURBINA A GAS
NATURAL”

Debido al constante aumento en los estándares de producción y calidad, así como las exigencias impuestas por estrictas normativas ambientales y de seguridad, existe en la industria moderna la necesidad de incrementar la eficiencia y complejidad de los sistemas de control asociados. Sin embargo, existen perturbaciones que no pueden ser eficazmente manejadas por los controladores a menos que exista una adecuada caracterización del modo de operación. He ahí la necesidad de incorporar sistemas de monitoreo para detección y aislamiento de anomalías que pudiesen condicionar el funcionamiento del proceso.

En este sentido, el objetivo general de este trabajo es diseñar e implementar un sistema supervisor en línea con modelos basados en similitud (SBM) que permita la detección y el aislamiento de anomalías en procesos industriales multivariados y en particular, en el caso de una turbina a gas natural. El diseño del sistema supervisor incluye una etapa preliminar fuera de línea y de tres etapas que operan en tiempo real. La etapa preliminar corresponde al análisis del proceso en estudio y de los modos de falla de interés, así como la generación de los modelos SBM que serán utilizados en la etapa de implementación en línea. A través de esta metodología de estudio de eventos se determinan las variables más relevantes en cada anomalía con el fin de optimizar el procedimiento de generación de modelos. Las etapas asociadas a la implementación en línea se han denominado “fases”. “Fase I” genera una señal de alarma si la salida estimada por un modelo clasificador SBM, diseñado en la etapa preliminar, supera un umbral determinado en la etapa preliminar. Si se produce una alarma, entonces se ejecuta la “Fase II” que determina el vector de residuos entre las salidas medidas y las salidas estimadas por el modelo SBM, para cada posible anomalía en gestación. “Fase III” calcula el indicador de Hotelling de cada vector de residuos. Se genera una alarma de anomalía si este indicador supera un umbral determinado en la etapa preliminar.

Este sistema es aplicado en una turbina a gas natural para el estudio de dos anomalías que ocasionan una disminución significativa en la potencia: (i) temperatura de un álabe fuera de rango (Anomalía #1) y, (ii) suciedad excesiva en el compresor (Anomalía #2). Los porcentajes de detección de las Anomalías #1 y #2 son 96% y 76%, respectivamente. Estos porcentajes se estiman en función del número de datos. Debido al tamaño de los conjuntos de datos disponibles, es imposible en esta aplicación utilizar probabilidades para caracterizar al sistema supervisor. Las principales ventajas de disponer de este sistema supervisor son: (1) con una metodología de estudio de eventos, se obtienen aquellas variables que son más relevantes en cada anomalía, (2) con la “Fase I” se logran descartar del análisis aquellas observaciones que representan un modo normal de operación y (3) a través de señales binarias “Normal” o “Alarma” se logra centralizar la información que entregan todos los modelos.

AGRADECIMIENTOS

A mis padres, Cecilia Toro y Manuel Fuentealba, y mi hermano Edgardo, porque sin ellos no habría llegado aquí, entregándome la confianza que necesité y enseñándome a que con esfuerzo se alcanzan los objetivos.

A Nathaly Romero (resumo porque o sino gastaría hojas y hojas) y a toda su familia, por su constante compañía, por su apoyo y por el cariño que me entregaron día a día.

A mis fieles amigos que siempre me recibían con sus ladridos y saltos: Dixie, Perseo y Darko.

A mi abuela por su la fuerza que demostró sobre todo estos últimos años. A mi abuelo por la energía que me mostró que hay que tener, hasta el último de sus días.

A mi profesor guía Marcos O. y a la profesora Doris S. por el tiempo, el conocimiento entregado y la ayuda brindada durante no solo este trabajo, sino también a lo largo de todos estos años en la Universidad.

Por todos esos gratos momentos vividos, a todos mis amigos que conocí a lo largo de mi carrera universitaria (en orden alfabético para que no hayan resentimientos): Agustín P., Alejandro A., Alexis A., Camilo L., César N., Cristian S., Daniel C., Eduardo P., Enrique G., Felipe U., Gabriela L., Javier P., Karen S., María José R., Matías M., Rocío R., Seba A.

A mis compañeros de proyecto y amigos Alejandro L. y Hugo N., que fueron parte fundamental para que todo esto resultara.

A todos los integrantes que componen el laboratorio de control automático que este último año fue como mi segunda casa.

ÍNDICE GENERAL

| | |
|--|----|
| CAPÍTULO 1: Introducción..... | 1 |
| 1.1 Antecedentes | 1 |
| 1.2 Objetivos | 2 |
| 1.2.1 Objetivo general | 2 |
| 1.2.2 Objetivos específicos..... | 3 |
| 1.3 Alcances | 3 |
| 1.4 Estructura del documento..... | 4 |
| CAPÍTULO 2: Revisión bibliográfica y estado del arte | 5 |
| 2.1 Sistemas de supervisión de procesos..... | 5 |
| 2.1.1 Conceptos básicos | 5 |
| 2.1.2 Descripción de un sistema supervisor | 7 |
| 2.2 Detección de fallas | 8 |
| 2.2.1 Minería de datos..... | 9 |
| 2.2.1.1 Métodos con comprobación de límites y tendencias | 9 |
| 2.2.1.2 Métodos basados en modelos de señales..... | 11 |
| 2.2.1.3 Métodos de análisis multivariado..... | 12 |
| 2.2.2 Métodos de modelos del proceso | 13 |
| 2.3 Aislamiento de fallas..... | 15 |
| 2.3.1 Métodos de clasificación..... | 16 |
| 2.3.2 Métodos de inferencia | 19 |
| 2.4 Detección y aislamiento de fallas (FDI) utilizando método de los residuos | 21 |
| 2.5 Herramientas utilizadas en sistemas FDI | 22 |
| 2.5.1 Error cuadrático medio..... | 22 |

| | | |
|--|--|----|
| 2.5.2 | Filtro de mediana móvil | 22 |
| 2.5.3 | Agrupación por k-medias (k-means clustering) | 23 |
| 2.5.4 | Análisis de componentes principales (PCA) | 24 |
| 2.5.5 | Test de estadístico de Hotelling..... | 26 |
| 2.5.6 | Modelos basados en similitud (SBM) | 27 |
| CAPÍTULO 3: Herramientas de generación de modelos basados en similitud | | 29 |
| 3.1 | Herramienta de generación de modelos de anomalías basados en similitud..... | 30 |
| 3.1.1 | Etapa 1: análisis inicial..... | 32 |
| 3.1.2 | Etapa 2: modelo preliminar | 35 |
| 3.1.3 | Etapa 3: fase iterativa de optimización..... | 35 |
| 3.1.4 | Etapa 4: reajuste del modelo | 36 |
| 3.2 | Modelo SBM clasificador | 38 |
| 3.2.1 | Etapa 1: salidas del conjunto de entrenamiento | 40 |
| 3.2.2 | Etapa 2: generación del modelo SBM..... | 41 |
| 3.2.3 | Etapa 3: determinación del umbral de decisión..... | 41 |
| CAPÍTULO 4: Implementación de sistema supervisor FDI utilizando modelos basados en similitud | | 43 |
| 4.1 | Esquema general del sistema supervisor para estudio de eventos..... | 43 |
| 4.2 | Metodología del estudio de eventos | 45 |
| 4.2.1 | Etapa 1: recopilación de datos..... | 47 |
| 4.2.2 | Etapa 2: análisis preliminar de variables..... | 49 |
| 4.2.3 | Etapa 3: creación de modelos..... | 50 |
| 4.2.4 | Etapa 4A: validación del modelo | 50 |
| 4.2.5 | Etapa 4B: prueba de eficacia del modelo en la detección de la anomalía | 50 |
| 4.2.6 | Etapa 5: reajuste de parámetros..... | 51 |
| 4.3 | Sistema supervisor en línea | 51 |
| 4.3.1 | Fase I: modelo SBM para detección de estados anormales..... | 52 |

| | | |
|--|--|-----|
| 4.3.1.1 | Salidas del modelo SBM clasificador | 53 |
| 4.3.1.2 | Determinación del estado del sistema: normal o alerta | 54 |
| 4.3.2 | Fase II: modelos del proceso para obtención de residuos | 55 |
| 4.3.3 | Fase III: análisis de residuos para determinación de falla | 55 |
| CAPÍTULO 5: Aplicación del sistema supervisor en línea en una turbina a gas..... | | 57 |
| 5.1 | Descripción del proceso industrial | 57 |
| 5.1.1 | Turbina a gas de la central térmica de ciclo combinado | 59 |
| 5.1.2 | Base de datos | 61 |
| 5.2 | Generación de modelos de anomalías registradas | 63 |
| 5.2.1 | Anomalía #1: temperatura excedida en álabe N°20..... | 64 |
| 5.2.2 | Anomalía #2: suciedad del compresor | 74 |
| 5.3 | Modelo clasificador..... | 80 |
| 5.4 | Sistema supervisor..... | 91 |
| 5.4.1 | Reajustes en los modelos del sistema supervisor | 91 |
| 5.4.2 | Resultados en la base de datos A..... | 97 |
| 5.4.3 | Resultados en la base de datos B..... | 102 |
| CAPÍTULO 6: Conclusiones | | 108 |
| Referencias..... | | 111 |

ÍNDICE DE TABLAS

| | |
|---|-----|
| Tabla 3.1: Argumentos de la función $ModeloSBM(\cdot)$ | 31 |
| Tabla 3.2: Porcentajes de observaciones seleccionadas en el ejemplo de la Figura 3.2b. | 33 |
| Tabla 3.3: Argumentos de la función $OfflineClass(\cdot)$ | 39 |
| Tabla 4.1: Argumentos de la función $systemcondition(\cdot)$ | 52 |
| Tabla 5.1: Bases de datos utilizadas para aplicar el sistema supervisor..... | 61 |
| Tabla 5.2: Variables de entrada y salida del modelo preliminar. | 66 |
| Tabla 5.3: Tabla de contingencia del modelo preliminar en los conjuntos de entrenamiento, validación y prueba en la anomalía..... | 70 |
| Tabla 5.4: Variables de entrada y salida del modelo M_1 | 71 |
| Tabla 5.5: Tabla de contingencia del modelo M_1 en los conjuntos de entrenamiento, validación y prueba en la anomalía. | 73 |
| Tabla 5.6: Error cuadrático medio en los modelos preliminar y M_1 en los conjuntos de entrenamiento y validación..... | 74 |
| Tabla 5.7: Variables de entrada y salida del modelo M_2 | 76 |
| Tabla 5.8: Tabla de contingencia del modelo M_2 en los conjuntos de entrenamiento, validación y prueba en la anomalía | 80 |
| Tabla 5.9: Variables de entrada del modelo clasificador. | 82 |
| Tabla 5.10: Tabla de contingencia de la Anomalía #1 en los conjuntos de entrenamiento y prueba de la base de datos A..... | 90 |
| Tabla 5.11: Tabla de contingencia de la Anomalía #2 en los conjuntos de entrenamiento y prueba de la base de datos B..... | 90 |
| Tabla 5.12: Tabla de estados del sistema supervisor..... | 96 |
| Tabla 5.13: Significado de las señales de alarma en cada modelo..... | 96 |
| Tabla 5.14: Tabla de contingencia del modelo clasificador en la base de datos A. | 99 |
| Tabla 5.15: Tabla de contingencia del modelo M_1 en la base de datos A..... | 100 |
| Tabla 5.16: Tabla de contingencia del modelo M_2 en la base de datos A..... | 100 |
| Tabla 5.17: Tabla de contingencia del modelo M_3 en la base de datos A..... | 101 |

| | |
|--|-----|
| Tabla 5.18: Porcentajes de detección de la señal de estado en la base de datos A..... | 101 |
| Tabla 5.19: Tabla de contingencia del modelo clasificador en la base de datos B. | 104 |
| Tabla 5.20: Tabla de contingencia del modelo M_1 en la base de datos B. | 105 |
| Tabla 5.21: Tabla de contingencia del modelo M_2 en la base de datos B. | 105 |
| Tabla 5.22: Tabla de contingencia del modelo M_3 en la base de datos B. | 106 |
| Tabla 5.23: Porcentajes de detección de la señal de estado en la base de datos B. | 106 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 2.1: Esquema general de diferentes métodos de supervisión con gestión de fallas [32]. | 7 |
| Figura 2.2: Métodos de detección de fallas con señales individuales. | 8 |
| Figura 2.3: (a) Método de comprobación de límites y (b) método de comprobación de tendencias [33]. | 9 |
| Figura 2.4: Densidad de probabilidad normal de la variable observada Y , μ_0 y σ_0 caracterizan el estado nominal, μ_1 y σ_1 el estado de falla del sistema [33]. | 10 |
| Figura 2.5: (a) Variable estocástica $Y(t)$ con función de densidad de probabilidad $p(Y)$ en estado normal, y $p'(Y)$ en estado de falla; (b) Variable estocástica $Y(t)$ como conjunto difuso $\mu_n(Y)$ en estado normal, y $\mu'(Y)$ en estado de falla [33]. | 11 |
| Figura 2.6: Esquema de los métodos de detección de falla con modelos de señales [32]. | 12 |
| Figura 2.7: Esquema de los métodos de detección de falla con modelos del proceso [32]. | 13 |
| Figura 2.8: Aislamiento de fallas por métodos de clasificación (a) y por métodos de inferencia (b) [32]. | 15 |
| Figura 2.9: Métodos de aislamiento de fallas. | 16 |
| Figura 2.10: Método de árboles de decisión para aislamiento de fallas [34]. | 17 |
| Figura 2.11: Ejemplo del clasificador geométrico <i>k-vecinos más cercanos</i> (k-nearest neighbors) [28]. | 18 |
| Figura 2.12: Método de razonamiento binario para aislamiento de fallas [34]. | 19 |
| Figura 2.13: Método híbrido neuro-difuso para aislamiento de fallas [34]. | 20 |
| Figura 2.14: Esquema de un sistema FDI utilizando método de los residuos [40]. | 21 |
| Figura 2.15: Señal de ejemplo X y su mediana móvil X_{MM} de largo 2. | 23 |
| Figura 2.16: Pasos del algoritmo de agrupación por k-medias, considerando $k = 3$. Los centros de los grupos están representados por círculos, y los datos por cuadrados [53]. | 24 |
| Figura 3.1: Etapas de la herramienta de generación de modelos basados en similitud. | 32 |
| Figura 3.2: (a) Datos en el espacio original y (b) y en el espacio de componentes principales. | 33 |
| Figura 3.3: Selección de observaciones por medio del indicador de Hotelling. | 34 |
| Figura 3.4: Etapa 3 de optimización del conjunto de observaciones. | 36 |
| Figura 3.5: Selección de observaciones en la etapa de reajuste del modelo. | 37 |
| Figura 3.6: Etapas para la generación del Modelo SBM Clasificador. | 39 |

| | |
|--|----|
| Figura 3.7: Histogramas para determinar el umbral de decisión <i>Classthreshold</i> | 42 |
| Figura 4.1: Esquema del sistema supervisor FDI..... | 44 |
| Figura 4.2: Tipos de modelos de anomalías registradas; (a) modelo del proceso en operación normal; (b) modelo del proceso operando en la anomalía. | 46 |
| Figura 4.3: Metodología para la generación de modelos de anomalías registradas. | 47 |
| Figura 4.4: Ejemplo de una base de datos utilizada para generar modelos..... | 48 |
| Figura 4.5: Obtención de <i>XoutClass</i> en base a la distancia a los centros de cada agrupación. | 54 |
| Figura 4.6: Modelos de anomalías registradas, con sus respectivas entradas, salidas y cálculo de residuos. | 55 |
| Figura 5.1: Central térmica de ciclo combinado San Isidro [51] [52]..... | 58 |
| Figura 5.2: Central térmica de ciclo combinado [52]..... | 59 |
| Figura 5.3: Esquema de una central térmica de ciclo combinado [54]. | 60 |
| Figura 5.4: Potencia activa en la base de datos A. | 62 |
| Figura 5.5: Porcentaje del flujo de combustible (diesel y gas) en la base de datos A..... | 62 |
| Figura 5.6: Potencia activa en la base de datos B. | 63 |
| Figura 5.7: Porcentaje del flujo de combustible (diesel y gas) en la base de datos B..... | 63 |
| Figura 5.8: Potencia activa de la base de datos A con los instantes en que ocurre la Anomalía #1..... | 65 |
| Figura 5.9: Potencia activa de la base de datos A con los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo). | 65 |
| Figura 5.10: Variable de salida “ <i>Max. delta temp. gases álabes</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar. | 67 |
| Figura 5.11: Variable de salida “ <i>Vibración XI</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar..... | 67 |
| Figura 5.12: Variable de salida “ <i>Vibración X2</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar..... | 68 |
| Figura 5.13: Variable de salida “ <i>Vibración Y1</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar..... | 68 |
| Figura 5.14: Variable de salida “ <i>Vibración Y2</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar..... | 68 |

| | |
|--|----|
| Figura 5.15: Variable de salida “ <i>Señal de control 1</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar. | 69 |
| Figura 5.16: Variable de salida “ <i>Señal de control 2</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar. | 69 |
| Figura 5.17: Potencia activa en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo). | 69 |
| Figura 5.18: Indicador de Hotelling de los residuos con el modelo preliminar en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo). | 70 |
| Figura 5.19: Variable de entrada “ <i>Señal de control</i> ” del modelo M_1 | 71 |
| Figura 5.20: Variable de entrada “ <i>Potencia activa</i> ” del modelo M_1 | 71 |
| Figura 5.21: Variable de entrada “ <i>Temp. prom. gases álabes</i> ” del modelo M_1 | 72 |
| Figura 5.22: Variable de salida “ <i>Max. delta temp. gases álabes</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_1 | 72 |
| Figura 5.23: Variable de salida “ <i>Vibración X2</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_1 | 72 |
| Figura 5.24: Variable de salida “ <i>Vibración Y2</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_1 | 73 |
| Figura 5.25: Indicador de Hotelling de los residuos con el modelo M_1 en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo). | 73 |
| Figura 5.26: Potencia activa de la base de datos B con los instantes en que ocurre la Anomalía #2. | 75 |
| Figura 5.27: Potencia activa de la base de datos B con los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo). | 76 |
| Figura 5.28: Variable de entrada “ <i>Flujo de combustible</i> ” del modelo M_2 | 77 |
| Figura 5.29: Variable de entrada “ <i>Señal de control</i> ” del modelo M_2 | 77 |
| Figura 5.30: Variable de entrada “ <i>Temperatura 1</i> ” del modelo M_2 | 77 |
| Figura 5.31: Variable de entrada “ <i>Posición Válvula 1</i> ” del modelo M_2 | 77 |
| Figura 5.32: Variable de entrada “ <i>Posición Válvula 2</i> ” del modelo M_2 | 78 |
| Figura 5.33: Variable de salida “ <i>Presión</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_2 | 78 |
| Figura 5.34: Variable de salida “ <i>Potencia activa</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_2 | 78 |

| | |
|---|----|
| Figura 5.35: Variable de salida “ <i>Temperatura 2</i> ” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_2 | 79 |
| Figura 5.36: Potencia activa en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo)..... | 79 |
| Figura 5.37: Indicador de Hotelling de los residuos con el modelo M_2 en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo). | 79 |
| Figura 5.38: Potencia activa de la base de datos A con los conjuntos de entrenamiento (azul) y prueba en la anomalía (rojo) | 81 |
| Figura 5.39: Potencia activa de la base de datos B con los conjuntos de entrenamiento (azul) y prueba en la anomalía (rojo). | 81 |
| Figura 5.40: Las tres agrupaciones y centros (negro) en el conjunto de entrenamiento. | 82 |
| Figura 5.41: Observaciones seleccionadas en cada agrupación para obtener las matrices D_i y D_o | 83 |
| Figura 5.42: Variable de salida “ <i>Agrupación 1</i> ” en el conjunto de entrenamiento (negro), salida estimada con SBM (azul) y umbral de decisión (verde). | 84 |
| Figura 5.43: Variable de salida “ <i>Agrupación 2</i> ” en el conjunto de entrenamiento (negro), salida estimada con SBM (azul) y umbral de decisión (verde). | 84 |
| Figura 5.44: Variable de salida “ <i>Agrupación 3</i> ” en el conjunto de entrenamiento (negro), salida estimada con SBM (azul) y umbral de decisión (verde). | 84 |
| Figura 5.45: Potencia activa en el conjunto de entrenamiento. | 85 |
| Figura 5.46: Señal de alarma en el conjunto de entrenamiento..... | 85 |
| Figura 5.47: Variable de salida “ <i>Agrupación 1</i> ” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde). | 86 |
| Figura 5.48: Variable de salida “ <i>Agrupación 2</i> ” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde). | 86 |
| Figura 5.49: Variable de salida “ <i>Agrupación 3</i> ” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde). | 87 |
| Figura 5.50: Potencia activa en la base de datos A con los conjuntos de entrenamiento (azul) y de prueba en la anomalía (rojo). | 87 |
| Figura 5.51: Señal de alarma en la base de datos A con los conjuntos de entrenamiento (azul) y de prueba en la anomalía (rojo). | 87 |
| Figura 5.52: Variable de salida “ <i>Agrupación 1</i> ” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde). | 88 |

| | |
|--|-----|
| Figura 5.53: Variable de salida “Agrupación 2” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde). | 88 |
| Figura 5.54: Variable de salida “Agrupación 3” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde). | 89 |
| Figura 5.55: Potencia activa en la base de datos B con los conjuntos de entrenamiento (azul) y de prueba en la anomalía (rojo). | 89 |
| Figura 5.56: Señal de alarma en la base de datos B con los conjuntos de entrenamiento (azul) y de prueba en la anomalía (rojo). | 89 |
| Figura 5.57: Potencia activa en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) en base de datos A. | 92 |
| Figura 5.58: Indicador de Hotelling de los residuos con el modelo M_2 en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) en base de datos A (Anomalía #1). | 92 |
| Figura 5.59: Potencia activa en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) en base de datos B. | 92 |
| Figura 5.60: Indicador de Hotelling de los residuos con el modelo M_1 en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) en base de datos B (Anomalía #2). | 93 |
| Figura 5.61: Potencia activa en los conjuntos de entrenamiento (azul) y validación (verde) en la base de datos B para el modelo M_3 | 94 |
| Figura 5.62: Indicador de Hotelling de los residuos con el modelo M_1 en los conjuntos de entrenamiento (azul), validación (verde) y prueba (rojo). | 94 |
| Figura 5.63: Indicador de Hotelling de los residuos con el modelo M_2 en los conjuntos de entrenamiento (azul), validación (verde) y prueba (rojo). | 94 |
| Figura 5.64: Indicador de Hotelling de los residuos con el modelo M_3 en los conjuntos de entrenamiento (azul), validación (verde) y prueba (rojo). | 95 |
| Figura 5.65: Potencia activa en la base de datos A en presencia de la Anomalía #1. | 97 |
| Figura 5.66: Señal de alarma del modelo clasificador en la base de datos A. | 97 |
| Figura 5.67: Señal de alarma del modelo M_1 en la base de datos A. | 97 |
| Figura 5.68: Señal de alarma del modelo M_2 en la base de datos A. | 98 |
| Figura 5.69: Señal de alarma del modelo M_3 en la base de datos A. | 98 |
| Figura 5.70: Señal de estado del proceso en la base de datos A. | 98 |
| Figura 5.71: Potencia activa en la base de datos B en presencia de la Anomalía #2. | 102 |
| Figura 5.72: Señal de alarma del modelo clasificador en la base de datos B. | 102 |

| | |
|---|-----|
| Figura 5.73: Señal de alarma del modelo M_1 en la base de datos B..... | 102 |
| Figura 5.74: Señal de alarma del modelo M_2 en la base de datos B. | 103 |
| Figura 5.75: Señal de alarma del modelo M_3 en la base de datos B. | 103 |
| Figura 5.76: Señal de estado del proceso en la base de datos B (observaciones 1-9567). | 103 |
| Figura 5.77: Señal de estado del proceso en la base de datos B (observaciones 9568-19134). | 103 |

ABREVIACIONES

| | |
|---------|--|
| ANN | Redes neuronales artificiales (del inglés: Artificial neural network). |
| ARMA | Modelo autorregresivo de media móvil (del inglés: Autoregressive-moving average). |
| ECM/MSE | Error cuadrático medio (del inglés: Mean square error). |
| FDA | Análisis discriminante de Fisher (del inglés: Fisher's discriminant analysis). |
| FDI | Detección y aislamiento de fallas (del inglés: Fault detection and isolation). |
| FFT | Transformación rápida de Fourier (del inglés: Fast Fourier transformation). |
| ICA | Análisis de componentes independientes (del inglés: Independent component analysis). |
| k-NN | k - vecinos más cercanos (del inglés: k-nearest neighbor). |
| LDA | Análisis de discriminante lineal (del inglés: Linear discriminant analysis). |
| PCA | Análisis de componentes principales (del inglés: Principal component analysis). |
| PID | Controlador proporcional, integral y derivativo. |
| PLS | Mínimos cuadrados parciales (del inglés: Partial least squares). |
| SBM | Modelación basada en similitud (del inglés: Similarity based modeling). |

CAPÍTULO 1: INTRODUCCIÓN

1.1 ANTECEDENTES

Debido al creciente aumento en las demandas de productividad y calidad de los productos, y a la necesidad de satisfacer las normativas ambientales y de seguridad, la industria moderna ha requerido aumentar el número y la complejidad de los sistemas de control asociados. Los controladores utilizados comúnmente (PID, controladores predictivos) son capaces de mantener el sistema en los puntos de operación deseados, aún frente a algunos tipos de perturbaciones o cambios en el proceso. Sin embargo, hay cambios o desviaciones que los controladores no pueden manejar. Entre estos, se encuentran cambios en los parámetros del proceso, problemas en los actuadores y problemas en los sensores [19].

Para asegurar que el proceso tenga un rendimiento acorde con las necesidades y mantener el sistema en un rango de seguridad apropiado, es necesario diagnosticar, a través de la detección y el aislamiento, estas desviaciones no permitidas. Estas tareas se encuentran asociadas con la supervisión del proceso. La información que entrega la supervisión no solo asegura que los operadores y el personal de mantención se encuentren mejor informados del estado del sistema, sino también los guía a realizar las acciones apropiadas para eliminar el comportamiento anormal del proceso.

Tal como se señala en [32], la supervisión de un proceso cuenta con diferentes etapas: detección, aislamiento, identificación y recuperación del proceso o intervención. La etapa de detección muestra, en base a ciertos criterios, si el sistema se encuentra fuera de los rangos normales de operación y por lo tanto, hay presencia de una anomalía. El objetivo en la etapa de aislamiento es encontrar el origen de la anomalía, aislando las componentes del proceso cuando el sistema no está en su operación nominal o normal. A continuación, con la identificación se desea estimar el tamaño, tipo y/o naturaleza de la anomalía. En base a la información obtenida en las etapas anteriores, en la etapa de recuperación o intervención los operadores del proceso ejecutan las acciones necesarias para eliminar, de manera total o parcial, la anomalía. Este trabajo se centra en los sistemas de supervisión para detección y aislamiento de fallas (fault detection and isolation o FDI), esto es, detección y aislamiento de anomalías.

La literatura muestra que los trabajos relativos a la detección de anomalías utilizan variados métodos basados en la minería de datos [5], [13], [18] o en la generación de modelos del proceso [9], [25], [39]. De entre los tipos de modelos del proceso, los modelos no paramétricos, como las redes neuronales artificiales (ANN) [3], [6] o los sistemas expertos difusos [37], tienen ventaja sobre los paramétricos cuando se trata de procesos con una gran cantidad de variables debido a que no requieren de un conocimiento sobre la estructura del proceso a modelar. Un modelo basado en similitud o SBM es un tipo de modelación no paramétrica que realiza estimaciones de las variables de salida utilizando los datos más representativos del proceso. Es una técnica de modelación más reciente que las otras. Sin embargo, existen trabajos donde se utilizan los modelos SBM en procesos multivariados con la finalidad de detectar anomalías [11], [40]. Con respecto al aislamiento de anomalías, los métodos se enfocan en la utilización de clasificadores [22], [38] o en técnicas de inferencia como los sistemas expertos difusos [8], [42].

Como ya fue mencionado, los modelos basados en similitud se han aplicado para detectar anomalías. Sin embargo, no se han encontrado referencias bibliográficas que indiquen su utilización en sistemas supervisores con la finalidad de detectar y aislar anomalías. He ahí la necesidad de no solo implementar algoritmos y metodologías para alcanzar este objetivo, sino también utilizarlos en una aplicación con datos reales. Así, se pretende implementar el sistema supervisor en un proceso que puede estar constantemente susceptible a anomalías, y que es de vital importancia prevenirlas lo más pronto posible, como lo son las turbinas a gas de una central termoeléctrica de ciclo combinado.

1.2 OBJETIVOS

1.2.1 OBJETIVO GENERAL

El objetivo general de este trabajo es el diseño e implementación de un sistema supervisor con modelos basados en similitud para detectar y aislar fallas en procesos industriales multivariados, en particular, en una turbina a gas natural.

1.2.2 OBJETIVOS ESPECÍFICOS

Se han planteado además los siguientes objetivos específicos:

- Diseñar e implementar la estructura del sistema supervisor que será empleado para detectar y aislar fallas en procesos industriales.
- Perfeccionar las herramientas de generación de los modelos que serán utilizados por el sistema supervisor.
- Diseñar una metodología de estudio de eventos para encontrar aquellas variables más relevantes en el proceso y así obtener los modelos basados en similitud a través de las herramientas creadas.
- Aplicar la metodología de estudio de eventos para generar modelos que detecten las anomalías que han sido registradas en las bases de datos de una turbina a gas.
- Aplicar el sistema supervisor en línea para detectar y aislar las anomalías que han sido registradas en la turbina a gas.

1.3 ALCANCES

El presente trabajo se enmarca en el proyecto de la empresa CONTAC INGENIERIOS LTDA. en colaboración con la UNIVERSIDAD DE CHILE y ENDESA CHILE. Este proyecto consiste en el diseño de herramientas de detección de anomalías en línea, en procesos industriales utilizando modelos basados en similitud (SBM). Esta herramienta, debe tener la facultad de ser utilizada en cualquier proceso industrial. En este proyecto en particular, se ha aplicado en una turbina a gas de la central termoeléctrica de ciclo combinado San Isidro I, perteneciente a ENDESA CHILE. Este es el segundo trabajo que ha sido desarrollado para este proyecto. En [1] se describe en detalle la herramienta de generación de modelos para detección de anomalías, y luego se utiliza para detectar la suciedad del compresor en la turbina a gas antes mencionada. Este trabajo prosigue con el proyecto a través del diseño de un sistema supervisor que sea capaz de detectar múltiples anomalías, y además pueda aislarlas. El proyecto se encuentra actualmente en su etapa de implementación en línea.

Además, existen restricciones de confidencialidad en los datos adquiridos y en las herramientas diseñadas. De este modo, se exponen las etapas de diseño, las metodologías utilizadas y los resultados obtenidos, pero no se entregan los algoritmos generados ni los nombres de las variables en forma detallada.

Para el desarrollo de este trabajo se requiere de un computador que disponga MATLAB para el diseño de los algoritmos y las simulaciones. Además, ENDESA CHILE facilita las bases de datos utilizadas y contribuye de forma importante a la definición de la estructura del sistema supervisor a través del conocimiento de sus expertos en la turbina a gas de la central de ciclo combinado San Isidro.

1.4 ESTRUCTURA DEL DOCUMENTO

La estructura del documento es la siguiente. El Capítulo 2 contiene el marco teórico necesario para un correcto entendimiento del tema, así como también los trabajos que han sido realizados en esta área.

El Capítulo 3 describe los algoritmos para generar los modelos basados en similitud que serán utilizados por el sistema supervisor para detectar y aislar anomalías. Las herramientas de generación de modelos, requieren de una selección adecuada de variables de entrada y salida. La metodología para obtener estos conjuntos de variables se describe en el Capítulo 4. Además, en este capítulo se describen las tres fases del sistema supervisor, que contempla la detección de múltiples anomalías a través de un modelo clasificador, y luego la aislación de estas para determinar su origen.

En el Capítulo 5 se realiza una descripción del proceso industrial en estudio (turbina a gas) y de las bases de datos que se utilizan en el sistema supervisor. A continuación se muestran los resultados del sistema supervisor en el estudio de dos eventos (la temperatura de un álabe de la cámara de combustión excede el límite de seguridad y un aumento en la suciedad del compresor), generando para dicho efecto los modelos con las herramientas descritas en el Capítulo 3 y utilizando la metodología expuesta en el Capítulo 4.

En base a los resultados obtenidos en los capítulos anteriores, en el Capítulo 6 se presentan las conclusiones de este estudio y el trabajo futuro a realizar.

CAPÍTULO 2: REVISIÓN BIBLIOGRÁFICA Y ESTADO DEL ARTE

El presente capítulo se enfoca en la descripción del estado del arte para la problemática descrita en la Sección 1.1. En la Sección 2.1 se encuentran los conceptos generales y básicos que se relacionan comúnmente con los sistemas de supervisión de procesos. En la Sección 2.2 y Sección 2.3 se describen en términos generales los variados métodos utilizados para detectar y aislar anomalías, respectivamente. El sistema de supervisión diseñado en este trabajo se basa en la utilización de los residuos de modelos. Esta metodología se describe en la Sección 2.4. Finalmente, la Sección 2.5 contiene las técnicas y algoritmos utilizados en el Capítulo 3 y Capítulo 4 para diseñar e implementar las herramientas de generación de modelos y el sistema supervisor.

2.1 SISTEMAS DE SUPERVISIÓN DE PROCESOS

Los sistemas de supervisión tienen como objetivo determinar la condición en la que se encuentra un proceso y posteriormente ejecutar las acciones apropiadas para evitar daños o accidentes en caso que exista alguna condición anómala. Para entender apropiadamente lo referente a los sistemas de supervisión y al monitoreo de procesos, en la Sección 2.1.1 se introducen algunos conceptos básicos y en la Sección 2.1.2 se realiza una descripción general de los sistemas de supervisión.

2.1.1 CONCEPTOS BÁSICOS

Los conceptos descritos a continuación han sido obtenidos desde [32], [36] y [41]. Estos conceptos corresponden a la terminología que históricamente se ha utilizado. Sin embargo, es necesario recalcar que otras fuentes como [19] cambian algunas de las definiciones aquí descritas.

- Modos de operación: son aquellas características que definen la dinámica de un sistema. Un modo de operación normal o nominal es cuando las características son las deseadas, y un modo de operación en falla es cuando las características involucran comportamientos inesperados.

- Síntomas (*symptom*): cambios inusuales, en comparación a un comportamiento aceptable o nominal, de las características o parámetros de un sistema. Los efectos observados en una falla son síntomas.
- Anomalía: datos del proceso que escapan del modo de operación normal. Las observaciones discordantes, valores atípicos, *outliers* son ejemplos de anomalías.
- Falla (*fault*): desviaciones no permitidas, en comparación a un comportamiento aceptable o nominal, en al menos una característica o parámetro del sistema.
- Evento crítico (*failure*): interrupción en las funcionalidades del sistema, bajo ciertas condiciones de operación específicas.
- Perturbación (*disturbance*): entrada desconocida y descontrolada que actúa en el sistema, desviándolo de su estado actual.
- Diagnóstico de fallas: esquema de supervisión cuyo objetivo es determinar el tipo de falla en un sistema con el mayor detalle posible: detectarla tan pronto como ocurre (detección de fallas), encontrar su origen aislando las componentes del sistema cuando el sistema no está en su operación nominal (aislamiento de fallas) y finalmente estimar el tamaño y tipo o naturaleza de la falla (identificación de fallas).
- Detección y aislamiento de fallas (*fault detection and isolation*): esquema de supervisión que involucra detectar la falla, y luego encontrar el origen.
- Detección de fallas (*fault detection*): tiene por objetivo encontrar patrones en el sistema que indiquen que este no está en su operación normal, tan pronto como sea posible.
- Aislamiento de fallas (*fault isolation*): tiene por objetivo encontrar la causa de la falla detectada, aislando las componentes del sistema cuando este no está en su operación normal o nominal.
- Identificación de fallas (*fault identification*): tiene por objetivo estimar el tamaño y el tipo de la falla encontrada.

- Residuos: diferencia entre las salidas estimadas obtenidas a través de un modelo del proceso y las salidas con datos reales. Las características de los residuos determinan si el proceso se encuentra en falla o no.

2.1.2 DESCRIPCIÓN DE UN SISTEMA SUPERVISOR

Como se ha dicho anteriormente, los sistemas supervisores de procesos tienen el objetivo de determinar el estado actual de un sistema encontrando estados indeseados o no permitidos, y además ejecutar las acciones apropiadas para evitar daños o accidentes como resultado de fallas o eventos críticos, que perjudiquen a corto o largo plazo los niveles de producción, la seguridad o el estado de las máquinas. En el mundo académico e investigativo no hay consenso con respecto a las definiciones que involucran un sistema supervisor. A continuación en la Figura 2.1 se muestra el esquema general obtenido desde [32] con los diferentes métodos de supervisión con gestión de fallas.

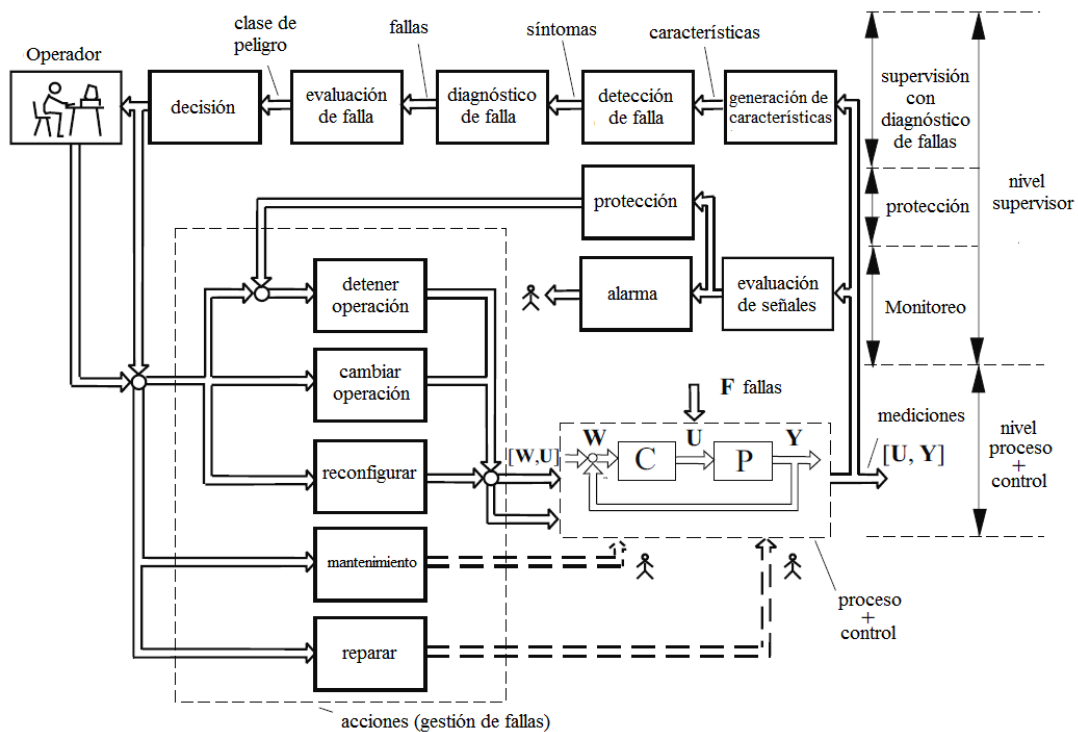


Figura 2.1: Esquema general de diferentes métodos de supervisión con gestión de fallas [32].

Este esquema general involucra tres niveles: monitoreo, protección automática y supervisión con diagnóstico de fallas. El monitoreo contempla verificar que las variables medidas cumplan con una cierta tolerancia. Cuando esto no ocurre, entonces se generan alarmas que alertan al operador, el cual deberá tomar las acciones pertinentes al caso. Si adicionalmente se detectan estados en el proceso que se encuentran por sobre un umbral y por lo tanto, son típicamente peligrosos, el sistema automáticamente ingresa en un modo a prueba de fallos con el objetivo de proteger los equipos. Esto corresponde al segundo nivel, protección automática. Ejemplos de fallas en los cuales se aplica la protección automática son un corto-circuito en un cable eléctrico o el sobrecalentamiento en un motor eléctrico. Finalmente, el tercer nivel corresponde a la supervisión con diagnóstico de fallas. Las tareas que involucra son la generación de características con los objetivos de detectar, diagnosticar fallas, y finalmente realizar las acciones a través de un operador o de un sistema de administración de fallas para eliminar este evento.

2.2 DETECCIÓN DE FALLAS

El principal objetivo de la detección es verificar si una falla ha ocurrido para así proveer esta información a un sistema supervisor o a un operador y realizar acciones apropiadas para evitar daños serios en el sistema, tan pronto como sea posible. Los métodos de detección de falla se clasifican en aquellos que utilizan minería de datos (data mining) y los que realizan detección a través de modelos del proceso, tal como se observa en la Figura 2.2. En esta sección, solo se introducen estos métodos basándose en la información obtenida en [33].

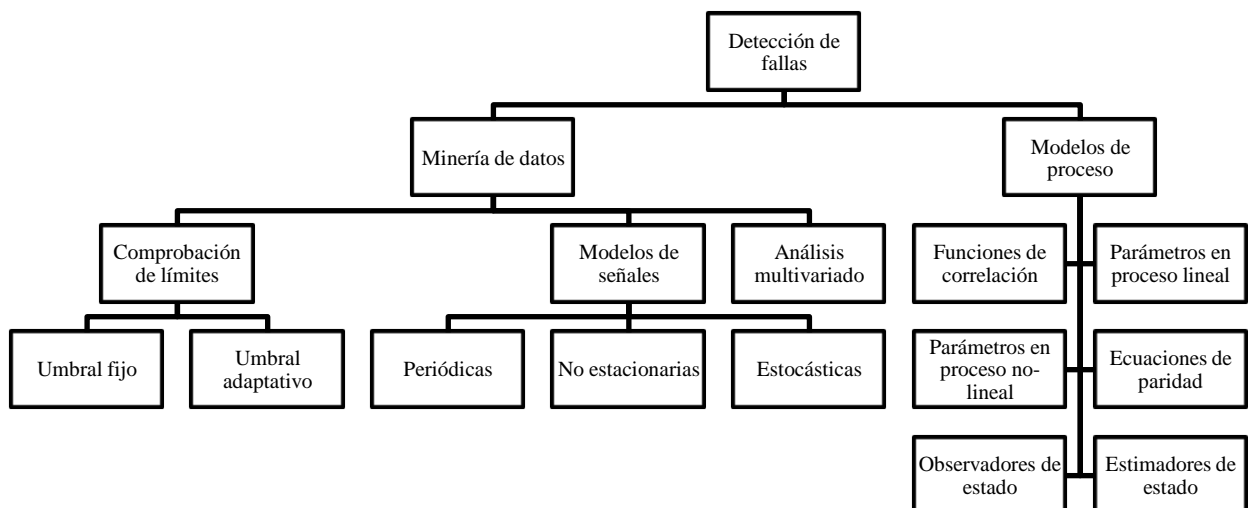


Figura 2.2: Métodos de detección de fallas con señales individuales.

2.2.1 MINERÍA DE DATOS

Los métodos de detección basados en minería de datos utilizan señales del sistema para generar características del proceso que muestren si hay presencia o no de algún evento anómalo. Estas características pueden ser los límites o tendencias de las señales, u obtenerse a través de modelos de señales o con métodos de análisis multivariado.

2.2.1.1 MÉTODOS CON COMPROBACIÓN DE LÍMITES Y TENDENCIAS

Uno de los métodos más frecuentes y simples de utilizar son aquellos que se enfocan en comprobar límites y tendencias con un umbral fijo [16]. Considerando una señal $y(t)$, la comprobación de límites se encarga de verificar que $y(t)$ se encuentre entre un valor mínimo y máximo, mientras que la comprobación de tendencias verifica que la derivada en el tiempo $\dot{y}(t)$ se encuentre en un rango dado por un valor mínimo y máximo tal como se observa en la Figura 2.3a y Figura 2.3b.

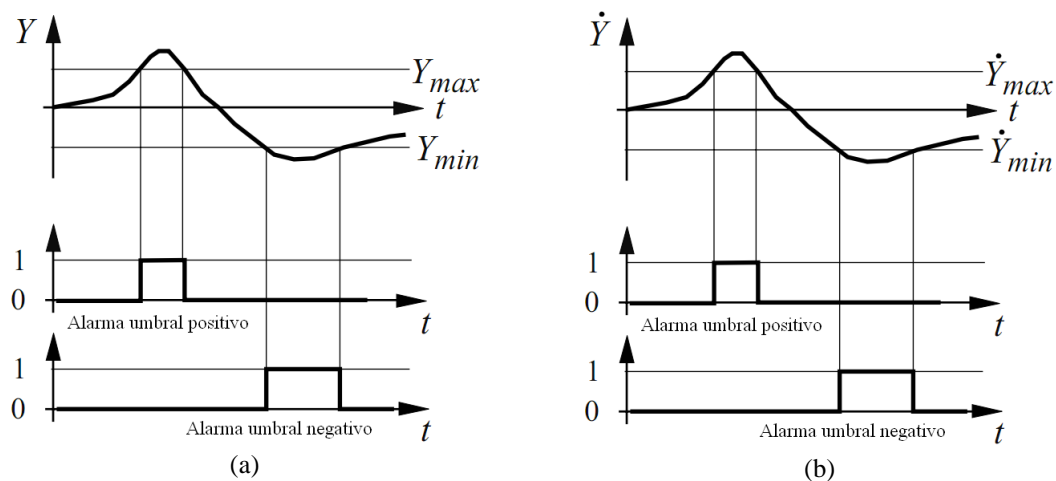


Figura 2.3: (a) Método de comprobación de límites y (b) método de comprobación de tendencias [33].

Los métodos por detección de cambios se enfocan en verificar los cambios que experimenta la media μ_0 y la desviación estándar σ_0 de una variable Y cuando se encuentra en un modo de operación nominal o normal (Figura 2.4). Considerando que posterior a la falla la media y la desviación estándar de la variable Y es μ_1 y σ_1 respectivamente, se distinguen tres casos:

- La media μ_0 cambia a $\mu_1 = \mu_0 + \Delta\mu$, la desviación estándar permanece constante $\sigma_1 = \sigma_0$.
- La media permanece constante $\mu_1 = \mu_0$ y la desviación estándar cambia a $\sigma_1 = \sigma_0 + \Delta\sigma$.
- La media μ_0 cambia a $\mu_1 = \mu_0 + \Delta\mu$, la desviación estándar cambia a $\sigma_1 = \sigma_0 + \Delta\sigma$.

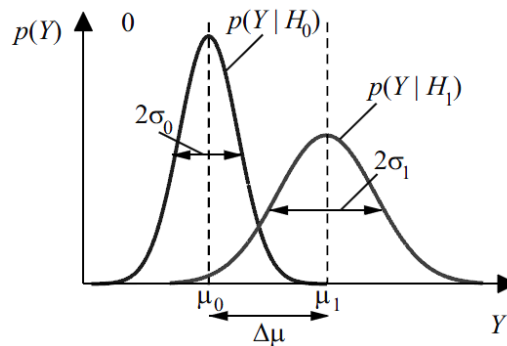


Figura 2.4: Densidad de probabilidad normal de la variable observada Y , μ_0 y σ_0 caracterizan el estado nominal, μ_1 y σ_1 el estado de falla del sistema [33].

La elección de un umbral depende exclusivamente de las fallas detectadas correctamente versus las falsas alarmas que puedan aparecer. En [5] se aplica una variante que utiliza un umbral difuso, donde no se consideran funciones de densidad para Y , si no, conjuntos difusos (Figura 2.5) con el objetivo de detectar cambios en imágenes en un sistema de radar.

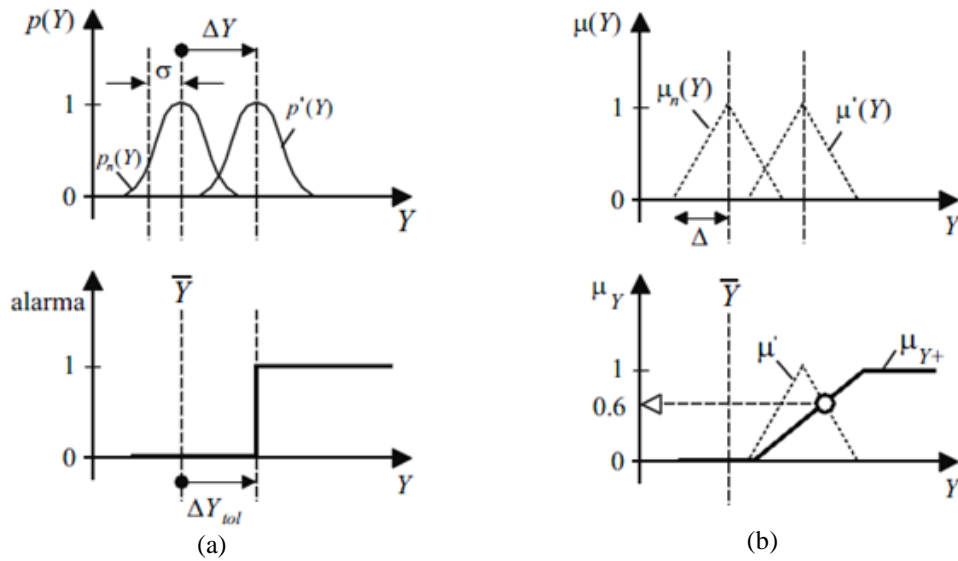


Figura 2.5: (a) Variable estocástica $Y(t)$ con función de densidad de probabilidad $p(Y)$ en estado normal, y $p'(Y)$ en estado de falla; (b) Variable estocástica $Y(t)$ como conjunto difuso $\mu_n(Y)$ en estado normal, y $\mu'(Y)$ en estado de falla [33].

2.2.1.2 MÉTODOS BASADOS EN MODELOS DE SEÑALES

Los métodos basados en modelos de señales pueden ser aplicados cuando en el proceso se presentan señales que muestran oscilaciones ya sea de naturaleza armónica, estocástica o ambas, y se producen anomalías en los actuadores que originan cambios en estas mediciones. Ejemplos de estos casos se dan en máquinas vibratorias como turbinas, motores a gas y máquinas de molienda, en las cuales las mediciones de posición, velocidad o aceleración, permiten detectar fallas. El esquema de la Figura 2.6 muestra las etapas de los métodos de modelos de señales para detectar fallas.

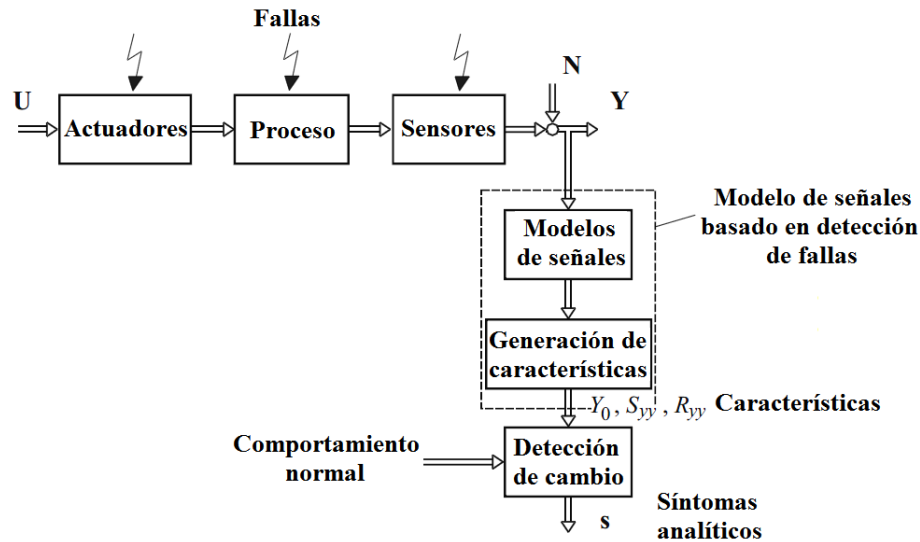


Figura 2.6: Esquema de los métodos de detección de falla con modelos de señales [32].

Tal como se observa en la Figura 2.2, los métodos de detección de fallas basados en modelos de señales se clasifican en tres tipos de acuerdo a las características de las señales: periódicas, no estacionarias o estocásticas. En el caso en que las señales sean periódicas se utilizan técnicas como el análisis de Fourier, funciones de correlación, la transformada rápida de Fourier (FFT) o la estimación del espectro de máxima entropía. La transformada wavelet es útil cuando las señales no son estacionarias. Finalmente, en el caso en que las señales son estocásticas se utiliza el análisis de correlación a través de la función de autocorrelación, el análisis del espectro de densidad obtenido como la transformada de Fourier de la función de autocovarianza o la estimación de parámetros de señales con modelos de tipo ARMA (Modelo autorregresivo de media móvil).

Diversos trabajos utilizan métodos con modelos de señales, como [18] donde se utiliza FFT para efectos de comparar con un método basado en ICA, o [13] por medio del análisis wavelet.

2.2.1.3 MÉTODOS DE ANÁLISIS MULTIVARIADO

Las técnicas alternativas a los métodos de detección basados en modelos de señales o procesos para problemas de gran escala, son las que se encuentran basadas en el análisis de datos lineales tales como PCA [14], [48], [50] y PLS [10], o análisis en datos no gaussianos como ICA [17], [18], [47], que son métodos de transformación que reducen el número de dimensiones del sistema.

Entre las opciones para aplicar estos métodos se encuentran:

- Detección de cambios utilizando las proyecciones que entrega PCA.
- Detección de cambios en las variables del espacio original, obtenidas a partir de las proyecciones.
- Análisis de residuos entre la variable original, y la variable obtenida a partir de las proyecciones.

2.2.2 MÉTODOS DE MODELOS DEL PROCESO

Los métodos basados en modelos de proceso para detección de falla encuentran relaciones matemáticas entre señales de entrada $U(k)$ y señales de salida $Y(k)$, con k representando un índice temporal, para extraer información y así encontrar cambios que podrían ser causados por fallas, tal como muestra el esquema general de la Figura 2.7. Estos métodos extraen características como parámetros θ , variables de estado x o residuos r , las que son comparadas con sus valores nominales para detectar eventuales cambios y fallas.

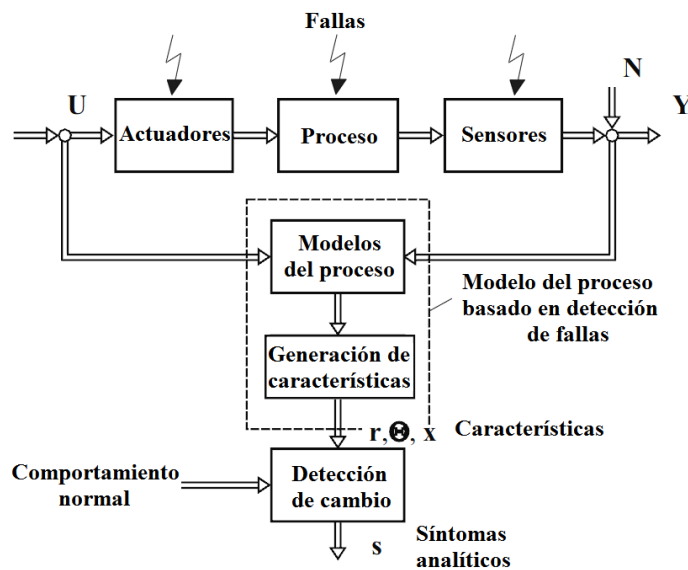


Figura 2.7: Esquema de los métodos de detección de falla con modelos del proceso [32].

Tal como se observa en la Figura 2.2, estos modelos pueden ser obtenidos a través de la identificación con funciones de correlación, con la estimación de parámetros en procesos lineales y no lineales, con ecuaciones de paridad, y finalmente, con observadores y estimadores de estado.

Las funciones de correlación son utilizadas bajo el supuesto que el proceso es estacionario y por lo tanto la respuesta al impulso puede ser determinada si son conocidas la función de autocorrelación de la señal de entrada y la función de correlación cruzada entre las señales de entrada y de salida.

Cuando se asume que el proceso se comporta linealmente, los modelos se pueden obtener utilizando el método de los mínimos cuadrados en el caso que los parámetros son constantes en el tiempo. También hay otras técnicas para modelos lineales dinámicos con señales continuas o discretas. Estas pueden verse en detalle en [33].

Además se encuentran los métodos de estimación de parámetros para aquellos procesos que presentan comportamientos no-lineales, ya sea estático o dinámico. En el caso estático, un método es utilizar un modelo de regresión lineal en sus parámetros, con el supuesto de que la señal de salida $Y(k)$ depende no-linealmente de la entrada $U(k)$ del proceso de acuerdo a (2.1), y por lo tanto se requiere tener un grado de conocimiento del comportamiento del sistema. En el caso dinámico, las técnicas están basadas en aproximadores polinomiales, como el modelo generalizado de Hammerstein o el modelo paramétrico de Volterra [26], [31]. Además, en los casos en que la estructura del proceso es desconocida, es conveniente utilizar modelos de caja negra como ANN [3], [6], los cuales pueden aplicarse en sistemas tanto estáticos como dinámicos.

$$Y(k) = k_0 + U(k)K_1 + U^2(k)K_2 + \dots + U^q(k)K_q \quad (2.1)$$

Las ecuaciones de paridad plantean diseñar un modelo nominal donde el proceso se encuentra sin fallas, y compararlo con el proceso real. Las diferencias entre estas señales son conocidas como residuos, los cuales describen las discrepancias que hay entre el proceso y el modelo, y de esta forma detecta si hay o no eventos anómalos o fallas. Los modelos pueden ser obtenidos a partir de funciones de transferencia o modelos en el espacio de estados [25].

Finalmente, los observadores de estado y los estimadores de estado, calculan el error entre mediciones de la salida del proceso, y un modelo ajustable de la salida o de los estados, respectivamente, y por lo tanto, son métodos que también se basan en la generación de residuos. Entre las alternativas más clásicas, están el utilizar el observador de Luenberger [4], [29], [39] y el estimador de estados Filtro de Kalman [9], [49].

2.3 AISLAMIENTO DE FALLAS

El aislamiento de fallas tiene por objetivo determinar el origen de la falla detectada, aislando las componentes del sistema cuando este no está en su operación nominal. Tal como se puede observar en la Figura 2.8, los métodos de aislamiento pueden ser por clasificación o por inferencia respectivamente, y se basan en disponer de síntomas analíticos, síntomas heurísticos y el conocimiento heurístico del proceso. La información obtenida en esta sección ha sido extraída desde [34]

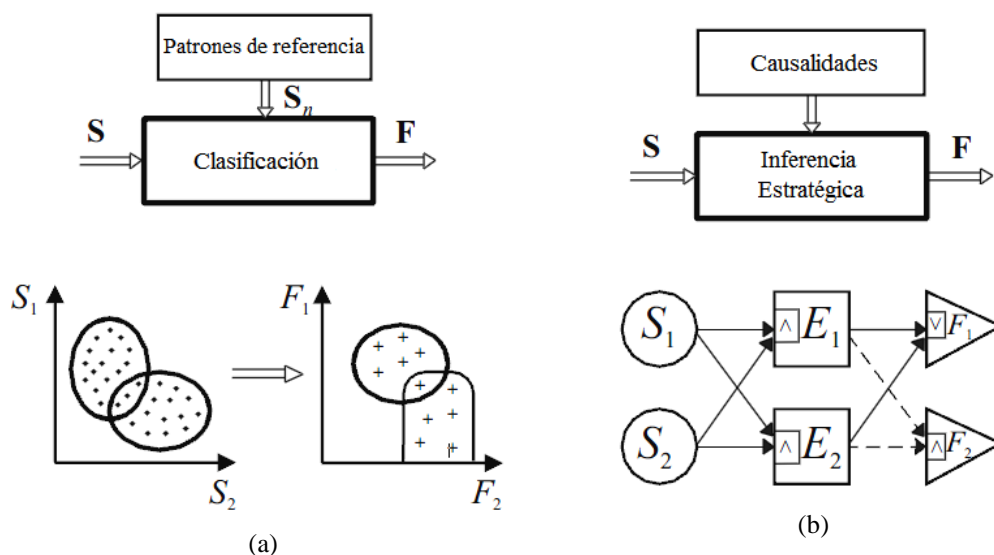


Figura 2.8: Aislamiento de fallas por métodos de clasificación (a) y por métodos de inferencia (b) [32].

Los síntomas analíticos son aquellos que se obtienen verificando los límites de las señales medidas en el proceso o señales obtenidas de los métodos de detección de fallas. Los síntomas heurísticos son mediciones cualitativas que se obtienen desde un operador del proceso y suelen ser expresiones de la forma “pequeño”, “mediano” o “mucho”.

En la Sección 2.3.1 se introducen los métodos de clasificación que han sido utilizados clásicamente. Estos métodos son preferibles cuando se desconocen las causas o los síntomas de la falla, y pueden ser desde reconocimiento de patrones, clasificadores geométricos, hasta clasificadores con redes neuronales o clasificadores difusos. En la Sección 2.3.2 se introducen los métodos por inferencia, los cuales son preferibles cuando se dispone de información adicional, como por ejemplo relaciones entre los síntomas que tiene el sistema cuando se encuentra en una falla, esto es, expresiones de la forma “*si-entonces*”.

La Figura 2.9 es un esquema que contiene los variados tipos de métodos por clasificación y por inferencia. [32].

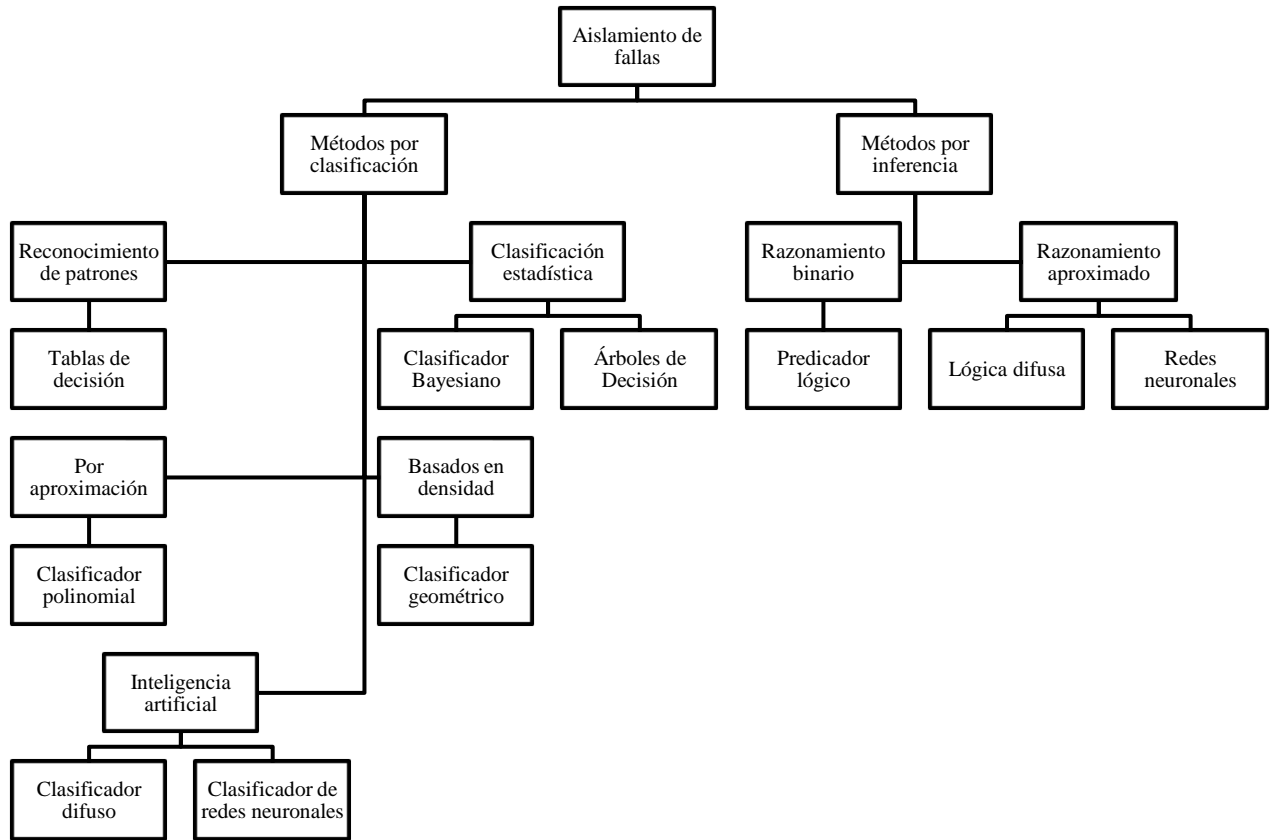


Figura 2.9: Métodos de aislamiento de fallas.

2.3.1 MÉTODOS DE CLASIFICACIÓN

El aislamiento de fallas por medio de métodos de clasificación se utiliza cuando no se dispone de conocimiento físico o estructural del sistema que permita relacionar los síntomas con las fallas observadas, y por tanto, basa sus técnicas principalmente en el análisis de las señales disponibles en el proceso (Figura 2.8a).

Históricamente, los métodos de clasificación estadística fueron los primeros en ser empleados en los problemas de aislamiento de fallas. Dichos métodos asumen que cada falla en el sistema tiene una distribución de probabilidad, típicamente Gaussiana, para la cual es necesaria fijar un umbral de decisión que separe una clase de la otra. Estos métodos son utilizados en diversas aplicaciones, como en [12] donde se detectan anomalías en submarinos utilizando el test de Hotelling (obtenido a partir de PCA), y posteriormente FDA para clasificar las fallas detectadas. Así mismo, en [13] se utilizan varianzas concatenadas de la transformada wavelet y LDA para efectos de detección y aislamiento de fallas. En [22] se propone aplicar PLS para detectar y FDA para identificar fallas en una planta que consta de dos tanques cilíndricos y cuyo objetivo es mantener un nivel de referencia fijo a partir de dos controladores digitales PID. En [46] se aplica PCA para reducir dimensiones y un clasificador Bayesiano ingenuo (del inglés: Naive Bayes Classification) para diagnosticar fallas en el proceso de Tennessee Eastman.

Dentro de los clasificadores estadísticos, se encuentran los árboles de decisión [44]. Un ejemplo común se da en la botánica, en la clasificación de las especies, donde por medio de una serie de preguntas cada vez más específicas se llega finalmente a la especie que cumple con todas las características y que ninguna otra tiene. En el aislamiento de fallas, estas características son los síntomas del sistema. Un ejemplo es el mostrado en la Figura 2.10 donde se distinguen dos clases: F_1 y F_2 , las cuales abarcan regiones en el espacio de síntomas del sistema. El árbol de decisión se va formando en base a límites de las regiones en las que estos síntomas se verifican.

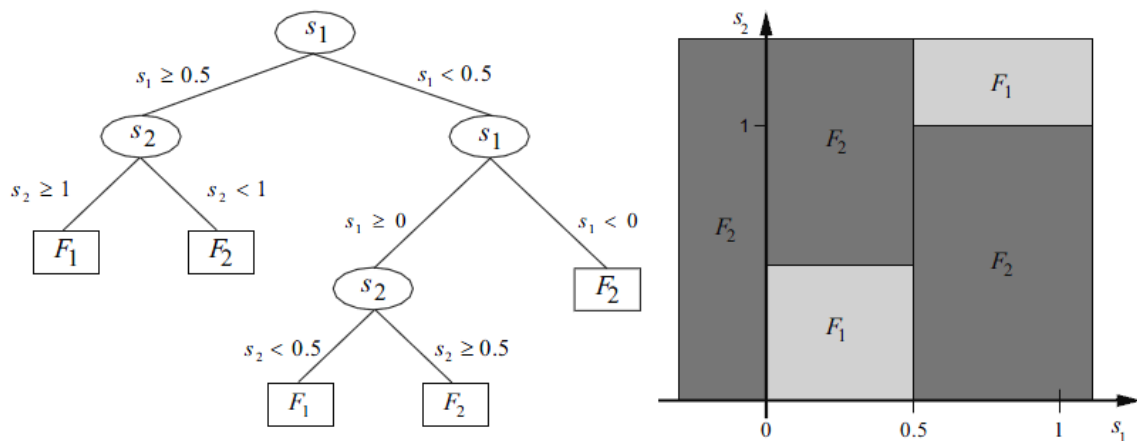


Figura 2.10: Método de árboles de decisión para aislamiento de fallas [34].

Los métodos basados en clasificadores geométricos son muy similares a los estadísticos y consisten en clasificar los datos en base a la distancia que tienen a datos de referencia, siendo uno de los más famosos el algoritmo de los *k-vecinos más cercanos* (k-NN). Un punto de la base de datos es asignado a la clase *C* si esta es la clase más frecuente entre los *k* elementos más cercanos del conjunto de entrenamiento. La Figura 2.11 muestra un ejemplo en el cual se desea clasificar el círculo en la clase de los cuadrados o en la clase de los triángulos. Así, si $k = 3$ entonces el círculo pertenecería a la clase de los triángulos, y si $k = 5$ a la clase de los cuadrados.

La Figura 2.11 muestra un ejemplo en el cual se desea clasificar el círculo etiquetado con una *X* en la clase *A* de los cuadrados o en la clase *B* de los triángulos. Así, si $k = 1$ entonces el círculo pertenecería a la clase *A* y si $k = 7$ a la clase *B*. Cuando $k = 10$ no hay decisión.

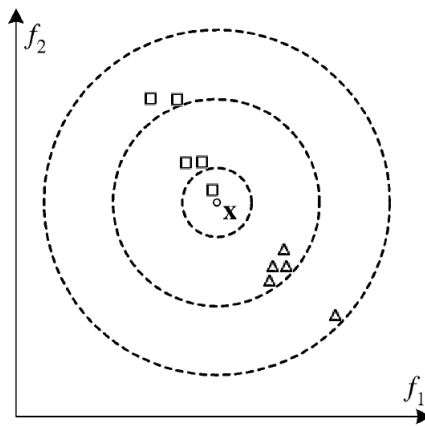


Figura 2.11: Ejemplo del clasificador geométrico *k-vecinos más cercanos* (k-nearest neighbors) [28].

Esta técnica, dependiendo del objetivo, es utilizada en diversas aplicaciones como en [2], [28] donde se desea clasificar para diagnosticar fallas, en el primero en la caja de cambios de un tractor y en el segundo en el proceso de fabricación de semiconductores, o en [45] para predecir el movimiento de objetos.

Finalmente, destacan los métodos basados en inteligencia artificial como lo son las redes neuronales y la lógica difusa. Estos han sido los últimos métodos en ser aplicados en los problemas de aislamiento de anomalías. A diferencia de los métodos de clasificación estadística que asumen una distribución Gaussiana, las redes neuronales están diseñadas para clasificar cualquier función, teniendo típicamente como función objetivo el minimizar la suma de los errores cuadráticos. En los métodos basados en inteligencia artificial, destacan los trabajos realizados en una planta de generación termoeléctrica [42], en un sistema de molienda [23] y en modelos teóricos de motores lineales [38].

2.3.2 MÉTODOS DE INFERENCIA

Estos métodos son utilizados cuando se tiene información adicional del sistema, y son al menos conocidas las relaciones entre fallas y síntomas en forma de relaciones causales: falla \rightarrow eventos \rightarrow síntomas (Figura 2.8b). Estos métodos de aislamiento son expresados cualitativamente en forma de reglas de la forma

SI < condición > ENTONCES < conclusión >

Las premisas son hechos expresados en la forma de síntomas s_i como entradas, y las conclusiones incluyen eventos e_i y fallas f_i como la causa lógica de los hechos. Los métodos por inferencia pueden ser de razonamiento binario o de razonamiento aproximado dependiendo si los síntomas son de naturaleza discreta o continua.

Los métodos de razonamiento binario, en particular los árboles de fallas, son una herramienta gráfica que muestra las relaciones que hay entre la falla que está ubicada en la parte superior del árbol, y los diferentes síntomas y condiciones en la parte inferior tal como se observa en la Figura 2.12. Son importantes durante las etapas tempranas de la identificación de los subsistemas que más contribuyen a una falla en el sistema. Si se agrega información estadística como las probabilidades de que ocurra cada evento, es posible obtener la probabilidad de que ocurra una falla en particular. En [24] se tiene un ejemplo de utilización de árboles en un sistema calentador, ventilador y aire acondicionado (HVAC, del inglés: Heating, Ventilation and Air Conditioning) para diagnosticar fallas.

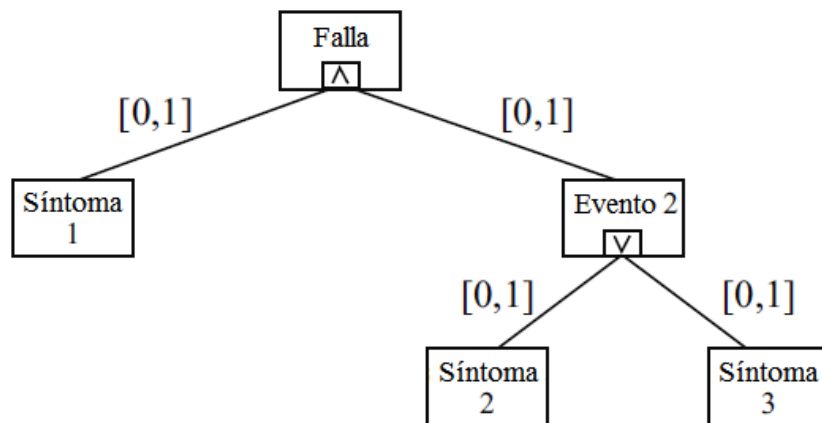


Figura 2.12: Método de razonamiento binario para aislamiento de fallas [34].

Los métodos por razonamiento aproximado son más apropiados cuando se tiene un proceso continuo donde los síntomas también son variables continuas. Un método es que los síntomas sean definidos como conjuntos difusos, dejando de ser variables "0" o "1", para convertirse en variables de la forma "normal", "incremento pequeño" o "incremento grande" [35]. En [7] se aplica este método basado en lógica difusa para diagnosticar fallas en un satélite.

Otro enfoque es utilizar métodos basados en sistemas híbridos neuro-difusos (Figura 2.13). Por un lado, las redes neuronales recuperan la información obtenida de los datos para aplicar técnicas de optimización, y por otro lado, los sistemas difusos incorporan los elementos intuitivos dando la posibilidad de integrar conocimiento de expertos. Así, estos sistemas híbridos incorporan información desde dos vertientes: conocimiento de expertos y datos experimentales.

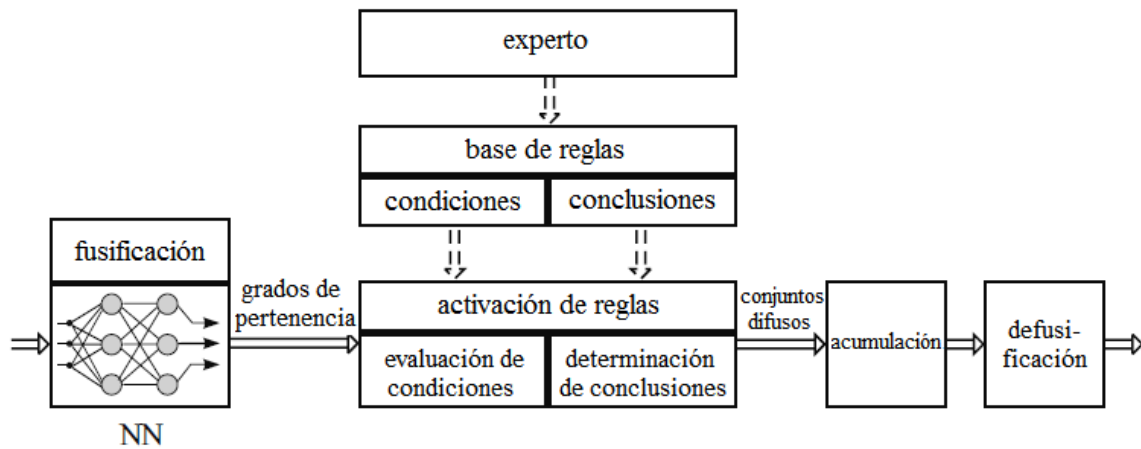


Figura 2.13: Método híbrido neuro-difuso para aislamiento de fallas [34].

La principal área de aplicación de estos sistemas es la aproximación de funciones no lineales con el objetivo de modelar e identificar sistemas. Sin embargo, son utilizados en el diagnóstico de fallas debido a su capacidad de generar sistemas que permiten tomar decisiones. En [8], se desarrolla un sistema de aislamiento de fallas para ser aplicado en una planta de generación termoeléctrica y que consiste en un sistema experto neuro-difuso utilizando ANN.

2.4 DETECCIÓN Y AISLAMIENTO DE FALLAS (FDI) UTILIZANDO MÉTODO DE LOS RESIDUOS

Un esquema de detección y aislamiento de fallas, utilizando análisis en los residuos (Figura 2.14), contempla principalmente dos etapas: obtención de residuos y determinación de anomalía. El esquema comienza con una selección de las características del proceso que serán modeladas con alguna de las técnicas de modelación ya vistas. A continuación, las salidas estimadas de estos modelos y_{f_est} se comparan con las salidas reales del proceso. A través de esta comparación se obtienen las variables residuales (2.2). En el caso ideal en que estos residuos son cero entonces no hay ningún conflicto, al menos dentro de los límites que explica el modelo.

$$r = y_f - y_{f_est} \quad (2.2)$$

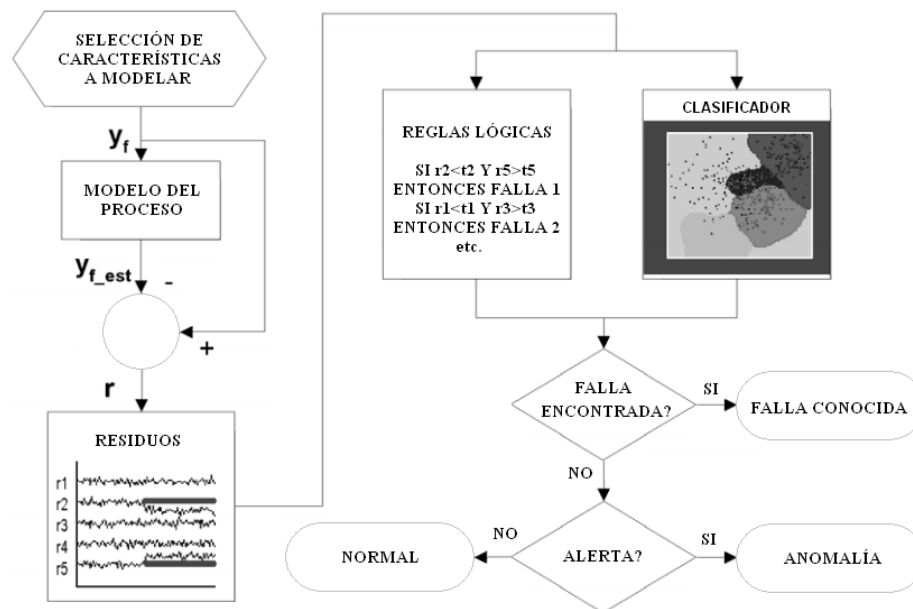


Figura 2.14: Esquema de un sistema FDI utilizando método de los residuos [40].

La siguiente etapa consiste en aplicar a los residuos los métodos de clasificación o de inferencia que fueron explicados en la Sección 2.3.1 y Sección 2.3.2 respectivamente. De acuerdo a los resultados que entregan estos métodos, se podrá concluir si corresponde a una falla conocida, a una anomalía desconocida hasta el momento, o se trata simplemente de una falsa alarma y el proceso sigue estando en modo nominal.

2.5 HERRAMIENTAS UTILIZADAS EN SISTEMAS FDI

En el Capítulo 3 y Capítulo 4 se diseñan e implementan algoritmos y metodologías con la finalidad de detectar y aislar anomalías dentro de la operación de un proceso. A continuación, y con el fin de facilitar la comprensión de los conceptos expuestos en dichos capítulos, se describen algunas de las herramientas que son comúnmente utilizadas en sistemas FDI, y en particular, por los algoritmos allí descritos.

2.5.1 ERROR CUADRÁTICO MEDIO

El error cuadrático medio (ECM) de un estimador de una variable es un concepto ampliamente utilizado para medir la diferencia entre dicha estimación \hat{X} y el valor real X . El ECM es una función que se calcula como el valor esperado del error cuadrático (2.3).

$$ECM = E \left[(X - \hat{X})^2 \right] \quad (2.3)$$

En la herramienta de generación de modelos (Sección 3.1) el error cuadrático medio se calcula con un número M de variables reales X_i y estimadas \hat{X}_i , para una observación fija k . El promedio sobre todas las observaciones es utilizado como criterio para determinar si un modelo es mejor que otro o no.

$$ECM(k) = \frac{1}{M} \sum_{i=1}^M \left(X_i(k) - \hat{X}_i(k) \right)^2 \quad (2.4)$$

2.5.2 FILTRO DE MEDIANA MÓVIL

El vector binario que determina si debe haber o no una alarma en el proceso, se obtiene a través de otras señales cuyo ruido genera una serie de falsas alarmas en donde el sistema supervisor indica que hay presente una anomalía, cuando en la realidad, esta no existe. Este filtro se utiliza con el objetivo de eliminar estos errores, y además presentar gráficos con un ruido menor para que puedan ser mejor interpretados. La mediana móvil X_{MM} de largo k para una señal X en la observación i -ésima es:

$$X_{MM}(i) = \text{mediana} (X(j) \quad i - k \leq j \leq i + k) \quad (2.5)$$

En la Figura 2.15 se observa un ejemplo con una señal X (línea punteada) y la mediana móvil de largo 2 de esta (línea continua). Se observa como la mediana tiende a suavizar los valores que se alejan del resto (observaciones 4 y 7) y que simulan el ruido que puede aparecer en una señal.

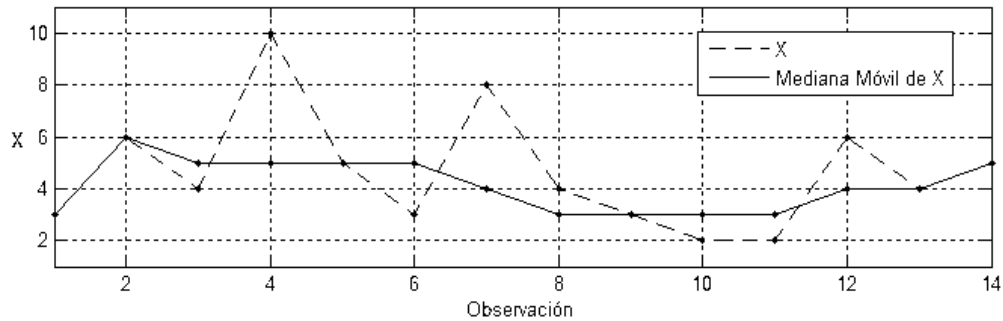


Figura 2.15: Señal de ejemplo X y su mediana móvil X_{MM} de largo 2.

2.5.3 AGRUPACIÓN POR K-MEDIAS (K-MEANS CLUSTERING)

Es un método que agrupa un conjunto de datos de n elementos en k grupos y que se basa en clasificar de acuerdo a la menor distancia que tiene un elemento a los k centros de los grupos. El algoritmo descrito en [30], es el siguiente:

1. Inicializar k medias μ_1, \dots, μ_k (Figura 2.16a). Estos serán inicialmente los centros de los grupos.
2. Asignar cada elemento x_j a un grupo S_i de acuerdo a la media μ_i más cercana (Figura 2.16b).
3. Recalcular la media μ_i (Figura 2.16c) como:

$$\mu_i = \frac{1}{|S_i|} \sum_{x_j \in S_i} x_j \quad (2.6)$$

4. Continuar hasta que no hayan cambios en las medias (Figura 2.16d).
5. Retornar las medias μ_i .
6. Fin del algoritmo.

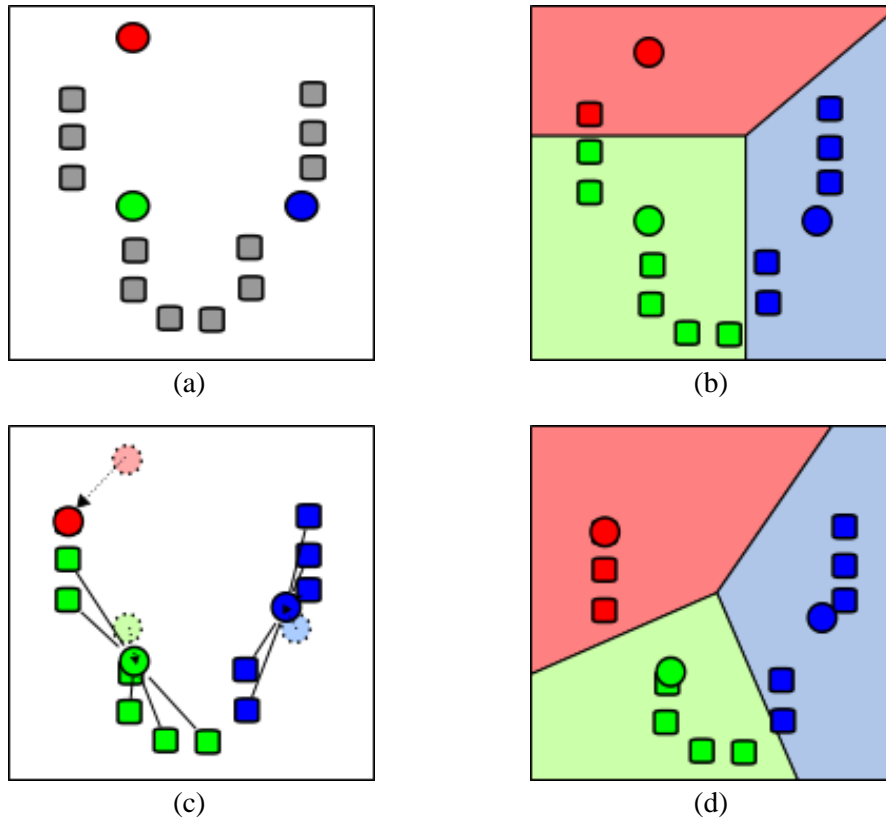


Figura 2.16: Pasos del algoritmo de agrupación por k-medias, considerando $k = 3$. Los centros de los grupos están representados por círculos, y los datos por cuadrados [53].

2.5.4 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

El análisis de componentes principales o PCA es una técnica lineal de reducción de dimensiones que captura la máxima variabilidad de los datos, obteniendo los llamados *vectores de carga* [20]. Dada una matriz de datos $X \in \mathbb{R}^{N \times M}$, con N el número de observaciones y M el número de variables, PCA resuelve el problema de optimización:

$$\max_{v \neq 0} \frac{v^T X^T X v}{v^T v} \quad (2.7)$$

Donde $v \in \mathbb{R}^M$.

Para obtener la solución de este problema, se debe obtener de la matriz de covarianza S , la matriz V que contiene en sus columnas los vectores propios (o vectores de carga) asociados a la matriz diagonal Λ que contiene los valores propios λ_i ordenados de mayor a menor ($\lambda_1 \geq \lambda_2 \dots \geq \lambda_M \geq 0$) siendo todos mayores o iguales a cero. Esta matriz se define como:

$$S = \frac{1}{n-1} X^T X = V \Lambda V^T \quad (2.8)$$

Con la finalidad de reducir el ruido que pueden representar las componentes asociadas a los valores propios menores, se selecciona una cantidad a de los primeros valores y vectores propios. Así, sea $P \in \mathbb{R}^{M \times a}$, con $a \leq M$, la matriz que contiene las primeras a columnas de V . Entonces, las proyecciones de X en un espacio de menor dimensión están contenidas en la matriz:

$$T = X P \quad (2.9)$$

La proyección de T en el espacio original de dimensión M es

$$\hat{X} = T P^T \quad (2.10)$$

La matriz residual E , que captura las variaciones en las observaciones asociadas a las componentes principales con valores propios que no fueron considerados en la matriz P , se calcula como:

$$E = X - \hat{X} \quad (2.11)$$

Las columnas t_i de la matriz T en el conjunto de entrenamiento, cumplen con las siguientes propiedades:

- $Var(t_1) \geq \dots \geq Var(t_a)$
- $media(t_i) = 0; \forall i$
- $t_i t_k^T = 0; \forall i \neq k$
- No existe otra expansión ortogonal de a componentes que capture mayor variación en los datos.

Cuando se incorpora un nuevo vector fila x_n ($1 \times M$) a la base de datos, y se desean obtener las primeras a componentes principales, se debe simplemente utilizar (2.12). Es muy importante notar que el vector x_n debe ser normalizado utilizando la media y desviación estándar de los datos originales, sin incorporar la información que agrega este vector.

$$t_n = x_n P \quad (2.12)$$

2.5.5 TEST DE ESTADÍSTICO DE HOTELLING

Sea la matriz de datos de entrenamiento $X \in \mathbb{R}^{N \times M}$, con N el número de observaciones y M el número de variables. Se define además S como en (2.8). El estadístico de Hotelling es una norma-2 de un vector de observación x de su media, escalada en la dirección de los vectores propios de S , y es inversamente proporcional a la desviación estándar a lo largo de los vectores propios. Permite caracterizar la variabilidad de los datos en el espacio de observación completo de dimensión M .

Sea $x \in \mathbb{R}^M$. El estadístico de Hotelling T^2 se calcula a partir de las matrices obtenidas con (2.8), (2.9) y a través del vector $z \in \mathbb{R}^a$ calculado en (2.13):

$$z = \Lambda^{-1/2} V^T x \quad (2.13)$$

$$T^2 = z^T z = x^T P \Lambda_a^{-1} P^T x \quad (2.14)$$

El indicador de Hotelling es utilizado en detección para verificar si el error de estimación se encuentra dentro de una región limitada por un umbral, y así determinar si es aceptable o no. El umbral de Hotelling basado en un nivel de confianza α , que sea adecuado para detección se determina asumiendo que las observaciones son muestras aleatorias de una distribución normal multivariable. Para detectar observaciones anómalas en el conjunto de entrenamiento (u *outliers*) se define el siguiente umbral:

$$T_\alpha^2 = \frac{(n-1)^2 \left(\frac{m}{(n-m-1)} \right) F_\alpha(m, n-m-1)}{n \left(1 + \left(\frac{m}{(n-m-1)} \right) F_\alpha(m, n-m-1) \right)} \quad (2.15)$$

Donde $F_\alpha(g, h)$ es el punto crítico superior $(100 * \alpha)\%$ de la distribución F de Fisher de g y h grados de libertad. Los eventos que ocurren en observaciones que no son del conjunto de entrenamiento pueden ser detectadas utilizando el umbral:

$$T_\alpha^2 = \frac{m(n-1)(n+1)}{n(n-m)} F_\alpha(m, n-m) \quad (2.16)$$

2.5.6 MODELOS BASADOS EN SIMILITUD (SBM)

SBM es una técnica de modelación no paramétrica [11], [21], y por tanto, no requiere a priori conocimiento de la estructura del sistema a modelar puesto que su implementación está basada en identificar similitudes y relaciones entre las variables de un conjunto de observaciones dado. Algunas aplicaciones de SBM para detección de fallas pueden verse en [1], [11], [40]. Sea el sistema estático siguiente:

$$y = f(x) \quad (2.17)$$

Donde $x \in \mathbb{R}^m$ e $y \in \mathbb{R}^p$ son las variables de entrada y salida del sistema respectivamente, y $f(\cdot)$ es una función desconocida.

A continuación, se definen las matrices de entrenamiento D_i y D_o a través de observaciones de variables de entrada y salida respectivamente:

$$D_i = [x_1 \quad \dots \quad x_n] \in \mathbb{R}^{m \times n} \quad (2.18)$$

$$D_o = [y_1 \quad \dots \quad y_n] \in \mathbb{R}^{p \times n} \quad (2.19)$$

Donde $y_i = f(x_i)$. Los pares $[x_i, y_i]_{i=1 \dots n}$ deben ser representativos de los puntos de operación del proceso que se desea modelar. En la Sección 3.1 se resumen los pasos de la herramienta de generación de modelos diseñada en [1], y que se encarga de encontrar observaciones representativas del proceso a través de una serie de etapas.

Dado un vector x^* , SBM encuentra una estimación \hat{y}^* de $y^* = f(x^*)$ por medio de una combinación lineal de las columnas de D_o :

$$\hat{y}^* = D_0 w \quad (2.20)$$

El vector w de (2.20) se obtiene como:

$$w = \frac{\hat{w}}{1^T \cdot \hat{w}} \quad (2.21)$$

$$\hat{w} = (D_i^T \Delta D_i)^{-1} (D_i \Delta x^*) \quad (2.22)$$

Donde Δ es el operador de similitud. Para dos elementos $A, B \in \mathbb{R}^n$, $A \Delta B \in \mathbb{R}^+$ debe ser simétrica, alcanzar su máximo en $A = B$ y ser monótonamente decreciente con $\|A - B\|$. De acuerdo a [11], luego de un estudio preliminar, se determinó que el operador de similitud que mejor captura la variabilidad de los datos es el operador triangular saturado definido como (2.23).

$$A \Delta B = \begin{cases} d - \|A - B\| & \|A - B\| \leq d + \varepsilon \\ \varepsilon & \|A - B\| > d + \varepsilon \end{cases} \quad (2.23)$$

A pesar que SBM asume que el sistema es estático, es posible adaptarlo para sistemas dinámicos discretos si se dispone de una secuencia temporal de observaciones. En este caso, el problema puede ser abordado desde dos perspectivas: abandonar las propiedades del sistema dinámico y abordarlo como si fuera estático, o incorporar regresores y considerarlos como una entrada o una salida.

CAPÍTULO 3: HERRAMIENTAS DE GENERACIÓN DE MODELOS BASADOS EN SIMILITUD

Los sistemas supervisores requieren de una etapa preliminar que consiste en el estudio de los eventos que se desean detectar, analizando las variables involucradas y las observaciones correspondientes al estado nominal y al estado anormal del sistema. Esta metodología es explicada en detalle en el Capítulo 4. En este capítulo, se describen las dos herramientas que son utilizadas por esta metodología: herramienta de generación de modelos basados en similitud y herramienta de generación de modelos clasificadores basados en similitud.

En la Sección 3.1 se describe la herramienta para generar modelos basados en similitud (SBM). Se compone de una serie de etapas iterativas que tienen el objetivo de optimizar el modelo, encontrando aquellas observaciones que son más representativas del proceso en base a un conjunto de variables de entrada y salida. Esta herramienta fue desarrollada en la primera fase del proyecto de detección de anomalías perteneciente a la empresa CONTAC INGENIEROS LTDA., en colaboración con la UNIVERSIDAD DE CHILE y ENDESA CHILE. Esta fase ha sido documentada en el trabajo de título [1], donde se realiza una descripción detallada de la herramienta y una aplicación de esta en una turbina a gas perteneciente a ENDESA CHILE. El procedimiento para obtener las variables utilizadas para modelar será descrito posteriormente en la Sección 4.2. La herramienta de generación de modelos para detección de anomalías, es posteriormente utilizada en la Fase II del sistema supervisor (Sección 4.3) para obtener los residuos de las variables de salida y así determinar si el proceso se encuentra en presencia de alguna de las anomalías en estudio.

En la Sección 3.2 se describe la herramienta para generar un modelo clasificador SBM. Este modelo, es utilizado en la Fase I del sistema supervisor para alertar si el sistema se encuentra en presencia de algún evento anómalo o no. Se basa en determinar la agrupación a la que pertenece una nueva observación. Si esta se encuentra fuera de los límites de todas las agrupaciones, entonces se genera una señal de alerta de un posible evento.

3.1 HERRAMIENTA DE GENERACIÓN DE MODELOS DE ANOMALÍAS BASADOS EN SIMILITUD

La herramienta de generación de modelos basados en similitud es utilizada por el sistema supervisor en la etapa fuera de línea para construir los modelos multivariantes SBM para detección de anomalías a partir de una base de datos con un conjunto de variables de entrada y de salida.

En [1], además de una descripción general de la estructura del algoritmo, se realiza una descripción detallada y una aplicación de este en la detección de la suciedad de un compresor en una turbina a gas. Tanto los datos de dicha turbina a gas, como el modelo generado en esa aplicación, serán utilizados en el Capítulo 5 para emplearlos en el sistema supervisor que ha sido diseñado.

La herramienta de generación de modelos basados en similitud consiste en una función de MATLAB llamada *ModeloSBM*(·) que cuenta con 4 etapas (análisis inicial, modelo preliminar, optimización y reajuste), tal como se observa en la Figura 3.1, y que utiliza una serie de otras funciones también generadas en ambiente MATLAB. En la Tabla 3.1 se encuentran las variables de entrada y salida de la función *ModeloSBM*(·).

El objetivo de esta función es encontrar, a partir en conjunto de observaciones, aquellas que son más representativas del proceso. Es importante destacar que esta metodología no necesariamente encontrará la solución óptima global, sino que se sigue un procedimiento que va mejorando al primer modelo generado. La función objetivo de este problema de optimización se obtiene con el promedio del error cuadrático medio de los residuos. Cada uno de los pasos comienza con una selección de observaciones las cuales se almacenan en las matrices D_i y D_o , y con la función SBM se estiman las salidas para todo el conjunto de entrenamiento. Los residuos obtenidos como la diferencia entre las variables de salida reales y estimadas, son transformados al espacio de componentes principales y con eso se obtiene el indicador de Hotelling. Con este indicador, y a través de un criterio de incorporación/eliminación, se agregan y eliminan observaciones en las matrices D_i y D_o .

Tabla 3.1: Argumentos de la función *ModeloSBM*(·).

| | Nombre | Descripción |
|-----------------|-------------------|---|
| Entradas | <i>Xmodel</i> | Matriz cuyas columnas son las variables de entrada del modelo SBM, y sus filas son las observaciones del proceso. |
| | <i>Ymodel</i> | Matriz cuyas columnas son las variables de salida del modelo SBM, y sus filas son las observaciones del proceso. |
| | <i>OBSpercent</i> | Es el porcentaje del total de observaciones que se utilizan para generar el modelo SBM. Típicamente: <i>OBSpercent</i> = 10%. |
| Salidas | <i>Di</i> | Matriz de $N \times M_1$ característica del modelo SBM. Las filas son las observaciones seleccionadas y las columnas son las variables de entrada del modelo. |
| | <i>Do</i> | Matriz de $N \times M_2$ característica del modelo SBM. Las filas son las observaciones seleccionadas y las columnas son las variables de salida del modelo. |
| | <i>ETrain</i> | Matriz de residuos del modelo, obtenida como la diferencia entre las variables de salida reales y estimadas. |
| | <i>meanETrain</i> | Media de cada columna de la matriz de residuos <i>ETrain</i> . |
| | <i>sdETrain</i> | Desviación estándar de cada columna de la matriz de residuos <i>ETrain</i> . |
| | <i>vpETrain</i> | Vectores propios que caracterizan el espacio de componentes principales de la matriz <i>ETrain</i> . |
| | <i>vlETrain</i> | Valores propios que caracterizan el espacio de componentes principales de la matriz <i>ETrain</i> . |

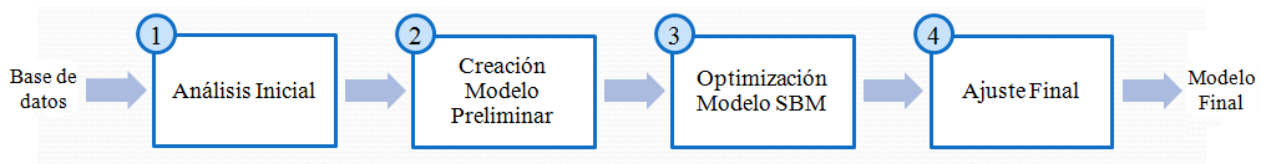


Figura 3.1: Etapas de la herramienta de generación de modelos basados en similitud.

Para generar un modelo SBM se requiere de una base de datos de entrenamiento del sistema en el cual se desea detectar anomalías. Esta base de datos es una matriz donde cada fila corresponde a observaciones del proceso, mientras que las columnas son las variables de entrada y salida. La selección de las variables de entrada y de salida se realiza por medio de una metodología de estudio de eventos para generar modelos de anomalías registradas, y que es explicado en detalle en la Sección 4.2. Es importante destacar que la base de datos debe ser representativa de la condición que se desea modelar: diferentes puntos de operación, si es para modelar una condición normal o anómala.

3.1.1 ETAPA 1: ANÁLISIS INICIAL

La Etapa 1 consiste en obtener las matrices D_i y D_o del primer modelo SBM seleccionando de las observaciones de la base de datos de entrenamiento:

$$40\%OBSpercent = 0.4 * OBSpercent \quad (3.1)$$

Típicamente, corresponde a un 4% del total de los datos de entrenamiento. El procedimiento se inicia transformando los datos en el espacio de componentes principales y luego se grafican las primeras dos componentes tal como se observa en la Figura 3.2. De este modo se determina cualitativamente el número de agrupaciones. Por medio de la técnica de *agrupación por k-medias* se obtienen los centros de cada una de las agrupaciones y con ello, las observaciones más representativas de cada punto de operación, es decir, aquellas que se encuentran más cerca de los centros.

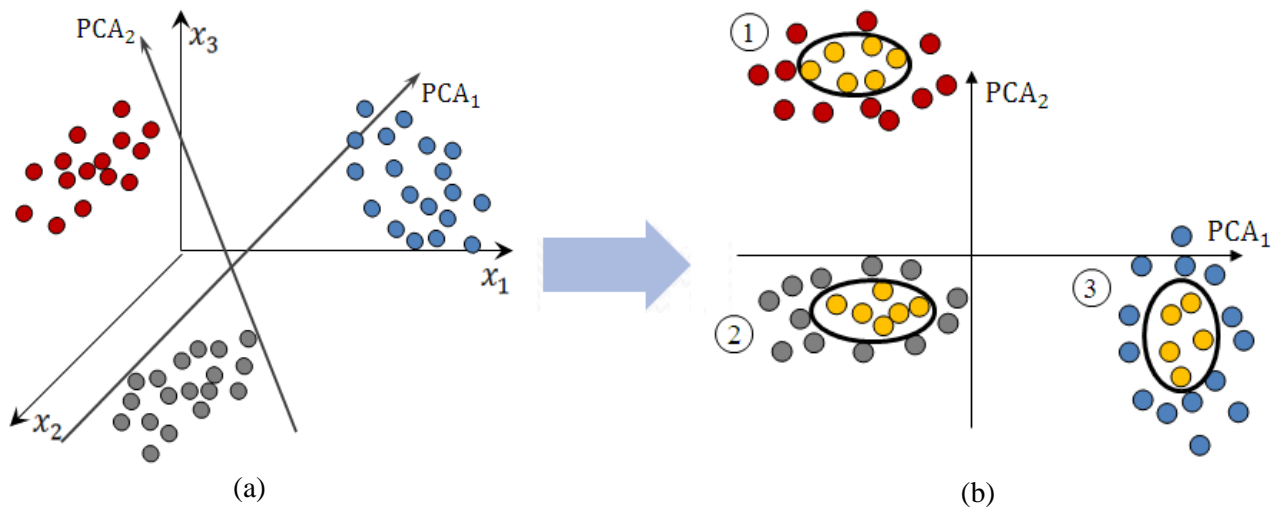


Figura 3.2: (a) Datos en el espacio original y (b) y en el espacio de componentes principales.

La selección de observaciones se realiza de acuerdo al tamaño que tiene cada agrupación, tal como se observa en la Tabla 3.2, asumiendo esa cantidad de elementos en cada agrupación de la Figura 3.2. Si bien el porcentaje de observaciones seleccionadas está dado por *OBSpercent*, debido a que se trazan circunferencias, hay un margen de error de $\pm 5\%$ al ser imposible en algunos casos alcanzar un porcentaje exacto.

Tabla 3.2: Porcentajes de observaciones seleccionadas en el ejemplo de la Figura 3.2b.

| Nº Agrupación | Observaciones | Nº óptimo de OBSselect (10%) | Nº mínimo de OBSselect (9.5%) | Nº máximo de OBSselect (10.5%) |
|---------------|---------------|------------------------------|-------------------------------|--------------------------------|
| 1 | 1000 (50%) | 100 | 95 | 105 |
| 2 | 600 (30%) | 60 | 57 | 63 |
| 3 | 400 (20%) | 40 | 38 | 42 |
| Total | 2000 (100%) | 200 | 190 | 210 |

El conjunto de observaciones seleccionadas se denomina *OBSinicial*. Así, las matrices D_i y D_o son subconjuntos de la base de datos cuyas columnas son las variables de entrada y salida respectivamente, y las filas son aquellas observaciones seleccionadas que están dadas por el conjunto *OBSinicial*.

Con las matrices D_i y D_o se genera el primer modelo SBM con un 4% de las observaciones. Al comparar las salidas reales con las salidas estimadas se obtiene la matriz de residuos E . El indicador de Hotelling se obtiene a partir de la matriz de residuos en el espacio de componentes principales. A través de este indicador, se incluyen observaciones no consideradas originalmente en el conjunto *OBSinicial* hasta completar un 10% de la base de datos, obteniéndose con ello un nuevo conjunto de observaciones *OBSpreliminar*, y las matrices D_i y D_o . El criterio de incorporación de observaciones es agregar aquellas que dentro de una ventana de tamaño fijo, se encuentren más cerca y por sobre el umbral de Hotelling. En la Figura 3.3 se observa el indicador de Hotelling, donde las observaciones en color verde son del conjunto *OBSinicial*, en rojo las que serán agregadas, en gris las que no estarán en *OBSpreliminar* y en azul el umbral de Hotelling.

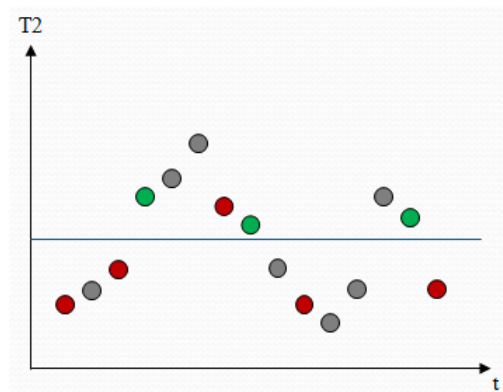


Figura 3.3: Selección de observaciones por medio del indicador de Hotelling.

Además, existe el criterio del promedio del error medio cuadrático para medir el desempeño de un modelo. Consiste en calcular el indicador R en cada uno de los modelos que son generados. Se dirá que un modelo tendrá mejor desempeño que otro si su valor de R es menor. Así, las observaciones solo se actualizarán si el desempeño del nuevo modelo es mejor que el del anterior. Este indicador R se calcula a partir de la matriz de errores $E \in \mathbb{R}^{N \times M}$ como:

$$R = \frac{1}{N} \sum_{i=1}^N S_i \quad (3.2)$$

Donde la componente i -ésima del vector S es:

$$S_i = \frac{1}{M} \sum_{j=1}^M E_{ij}^2 \quad (3.3)$$

3.1.2 ETAPA 2: MODELO PRELIMINAR

Las matrices D_i y D_o obtenidas en la Sección 3.1.1 se utilizan para generar un modelo SBM preliminar con la totalidad de observaciones (10% del total de la base de datos). Este modelo es solo preliminar ya que las etapas siguientes se encargan de optimizar y reajustar dicho modelo. A partir de las salidas estimadas a través de este modelo, se calcula la matriz de residuos como la diferencia entre estas salidas y las salidas reales del proceso.

Por otro lado, las observaciones del conjunto $OBS_{inicial}$ se ordenan de acuerdo al indicador de Hotelling de mayor a menor. Luego, este conjunto se divide consecutivamente en subconjuntos P_i con $i = 1 \dots 4$ de igual tamaño, los cuales serán utilizados por la Etapa 3 de optimización en la Sección 3.1.3.

3.1.3 ETAPA 3: FASE ITERATIVA DE OPTIMIZACIÓN

La fase de optimización se compone de 4 iteraciones que consisten en eliminar uno a uno los subconjuntos P_i de observaciones iniciales (Figura 3.4a), y reemplazarlos por nuevas observaciones A_i (Figura 3.4b) según el criterio del indicador de Hotelling visto en las etapas previas. Las razones de esta eliminación son debido a que algunas de estas observaciones son redundantes en cuanto a la información que entregan del proceso, o porque no son realmente representativas de las operaciones del sistema.

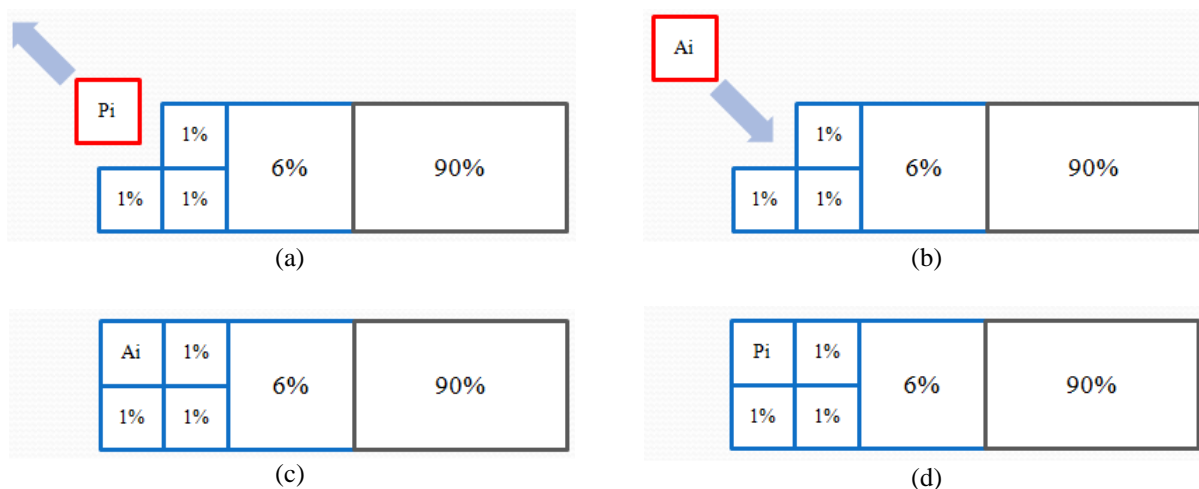


Figura 3.4: Etapa 3 de optimización del conjunto de observaciones.

Con este nuevo conjunto de observaciones (Figura 3.4c) que incluye a A_i , se genera un nuevo modelo SBM, y con ello se calcula la matriz de residuos y el indicador de Hotelling. A continuación, se compara el desempeño de este nuevo modelo con el anterior por medio del criterio del promedio del error cuadrático medio a través del indicador R explicado en la Sección 3.1.1. Las observaciones $OBS_{iteración}$ elegidas serán las de la Figura 3.4c o Figura 3.4d, dependiendo de cuál de esos modelos tiene un mejor desempeño.

A partir del modelo con mejor desempeño se calcula la matriz de residuos E_{Train} y con (3.3), el vector del error cuadrático medio S_{Train} .

3.1.4 ETAPA 4: REAJUSTE DEL MODELO

La Etapa 4 de la herramienta de generación de modelos basados en similitud es opcional, y consiste en agregar un número fijo de aquellas observaciones donde el vector S_{Train} es mayor. Para ello, se determinan dos variables:

- $OBS_{agregar}$: es el máximo número de observaciones que serán agregadas en esta etapa.
- $OBS_{intervalo}$: es el número de observaciones que no podrán ser agregadas antes y después de una observación perteneciente a $OBS_{agregar}$.

La variable $OBSintervalo$ está bajo el supuesto que si se agrega al modelo una observación con un alto error cuadrático medio, entonces las que están a su alrededor también debieran disminuir su error sin que sean incorporadas al modelo. Así, en el reajuste se incorporan al modelo aquellas observaciones con mayor error cuadrático medio, y se van descartando las que se encuentran en su vecindad.

La Figura 3.5 es un ejemplo donde se considera que $OBSagregar = 3$ y $OBSintervalo = 3$. En verde se observan las observaciones que han sido consideradas para el modelo, en azul la que fue incorporada en cada iteración, y en rojo las que han sido descartadas por ser adyacentes.

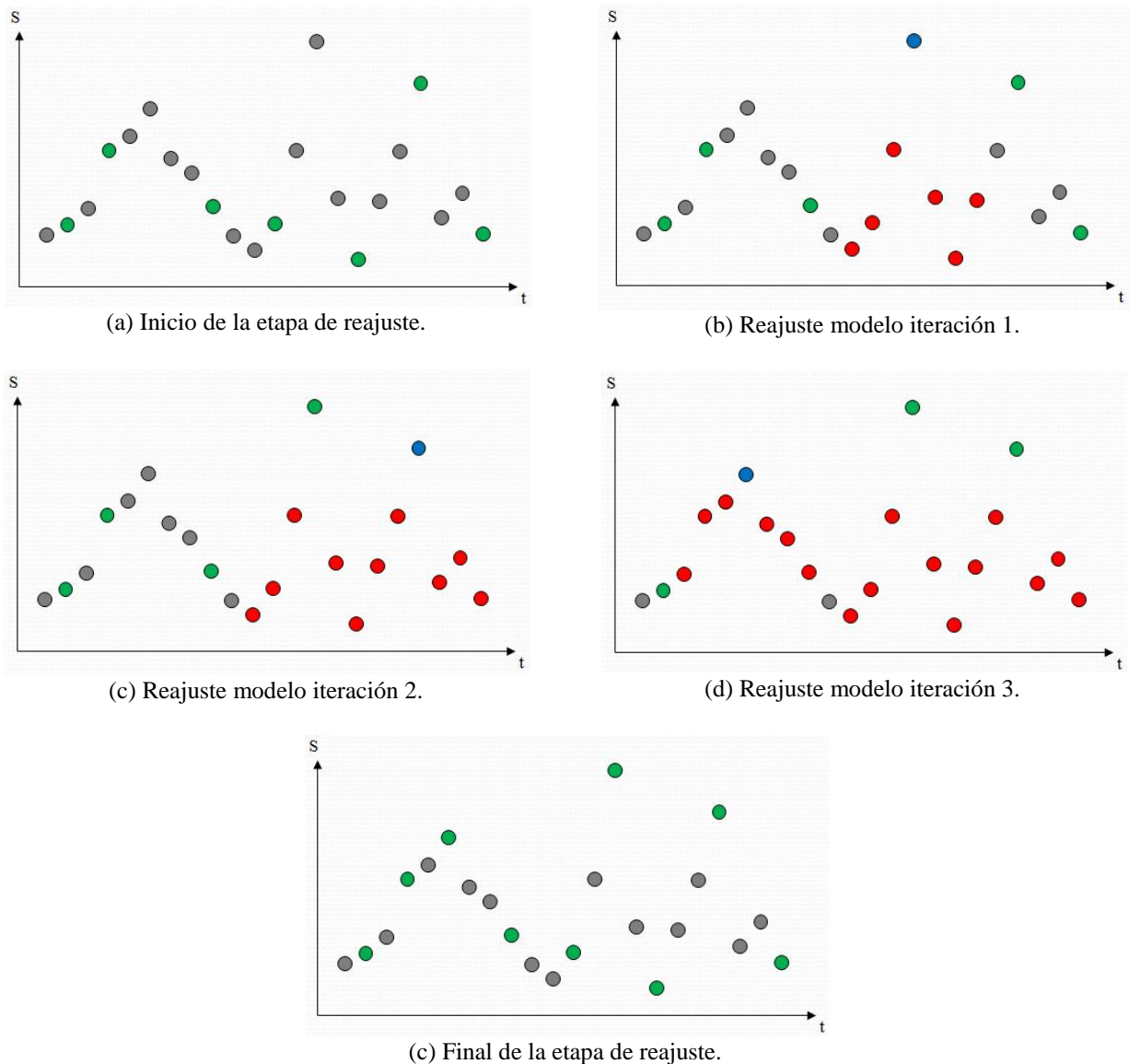


Figura 3.5: Selección de observaciones en la etapa de reajuste del modelo.

Con las observaciones *OBSfinal* ya determinadas, es posible obtener las salidas de la función *ModeloSBM(.)*: las matrices *Di* y *Do* propias del modelo; la matriz de residuos *ETrain* como la diferencia entre las salidas reales y las salidas estimadas; y la media, desviación estándar, valores propios y vectores propios de la matriz *ETrain*.

3.2 MODELO SBM CLASIFICADOR

El modelo SBM clasificador es utilizado en la Fase I del sistema supervisor FDI, para determinar si se está o no en presencia de algún evento anómalo, y así generar una señal de estado normal o alarma según sea el caso. Este modelo solo cumple el objetivo de advertir que algo extraño sucede en el proceso, sin indicar cuál es la anomalía.

La metodología para obtener este modelo, se basa en utilizar la función *OfflineClass(.)* cuyas entradas y salidas se describen en la Tabla 3.3. Tal como se observa en la Figura 3.6, esta función está contemplada en 3 etapas que se explicarán en la Sección 3.2.1, Sección 3.2.2 y Sección 3.2.3 respectivamente. La Etapa 1 consiste en transformar el espacio de datos en un espacio de componentes principales por medio de PCA, y así seleccionar un porcentaje de aquellas observaciones más representativas de cada punto de operación. En la Etapa 2, se genera un modelo SBM cuyas entradas son variables de la base de datos y la salida es un vector en el que sus componentes indican la agrupación en la que se encuentra el sistema en cada instante. Finalmente, en la Etapa 3 se escoge un umbral de modo tal que se produzcan alarmas cuando ocurren las anomalías registradas y disminuir el número de falsos positivos.

Tabla 3.3: Argumentos de la función *OfflineClass*(·).

| | Nombre | Descripción |
|-----------------|-----------------------|---|
| Entradas | <i>XClass</i> | Matriz cuyas columnas son las variables de entrada del modelo SBM clasificador, y sus filas son las observaciones del proceso. |
| | <i>OBSpercent</i> | Es el porcentaje del total de observaciones que se utilizan para generar el modelo SBM clasificador. Típicamente: <i>OBSpercent</i> = 10%. |
| Salidas | <i>Di</i> | Matriz de $N \times M_1$ característica del modelo SBM. Las filas son las observaciones seleccionadas y las columnas son las variables de entrada del modelo. |
| | <i>Do</i> | Matriz de $N \times M_2$ característica del modelo SBM. Las filas son las observaciones seleccionadas y las columnas son las variables de salida del modelo. |
| | <i>meanClass</i> | Media de la matriz de datos <i>XClass</i> . |
| | <i>sdClass</i> | Desviación estándar de la matriz de datos <i>XClass</i> . |
| | <i>vpClass</i> | Vectores propios que caracterizan el espacio de componentes principales. |
| | <i>Centroid</i> | Matriz de $N \times 2$ donde N corresponde al número de centroides en el espacio de componentes principales. |
| | <i>Classthreshold</i> | Vector con los umbrales de decisión de cada centroide. |

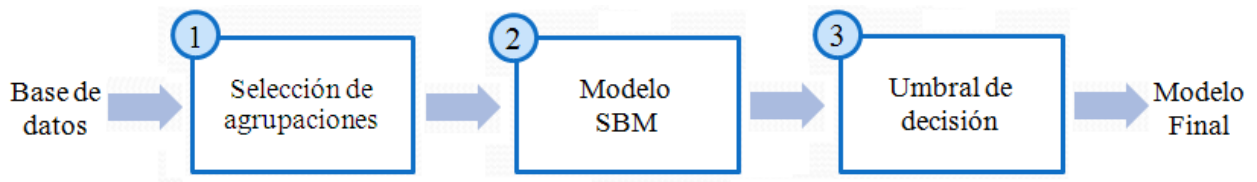


Figura 3.6: Etapas para la generación del Modelo SBM Clasificador.

Los datos de entrada $XClass$ del modelo clasificador son observaciones de aquellas variables utilizadas por los N modelos de anomalías registradas. Es decir, si el i -ésimo modelo M_i utiliza un conjunto $XModel_i$ como datos de entrada y un conjunto $YModel_i$ como datos de salida, entonces los datos utilizados para generar el modelo clasificador son:

$$XClass = \bigcup_{i=1}^N \{XModel_i \cup YModel_i\} \quad (3.4)$$

Las salidas utilizadas para generar el modelo son obtenidas en la Etapa 1 del algoritmo, en la Sección 3.2.1. Con respecto a las observaciones, estas deben ser con el sistema en operación normal. Se excluyen así, todos aquellos instantes donde el sistema se encuentra en presencia de alguna anomalía.

3.2.1 ETAPA 1: SALIDAS DEL CONJUNTO DE ENTRENAMIENTO

La Etapa 1 de la herramienta de generación de modelos clasificadores basados en similitud tiene por objetivo obtener la variable de salida que será utilizada para generar el modelo SBM clasificador.

El espacio de datos $XClass$ (Figura 3.2a) con tantas dimensiones como variables de entrada es reducido con PCA a uno de dos o tres dimensiones (Figura 3.2b). Con en el espacio de componentes principales se establece cualitativamente el número de agrupaciones. En la Figura 3.2b se observan tres agrupaciones, destacadas con color rojo, gris y azul. La clasificación se realiza con la técnica de *agrupación por K-medias* con la función de Matlab *k-means*, que entrega las coordenadas de los centros de cada agrupación en la matriz *Centroid* y un entero $XoutAux$ que determina la agrupación a la que pertenece cada observación.

El vector de enteros $XoutAux$ que establece la agrupación a la que pertenece cada observación seleccionada, se transforma a una matriz binaria $XoutClass$ con tantas columnas como centroides tenga el sistema. La transformación para la observación i -ésima es:

$$XoutAux(i) = n \rightarrow XoutClass(i, j) = \begin{cases} 1 & j = n \\ 0 & j \neq n \end{cases} \quad (3.5)$$

A continuación, se seleccionan aquellas observaciones que se encuentran en torno a una circunferencia centrada en cada centroide. La distribución de estos datos va de acuerdo al tamaño que tiene cada agrupación. Además el porcentaje de observaciones seleccionadas *OBSselect* está dado por la variable *OBSpercent*, cuyo valor típico es 10%. Sin embargo, debido a que se trazan circunferencias, hay un margen de error de $\pm 5\%$ al ser imposible en algunos casos alcanzar un porcentaje exacto.

En la Tabla 3.2, se observa el número de observaciones asumiendo esa cantidad de elementos en cada agrupación de la Figura 3.2b.

Las matrices *Di* y *Do* corresponden a las matrices de datos *XClass* y *XoutClass* en las observaciones seleccionadas *OBSselect*. Finalmente, en la etapa online del sistema supervisor son necesarios la media *meanClass* y la desviación estándar *sdClass* de los datos, y los vectores propios *vpClass* para poder transformar al mismo espacio de componentes principales los nuevos datos ingresados. También se requieren las coordenadas de los centros *Centroid* para determinar a que agrupación pertenecen los nuevos datos.

3.2.2 ETAPA 2: GENERACIÓN DEL MODELO SBM

La Etapa 2 consiste en generar, a través de las matrices *Di* y *Do* obtenidas en la Sección 3.2.1, el modelo SBM clasificador utilizando la herramienta de generación de modelos explicada en la Sección 3.1. Es importante destacar que de la herramienta explicada en la Sección 3.1 solo se utiliza la función *SBM(·)* con el objetivo de obtener las salidas estimadas *XoutClassSBM*, y no el resto del algoritmo que optimiza las matrices *Di* y *Do*. La razón de esto, es que las variables de salida que han sido generadas a través de agrupaciones no necesariamente representan la realidad, principalmente aquellas que se encuentran cerca de dos o más agrupaciones. Es por ello que se han determinado como las más representativas las que se encuentran más cerca de los centros, asegurando así, que efectivamente pertenezcan a dicha agrupación. Además, las salidas estimadas son filtradas con la mediana móvil para efectos de interpretar mejor los resultados.

3.2.3 ETAPA 3: DETERMINACIÓN DEL UMBRAL DE DECISIÓN

Finalmente, en esta etapa se determina un umbral de decisión *Classthreshold* que establece cuando el sistema se encuentra en condición normal o anómala.

Las componentes del vector *Classthreshold* corresponden al umbral de decisión en cada una de las agrupaciones encontradas anteriormente. Para obtener *Classthreshold(i)*, se seleccionan en primer lugar, por medio de la matriz *XoutClass*, todas aquellas observaciones en las cuales el sistema se encuentra en la agrupación *i*-ésima. A continuación, se obtiene el histograma de cada columna de *XoutClassSBM*, tal como se observa en la Figura 3.7. Hay que recordar que cada columna de esta matriz está asociada a una agrupación de la base de datos en el espacio de componentes principales.

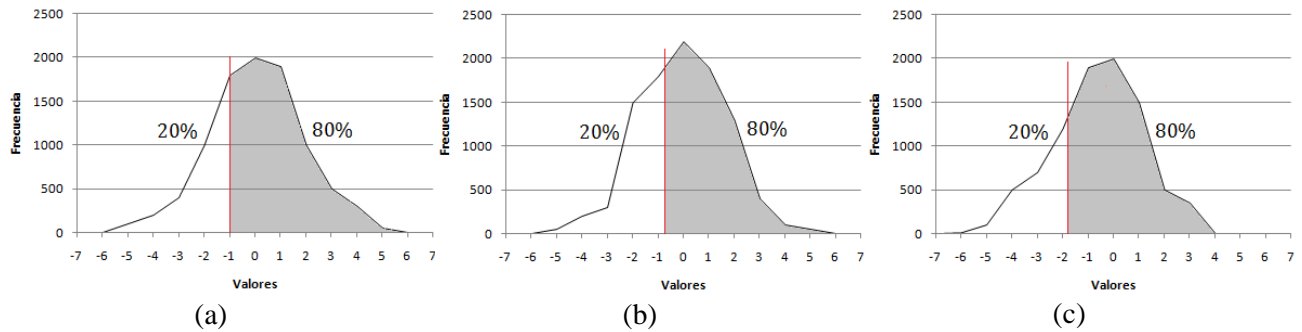


Figura 3.7: Histogramas para determinar el umbral de decisión *Classthreshold*.

Al recorrer el histograma de izquierda a derecha, se determina que el umbral es aquel en el que se alcanza cierto porcentaje del total de datos. En el sistema analizado se ha determinado que este porcentaje sea del 20%. Sin embargo, este es un parámetro de diseño que de ser necesario, puede ser reajustado acorde a cada caso.

Los algoritmos descritos en este capítulo son herramientas utilizadas para generar el modelo clasificador y los modelos de anomalías registradas para un conjunto de variables de entrada y salida previamente determinado. Para obtener este conjunto, se ha diseñado una metodología de estudio de eventos que se describe a continuación en el Capítulo 4. Además, en este capítulo se describe el sistema supervisor utilizado para detectar y aislar las anomalías en estudio.

CAPÍTULO 4: IMPLEMENTACIÓN DE SISTEMA SUPERVISOR FDI UTILIZANDO MODELOS BASADOS EN SIMILITUD

Este capítulo describe la implementación de un sistema supervisor para detección y aislamiento de fallas en base a las herramientas de generación de modelos basados en similitud (SBM) que fueron explicados en el Capítulo 3. El sistema supervisor utiliza información histórica del proceso para aprender de las anomalías que ya han ocurrido, y en base a ese estudio, es posible detectar esos futuros eventos. De este modo, en la Sección 4.1 se introduce el sistema supervisor con las etapas que lo componen. Las fases del sistema supervisor requieren de un análisis preliminar ejecutado fuera de línea, y que es explicado en la Sección 4.2: estudio de las variables involucradas, análisis de los instantes con operación normal y anómala, y diseño de un modelo clasificador y de modelos generadores de residuos. Finalmente, la Sección 4.3 describe las tres fases del sistema supervisor: clasificación para detección de estados anormales, generación de residuos y análisis de residuos.

4.1 ESQUEMA GENERAL DEL SISTEMA SUPERVISOR PARA ESTUDIO DE EVENTOS

El sistema supervisor de detección y aislamiento de fallas (Figura 4.1) es una metodología que consiste en determinar, a partir de un conjunto de variables, si el proceso en estudio se encuentra en un modo de operación normal o no, y si no lo está, determinar cuál es la causa dentro de un conjunto de posibilidades, dadas por los modelos de anomalías registradas previamente generados. Una anomalía es considerada registrada si se dispone de algún documento, o los operadores conocen al menos las consecuencias locales que tiene en el sistema. Por ejemplo en un sistema de generación en una central termoeléctrica: disminución en la potencia, aumento en alguna temperatura, cambio en la velocidad de rotación de alguna turbina, etc. Es importante destacar que este sistema requiere un análisis y un trabajo previo fuera de línea del proceso en estudio: estudiar las variables involucradas, análisis de instantes con operación normal y

anómala, creación de modelo clasificador y modelos generadores de residuos para cada anomalía en estudio, y determinación de umbrales de decisión.

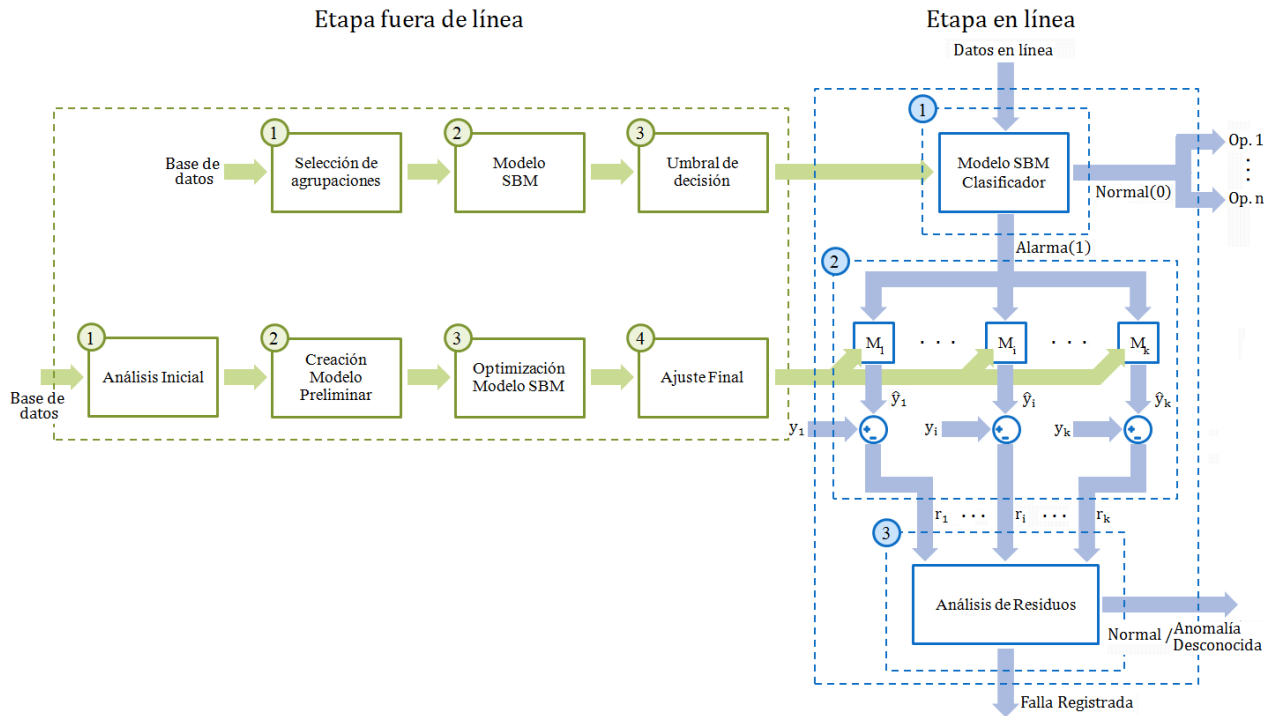


Figura 4.1: Esquema del sistema supervisor FDI.

Tal como se aprecia en la Figura 4.1, este esquema consta de la etapa fuera de línea a través de la generación del modelo clasificador y de los modelos de anomalías registradas. Estos modelos son utilizados posteriormente en la etapa en línea para verificar si existen diferencias en el proceso con respecto a una condición normal que puedan derivar en alguna de las anomalías ya registradas con anterioridad. El sistema supervisor requiere de un modelo clasificador y de al menos un modelo de anomalías registradas por cada uno de los eventos que se quieren detectar. En la aplicación del sistema supervisor en una turbina a gas en el Capítulo 5, se utiliza un modelo clasificador y tres modelos para explicar dos anomalías registradas (el tercer modelo proviene de modelar con las observaciones en que ocurre la anomalía).

La etapa en línea consiste en tres fases que transcurren de manera sucesiva y que son explicadas en la Sección 4.3.1, Sección 4.3.2 y Sección 4.3.3. En la Fase I, las variables ingresan a un modelo clasificador que tiene por finalidad determinar en qué punto de operación está operando el proceso en base a un cierto umbral de decisión, determinado en la etapa fuera de línea. Si el sistema no se encuentra en alguno de los puntos de operación normal, entonces se produce una señal de alarma con la que se activan los modelos M_i para generar los residuos, y que corresponde a la Fase II del sistema supervisor. Finalmente, en la Fase III se realiza un análisis sobre estos residuos para determinar si se está en presencia de alguna de las fallas registradas, si corresponde a un evento anómalo desconocido, o si el sistema se encuentra operando de forma normal y se había producido una falsa alarma en la Fase I.

El sistema supervisor tiene un carácter dinámico. Esto significa que tiene la facultad de ser modificado conforme a las necesidades que vaya teniendo el proceso. Las principales razones que obligan a actualizar el sistema son el registro de una nueva anomalía y los cambios físicos del proceso. En ambos casos, se debe realizar nuevamente el estudio previo del sistema: verificar que el modelo clasificador entrega una alarma para todas las anomalías y el análisis de las variables del nuevo modelo.

4.2 METODOLOGÍA DEL ESTUDIO DE EVENTOS

La metodología del estudio de eventos corresponde a la etapa fuera de línea del sistema supervisor y se encarga del estudio histórico del proceso. De este modo, se generan los modelos con las herramientas vistas en el Capítulo 3 y que son utilizados en la etapa en línea para detección e identificación de anormalidades. Esta metodología también se enmarca en el proyecto de generación de modelos basados en similitud mencionado anteriormente.

La metodología no solo requiere bases de datos con información actualizada del proceso, sino también datos históricos que permitan mostrar el comportamiento que ha tenido el sistema anteriormente. Estos datos históricos son de suma importancia en todos los modelos generados, ya que se deben analizar aquellos instantes donde el sistema opera normalmente y conjuntamente cuando se han registrado anomalías. Además, si ocurren cambios en el proceso o en la operación, se requiere de datos actualizados del sistema ya sea para validar o de ser necesario, actualizar los modelos generados. Por otro lado, un supuesto de la metodología es que las bases de datos no contienen información errónea por problemas con los sensores.

A pesar que el sistema supervisor ha sido desarrollado para procesos multivariables, no todas las variables disponibles son útiles en cada modelo. Por lo tanto, se hace necesario realizar un análisis previo para reducir el conjunto de variables, descartando aquellas que fallan constantemente y que se encuentren repetidas. De esta forma, se logra eliminar la información poco relevante y redundante. El conocimiento de los operadores del proceso y los informes de eventos registrados juegan un rol trascendental en esta etapa del estudio. Esto debido a que están siempre monitoreando aquellas variables que son más relevantes en cada anomalía y al registro que hay sobre los instantes en que estos eventos han ocurrido. Además, hay otros métodos de reducción de variables, que son explicados en la Sección 4.2.2 y que fueron posteriormente aplicados en el proceso en estudio en el Capítulo 5.

El sistema supervisor en línea utiliza el modelo clasificador para la Fase I y luego los modelos de cada anomalía registrada para la Fase II. Sin embargo, en la etapa fuera de línea del sistema supervisor, los modelos de anomalías registradas se generan antes que el modelo clasificador debido a que las variables involucradas en este último están relacionadas con las utilizadas por los modelos de anomalías registradas.

Tal como se observa en la Figura 4.2, los modelos de anomalías registradas pueden no solo representar al sistema en un estado de operación normal, sino que pueden modelar aquellos instantes en que en el proceso hay presencia de algún evento anómalo. En la Figura 4.2a se tiene un modelo del sistema en operación normal. Esto se traduce en que cuando el indicador de residuos esté por debajo del umbral establecido, entonces el sistema se encuentra en estado normal, y en caso contrario, en estado anómalo. El caso de la Figura 4.2b es completamente opuesto ya que se modela al sistema cuando existe una anomalía en particular. Así, cuando esta anomalía se encuentra presente en el sistema, entonces el indicador de residuos está por debajo del umbral, y en caso contrario, el sistema se encuentra en una condición normal.

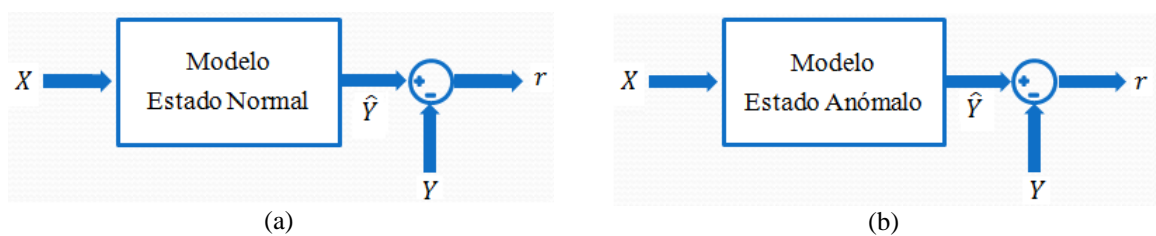


Figura 4.2: Tipos de modelos de anomalías registradas; (a) modelo del proceso en operación normal; (b) modelo del proceso operando en la anomalía.

Es importante destacar que tanto los modelos del proceso en estado normal, como en estado anómalo, solo explican un tipo de anomalía. Por lo tanto, si se tiene conocimiento de N anomalías en el proceso, entonces, el sistema supervisor requiere al menos N modelos de anomalías, independiente si representan al sistema en operación normal, o en operación anómala. Por otro lado, es necesario que el modelo clasificador sea capaz de detectar todas las anomalías debido a que los resultados que entrega son concluyentes sobre si el sistema supervisor pasa la Fase II o no.

La generación de cada uno de estos modelos consta de 6 etapas que van desde la recopilación de los datos hasta la obtención del modelo que será utilizado posteriormente en la Fase II del sistema supervisor en línea. Tal como se observa en la Figura 4.3, los parámetros del modelo (variables de entrada y salida, umbral de Hotelling) son reajustados de acuerdo a los resultados obtenidos en las etapas de validación y de prueba en la anomalía.

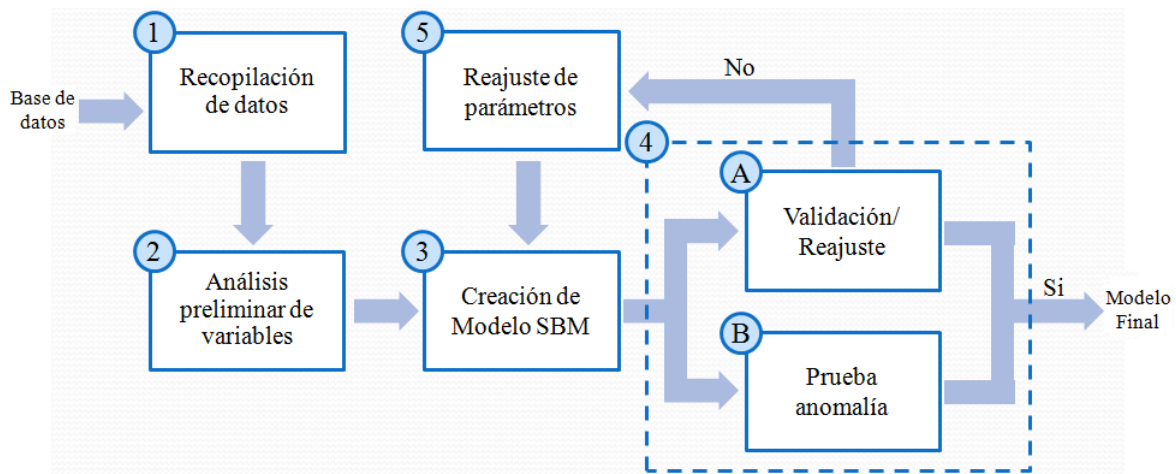


Figura 4.3: Metodología para la generación de modelos de anomalías registradas.

4.2.1 ETAPA 1: RECOPIACIÓN DE DATOS

La generación del modelo clasificador y de los modelos de anomalías requiere de datos históricos del proceso que sean abundantes en información. Esto quiere decir que la base de datos debe abarcar los diferentes puntos de operación del proceso y además disponer de mediciones de la anomalía registrada a explicar por dicho modelo. El tiempo de muestro es otra arista importante para analizar debido a que cada anomalía se prolonga en diferentes cantidades de tiempo. Las bases de datos corresponden a mediciones de múltiples variables, y por lo tanto, se representan como matrices de $N \times M$ (Figura 4.4), donde N corresponde al número de muestras y M el número de variables del proceso.

| | X_1 | X_2 | ... | X_M |
|------------|------------|------------|-----|------------|
| $X_1(t_1)$ | $X_1(t_1)$ | $X_2(t_1)$ | ... | $X_M(t_1)$ |
| $X_1(t_2)$ | $X_1(t_2)$ | $X_2(t_2)$ | ... | $X_M(t_2)$ |
| ... | ... | ... | ... | ... |
| $X_1(t_N)$ | $X_1(t_N)$ | $X_2(t_N)$ | ... | $X_M(t_N)$ |

Figura 4.4: Ejemplo de una base de datos utilizada para generar modelos.

Las bases de datos se dividen en los siguientes tres conjuntos de datos.

- Datos de entrenamiento: son utilizados para generar los modelos del proceso. En los modelos utilizados en la Fase II del sistema supervisor, estos pueden ser con el proceso en estado normal o anómalo, dependiendo del tipo de modelo que se desee obtener.
- Datos de validación: son utilizados para validar los modelos en la Etapa 4A (Sección 4.2.4). Deben contener al menos las operaciones que hay en los datos de entrenamiento. Los datos de validación no deben usarse para entrenar el modelo, sino solo para verificar su respuesta frente a estos nuevos datos. Estos datos cumplen el objetivo de evaluar el desempeño de la modelación.
- Datos de prueba en la anomalía: el estado de estos datos deben diferir a los de entrenamiento y validación. Es decir, si el modelo explica el sistema funcionando en estado normal, entonces estos datos deben ser del sistema funcionando en estado anómalo. Por el contrario, si el modelo se genera para explicar una anomalía, entonces estos datos deben ser del sistema operando de forma normal. Son utilizados en la Etapa 4B (Sección 4.2.5). Estos datos cumplen el objetivo de evaluar el desempeño de la detección.

4.2.2 ETAPA 2: ANÁLISIS PRELIMINAR DE VARIABLES

Esta etapa se encarga de determinar cuáles serán las variables de entrada y salida que serán utilizadas preliminarmente para generar los modelos de las anomalías registradas. Las variables de entrada del modelo clasificador se obtienen también en esta etapa como la unión de las variables de entrada y salida de todos los modelos de anomalías registradas, tal como se explica en la Sección 3.2. Posteriormente, en la Sección 4.2.6 se realiza una selección de variables más detallada en base a los resultados de la Etapa 4A y Etapa 4B. El análisis preliminar de variables se realiza con la base de datos en su totalidad. Así, se consideran datos del sistema en operación normal y también en presencia de la anomalía que quiere ser detectada con este modelo.

El objetivo de esta etapa es seleccionar aquellas variables que por los tres motivos que se describen a continuación, debiesen ser incorporadas o descartadas en el modelo:

- En base a un análisis cualitativo, se eliminan aquellas variables que en la base de datos tienen un comportamiento constante en los diferentes puntos de operación, inclusive en la anomalía. Lo mismo ocurre con aquellas variables que se encuentran repetidas.
- Los operadores son fundamentales en esta etapa de la metodología, ya que con su conocimiento se determinan aquellas variables que son más importantes, que están siempre monitoreando, y por tanto debiesen incluirse en el modelo.
- Los datos del conjunto de variables de entrada/salida que fueron seleccionados con los dos criterios antes mencionados, son analizados con regresión PLS. Para ello, se grafican las primeras dos componentes de los vectores de carga y luego se obtienen las distancias de cada una de las variable al origen. Se eliminan aquellas variables que están ubicadas más cerca del origen, ya que no pueden ser explicadas por los vectores de carga, y por lo tanto no debiesen aportar a la generación del modelo.

Como consecuencia de esta eliminación de variables, adicionalmente se produce una disminución en los tiempos de cómputo de los algoritmos de generación de modelos basados en similitud.

4.2.3 ETAPA 3: CREACIÓN DE MODELOS

Con las variables escogidas y los instantes de entrenamiento definidos se procede a la creación del modelo con las herramientas de generación de modelos descritos en el Capítulo 3. Tal como se observa en el diagrama de la Figura 4.3, se recurre a esta etapa tantas veces como se actualicen las variables de entrada y de salida en la Etapa 5 de reajuste de parámetros.

4.2.4 ETAPA 4A: VALIDACIÓN DEL MODELO

Una vez que el modelo ha sido generado en la Sección 4.2.3 con las variables de entrada y salida antes determinadas, se utilizan los datos de validación para comprobar el desempeño de la modelación frente a nuevos datos. Este desempeño se mide a través del indicador de Hotelling, comparándolo cualitativamente con el obtenido con los datos de entrenamiento. Las estimaciones de las variables de salida son utilizadas en la Sección 4.2.6 para reajustar si son necesarios el umbral de Hotelling y las variables del modelo.

4.2.5 ETAPA 4B: PRUEBA DE EFICACIA DEL MODELO EN LA DETECCIÓN DE LA ANOMALÍA

En esta etapa se estiman las salidas del modelo con los datos de prueba en la anomalía. En el modelo clasificador se observa el comportamiento de estas salidas y del umbral de decisión. En los modelos de anomalías registradas se calculan además los residuos. El objetivo de esta etapa es que en los instantes en que ocurre una anomalía esta logre ser detectada ya sea con el indicador de Hotelling en los modelos de anomalías registradas, o con las salidas estimadas en el modelo clasificador, y con la utilización de los umbrales de decisión. Si no se cumple con los objetivos de detección, se realizan los reajustes descritos en la Sección 4.2.6. El modelo final es aquel que cumple con los objetivos de modelación y detección planteados en la Etapa 4A y Etapa 4B, y en tal caso, no será necesario realizar más reajustes. Finalmente, cabe destacar que los resultados son filtrados con una mediana móvil para una mejor interpretación de estos.

4.2.6 ETAPA 5: REAJUSTE DE PARÁMETROS

Esta etapa se encarga de reajustar los umbrales de decisión que limitan si se genera o no una alarma en el proceso, aumentándolos o disminuyéndolos dependiendo si se desea aumentar la capacidad de modelación o de detección.

En el caso del modelo clasificador, además se agregan observaciones al conjunto de entrenamiento que en las etapas previas no fueron modeladas correctamente. De este modo, se espera que el error disminuya y con ello, la modelación tenga un mejor desempeño.

Con respecto a los modelos de anomalías registradas, esta etapa se encarga de eliminar cualitativamente aquellas variables de entrada o de salida que no aportan a la modelación ni a la detección de las anomalías. Con esta nueva selección de variables se regresa a la Etapa 3 para generar un nuevo modelo. Esta selección utiliza los residuos obtenidos en la Sección 4.2.4 y Sección 4.2.5 con los datos de validación y de prueba en la anomalía respectivamente. Los datos de validación se utilizan para eliminar aquellas variables de salida que no fueron estimadas correctamente. Para ello, los residuos obtenidos con los datos de validación se comparan con los obtenidos con los datos de entrenamiento, y se observan si hay cambios significativos. Los datos de prueba en la anomalía se utilizan para eliminar aquellas variables que no aportan en la detección de la anomalía al no revelar cambios entre los instantes de operación normal y aquellos en que ocurre la anomalía.

4.3 SISTEMA SUPERVISOR EN LÍNEA

Como ya se mencionó, el sistema supervisor consta de tres fases. Estas se ejecutan con la función de MATLAB *systemcondition*(·) cuyas entradas y salidas se describen en la Tabla 4.1.

Esta función simula al proceso real recibiendo uno a uno los nuevos vectores de datos del sistema X_n para los cuales se desconoce su estado. Además del estado en el que se encuentra el sistema, esta función entrega las salidas del modelo clasificador y el indicador de Hotelling obtenido en cada modelo. Adicionalmente, a las señales obtenidas se les aplica un filtro exponencial para interpretar mejor los resultados.

Tabla 4.1: Argumentos de la función *systemcondition*(·).

| | Nombre | Descripción |
|-----------------|------------------------|--|
| Entradas | <i>XClass</i> | Vector de nuevos datos del sistema utilizado como entrada del modelo clasificador. |
| | <i>XModel</i> | Matriz de celdas donde $XModel(i, j)$ es un vector de nuevos datos del sistema utilizado como entrada ($j = 1$) o salida ($j = 2$) del i -ésimo modelo de anomalías. |
| | <i>ClassifierModel</i> | Estructura del modelo clasificador. |
| | <i>FaultModel</i> | Vector de celdas donde $FaultModel(i)$ contiene la estructura del i -ésimo modelo de anomalías registradas. |
| | <i>FaultThreshold</i> | Vector donde $FaultThreshold(i)$ contiene el umbral de Hotelling del i -ésimo modelo de anomalías registradas. |
| Salidas | <i>XoutClass</i> | Salida del modelo clasificador utilizando centroides. |
| | <i>XoutClassSBM</i> | Salida del modelo SBM clasificador. |
| | <i>THotelling</i> | Vector donde $THotelling(i)$ es el índice de Hotelling obtenido a partir de los residuos del i -ésimo modelo de anomalías registradas. |
| | <i>stateClass</i> | Escalar binario con el estado del proceso al final de la Fase I. |
| | <i>stateFault</i> | Escalar discreto con el estado del proceso al final de la Fase III. |

4.3.1 FASE I: MODELO SBM PARA DETECCIÓN DE ESTADOS ANORMALES

En esta fase se verifica si los nuevos datos ingresados se encuentran en alguna de las operaciones normales ya conocidas o en alguna operación desconocida hasta el momento, y que podría llegar a ser una anomalía. Es por ello que esta etapa genera una alarma de precaución y no una de falla, ya que aún se desconoce si este evento corresponde o no una anomalía ya registrada. Se utiliza el modelo clasificador explicado anteriormente en la Sección 3.2, el cual viene disponible en la estructura *ClassifierModel*. Esta fase se divide en dos etapas: obtener las salidas del modelo clasificador y determinar con el umbral clasificador *Classthreshold* si el sistema presenta alguna anormalidad.

4.3.1.1 SALIDAS DEL MODELO SBM CLASIFICADOR

En esta etapa se obtienen las salidas estimadas del modelo clasificador utilizando el vector de datos de entrada $XClass$, la función $SBM(\cdot)$ y las matrices Di y Do propias del modelo. Estas salidas se representan como un vector $XoutClassSBM$ cuyo largo es la cantidad de agrupaciones establecidas en la etapa fuera de línea descrita en la Sección 3.2.

A diferencia de los modelos de anomalías registradas, no se dispone de una salida real con la cual comparar las salidas estimadas del modelo SBM y así obtener un residuo para analizar. Esta salida se genera determinando la agrupación cuyo centro esté a la menor distancia del vector $XClass$ en el espacio de componentes principales.

Para ello, $XClass$ se normaliza con la media $meanClass$ y la desviación estándar $sdClass$ (4.1) del modelo clasificador $ClassifierModel$, y luego se transforma con los vectores propios $vpClass$ (2.9) al espacio de componentes principales de los datos de entrenamiento (4.2).

$$\overline{XClass} = \frac{(XClass - \text{eanClass})}{sdClass} \quad (4.1)$$

$$XClass_{PCA} = vpClass * \overline{XClass} \quad (4.2)$$

En el espacio de componentes principales se calcula la distancia del punto P a los centros de cada agrupación. En la Figura 4.5 se tienen tres agrupaciones, donde estas distancias están representadas como d_1 , d_2 y d_3 .

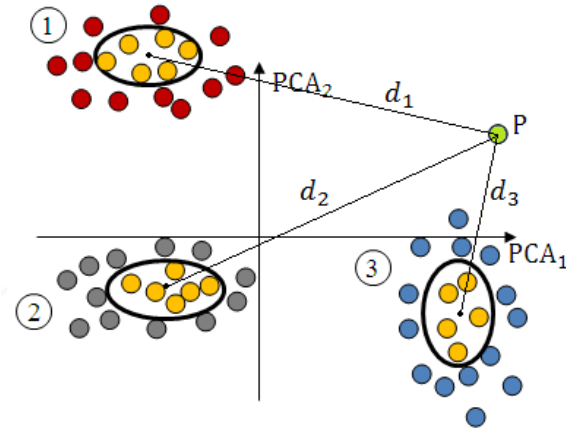


Figura 4.5: Obtención de *XoutClass* en base a la distancia a los centros de cada agrupación.

Así, si se tienen n agrupaciones, entonces:

$$d_i = \min_{j=1 \dots n} d_j \rightarrow P \in i - \text{ésima agrupación} \rightarrow XoutClass = i \quad (4.3)$$

Finalmente, la variable de salida *XoutClass* se transforma en un vector binario según la regla establecida en (3.5).

4.3.1.2 DETERMINACIÓN DEL ESTADO DEL SISTEMA: NORMAL O ALERTA

En esta etapa se determina si el sistema se encuentra en un estado de alarma o no. Para ello, se utilizan las salidas obtenidas en la Sección 4.3.1.1 y el umbral del modelo clasificador *Classthreshold*. El estado del proceso es guardado en un escalar binario *stateClass* que indica si se ha generado o no esta alarma. El análisis se realiza sobre cada componente del vector *XoutClassSBM*, teniéndose dos posibilidades:

- $XoutClass(i) = 1$: si $XoutClassSBM(i)$ se encuentra bajo el umbral *Classthreshold*, entonces se produce alarma
- $XoutClass(i) = 0$: si $XoutClassSBM(i)$ se encuentra sobre el umbral *Classthreshold*, entonces se produce alarma.

4.3.2 FASE II: MODELOS DEL PROCESO PARA OBTENCIÓN DE RESIDUOS

La Fase II del sistema supervisor solo se ejecuta si en la Fase I se declara estado de alarma en el sistema. Esta etapa consiste en determinar las salidas de los modelos de anomalías \hat{y}_i y así obtener los residuos que se utilizarán en la Fase III.

La salida estimada \hat{y}_i del modelo M_i se obtiene con el vector de datos de entrada $x_i = XModelo\{i, 1\}$, la función $SBM(\cdot)$ y las matrices D_i y D_o propias del modelo. Tal como se observa en la Figura 4.6, los residuos del modelo M_i se obtienen como

$$r_i = y_i - \hat{y}_i = Xmodel\{i, 2\} - \hat{y}_i \quad (4.5)$$

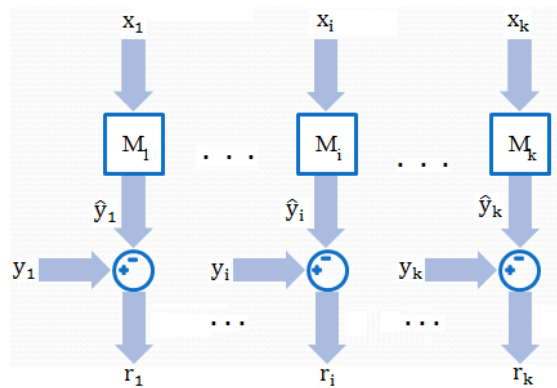


Figura 4.6: Modelos de anomalías registradas, con sus respectivas entradas, salidas y cálculo de residuos.

4.3.3 FASE III: ANÁLISIS DE RESIDUOS PARA DETERMINACIÓN DE FALLA

Finalmente, en la Fase III del sistema supervisor se analizan los residuos obtenidos en la Fase II para determinar el estado del sistema. A pesar de que hay otras alternativas, se decidió utilizar el estadístico de Hotelling como indicador de fallas. Para ello, cuando este indicador supere el umbral *FaultThreshold* obtenido anteriormente en la metodología de estudio de eventos de la Sección 4.2, se dirá que en el proceso hay presencia de una falla. Una alternativa a Hotelling, es utilizar directamente la norma de los residuos y verificar si se supera cierto umbral de decisión. Sin embargo, en ese caso se pierde información debido a que no se tiene en cuenta la variabilidad de los datos.

Con (2.13), (4.7) y (4.8) se calcula el estadístico de Hotelling para cada uno de los residuos, obteniéndose un vector de estadísticos que permitirán concluir sobre lo que podría estar sucediendo en el sistema. Así, para el residuo obtenido a partir del i -ésimo modelo, el estadístico de Hotelling T_i se obtiene como:

$$t_i = r_i * vpFault \quad (4.6)$$

$$T_i = t_i * vlFault^{-1} * t_i^T \quad (4.7)$$

El vector T se compara con el vector de umbrales de Hotelling para cada anomalía registrada (4.9).

$$T = [T_1 \quad \dots \quad T_i \quad \dots \quad T_k] \quad (4.8)$$

$$d = (T - FaultThreshold) > 0 \quad (4.9)$$

El vector d es binario donde $d(i) = 1$ si esa diferencia es mayor a 0, y $d(i) = 0$ en caso contrario. Con este vector se interpretan diferentes conclusiones dependiendo íntegramente del proceso analizado. En general, debe cumplirse lo siguiente:

- Si $d(i) = 0, \forall i$, entonces el sistema no debiese presentar ninguna de las anomalías registradas.
- Si $\exists! i / d(i) = 1$, entonces el sistema posiblemente presenta la falla que representa el modelo i -ésimo.
- En el caso contrario en que hay más de una componente de d igual a 1, entonces hay variadas posibilidades: el sistema presenta múltiples fallas; o solo ocurre una falla y los otros resultados son falsos positivos; o todos los resultados son falsos positivos.

Las conclusiones que se obtienen del vector d son guardadas en una variable escalar llamada *stateFault*.

A continuación, en el Capítulo 5 se procede con una aplicación de los algoritmos y las herramientas descritos en el Capítulo 3 y en el Capítulo 4, para detectar y aislar dos anomalías presentes en una turbina a gas natural de una central térmica de ciclo combinado.

CAPÍTULO 5: APLICACIÓN DEL SISTEMA SUPERVISOR EN LÍNEA EN UNA TURBINA A GAS

El sistema supervisor explicado en el Capítulo 4 se ha diseñado para encontrar anomalías en los procesos industriales. Antes de estar en completo funcionamiento, este sistema considera que las anomalías ya han sido previamente estudiadas con la determinación de las variables que se ven mayormente afectadas por dichas anomalías y a través de la construcción del modelo clasificador y de los modelos de anomalías registradas. Este análisis previo, requiere adicionalmente de información histórica del proceso, en instantes en que el sistema opera de manera normal, y también anómala. Así, para cada una de las anomalías que se desee detectar en línea, se debe disponer de un registro histórico para poder estudiarlas antes de generar los modelos.

En este capítulo se muestran los resultados de la aplicación del sistema supervisor en un proceso industrial consistente en una turbina a gas de la central térmica de ciclo combinado San Isidro. Para entender adecuadamente la metodología que se ha desarrollado, la Sección 5.1 describe el proceso industrial con las principales variables involucradas, y las bases de datos que se encuentran disponibles. La Sección 5.2 se encarga de explicar las dos anomalías que se han estudiado y de generar los modelos de anomalías registradas. En la Sección 5.3 se diseña el modelo clasificador que será utilizado por la Fase I del sistema supervisor. Finalmente, la Sección 5.4 muestra los resultados del sistema supervisor en el proceso industrial en estudio.

5.1 DESCRIPCIÓN DEL PROCESO INDUSTRIAL

En todo el mundo, los procesos industriales requieren de múltiples sistemas de protección, y de detección y prevención de anomalías para evitar que ocurran daños en sus componentes o detenciones no programadas que se traduzcan en altos costos para la industria. En particular en los sistemas de energía eléctrica, los procesos de generación son los más complejos e interesantes de estudiar debido a las diversas máquinas que interactúan y a los costos económicos que implicaría una falla en alguno de estos.

El proceso en estudio consiste en una turbina a gas perteneciente a la termoeléctrica de ciclo combinado San Isidro [51], ubicada a 8[Km] al oeste de la ciudad de Quillota, V Región. Esta central se puso en servicio en el año 1998 y actualmente cuenta con una segunda central, San Isidro 2, también de ciclo combinado. Cada central de ciclo combinado cuenta con dos unidades que han sido fabricadas por *Mitsubishi Heavy Industries*: una turbina a gas y una turbina a vapor. Estas tienen la facultad de operar con gas natural o con petróleo diesel según sean las necesidades y disponibilidades. Además, la potencia instalada es de 379[MW] (240[MW] por la turbina de gas y 139[MW] por la turbina de vapor); y se dispone de una generación media anual de 2045[GWh].



Figura 5.1: Central térmica de ciclo combinado San Isidro [51] [52].

Tal como se ha descrito en los capítulos previos, las herramientas de generación de modelos y el diseño del sistema supervisor requieren de conocimiento del proceso estudiado. A continuación se describen las componentes principales que constituyen la turbina a gas y el modo de funcionamiento.

5.1.1 TURBINA A GAS DE LA CENTRAL TÉRMICA DE CICLO COMBINADO

En una central térmica de ciclo combinado [52] la electricidad se genera mediante la utilización conjunta de dos turbinas: un turbogruppo de gas y un turbogruppo de vapor. Es decir, para la transformación de la energía del combustible en electricidad se superponen dos ciclos termodinámicos:

- Ciclo de Brayton o Joule: está presente en la turbina de gas. El aire se recibe directamente desde la atmósfera y se somete a un calentamiento y compresión para aprovecharlo como energía mecánica o eléctrica.
- Ciclo de Rankine: está presente en la turbina de vapor. Se relaciona el consumo de calor con la producción de trabajo o creación de energía a partir de vapor de agua.

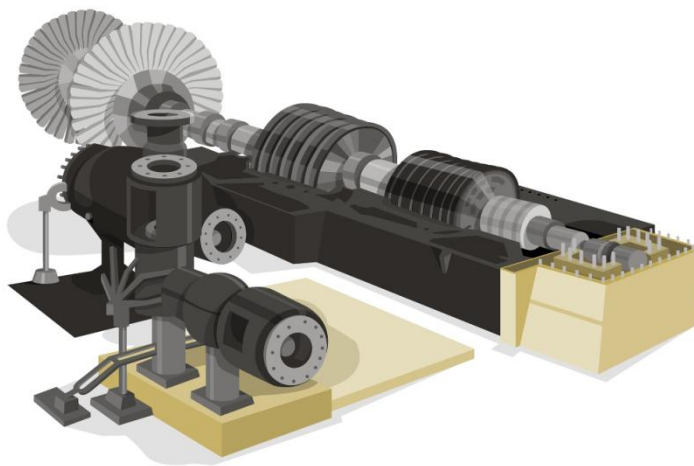


Figura 5.2: Central térmica de ciclo combinado [52].

Las principales características de las centrales térmicas de ciclo combinado son su flexibilidad, ya que la central puede operar a plena carga o cargas parciales hasta un mínimo de aproximadamente el 45% de la potencia máxima; la eficiencia elevada, debido a que el ciclo combinado proporciona mayor eficiencia por un margen más amplio de potencias; las emisiones son más bajas que en las centrales térmicas convencionales; un bajo coste de inversión por MW instalado; cortos periodos de construcción; ocupan una menor superficie por MW instalado al compararlas con las centrales termoeléctricas convencionales; tienen un bajo consumo de agua de refrigeración; y existe ahorro energético en forma de combustible.

Una central térmica de ciclo combinado (Figura 5.3) está formada por las siguientes partes:

1. Turbina de gas.
 - Compresor.
 - Cámara de combustión.
 - Turbina de gas.
2. Caldera de recuperación.
3. Turbina de vapor.

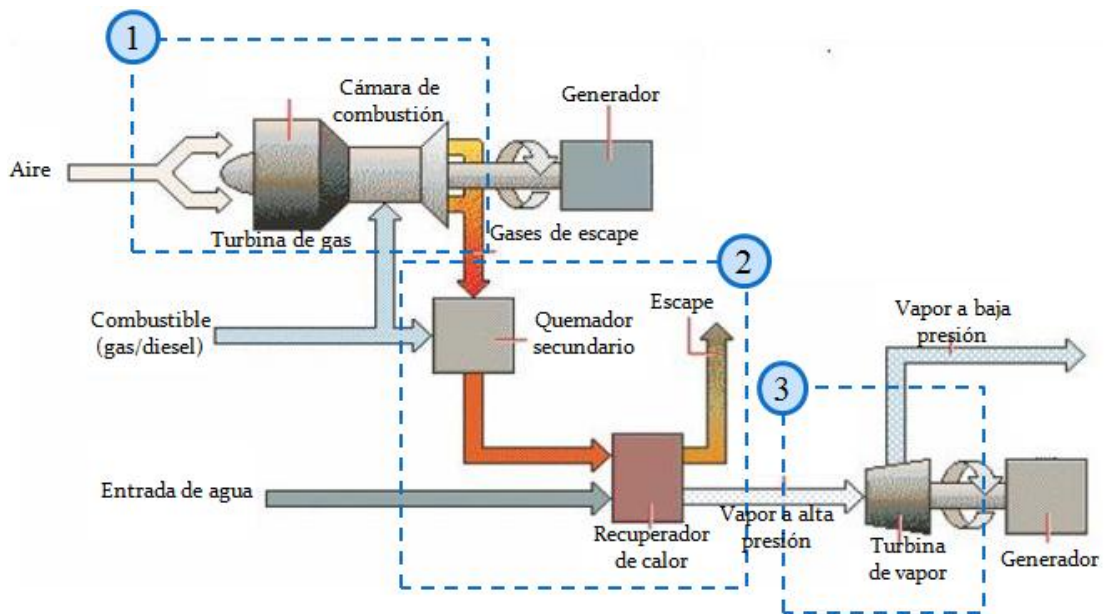


Figura 5.3: Esquema de una central térmica de ciclo combinado [54].

El funcionamiento de una central térmica de ciclo combinado se inicia cuando el aire es comprimido a alta presión en el compresor, pasando a la cámara de combustión donde se mezcla con el combustible. El compresor además cumple la función de inyectar aire para la refrigeración de las zonas calientes. Los gases de la combustión, cuya temperatura está alrededor de 1.400°C , pasan por la turbina de gas donde se expanden y su energía calorífica se transforma en energía mecánica, transmitiéndola al eje. Los gases que salen de la turbina de gas superan los 600°C .

En la caldera de recuperación de calor se produce vapor aprovechando los gases que salen de la turbina a gas. A la salida de la turbina el vapor se condensa (transformándose nuevamente en agua) y vuelve a la caldera para empezar un nuevo ciclo de producción de vapor. Actualmente la tendencia es acoplar la turbina de gas y la turbina de vapor a un mismo eje, de manera que accionan conjuntamente un mismo generador eléctrico.

5.1.2 BASE DE DATOS

Las bases de datos utilizadas para estudiar el proceso y luego aplicar en el sistema supervisor a través del análisis de dos anomalías, han sido facilitadas por ENDESA CHILE en el marco del proyecto de CONTAC INGENIEROS LTDA. en colaboración con la Universidad de Chile. El estudio se enfoca en aplicar el sistema supervisor de anomalías en la turbina a gas de la central térmica San Isidro, y no en sus otras etapas (caldera de recuperación y turbina a vapor). Las bases de datos están disponibles en dos archivos cuyas características se encuentran en la Tabla 5.1, y que recopilan observaciones de las variables de la turbina de gas de la central térmica de ciclo combinado.

Tabla 5.1: Bases de datos utilizadas para aplicar el sistema supervisor.

| | Base de datos A | Base de datos B |
|---|-------------------------------------|------------------------------------|
| N° Variables | 195 | 195 |
| N° Observaciones | 25923 | 19579 |
| Tiempo de muestreo | 1[min] | 10[min] |
| Fechas | 03-11-2008 / 06-11-2008 | 20-05-2011 / 03-10-2011 |
| | 21-11-2008 / 22-11-2008 | |
| | 02-12-2008 / 13-12-2008 | |
| Anomalías registradas | Temperatura excedida en álabe N°20. | Suciedad en el compresor. |
| Intervalos de observaciones en condición anómala | 14545:14770 | 1:3800 3830:5400 11500:13100 |

Los modelos generados en esta aplicación son estáticos. Así, a pesar que los tiempos de muestreo difieren en las dos bases de datos, este no será un parámetro relevante al momento de diseñar los modelos. Adicionalmente, en la Sección 5.4, estas bases de datos se acoplarán en una sola para generar los modelos del sistema supervisor.

En la Figura 5.4 y Figura 5.5 se tienen las mediciones en la base de datos A de la potencia activa y de los porcentajes de flujo de combustible, tanto de diesel como gas natural, respectivamente. En los instantes en que la potencia es nula la máquina no está operando. Además, se puede apreciar que opera con petróleo diesel en solo dos momentos (alrededor de las observaciones 5000 y 25000), y que en los otros instantes el combustible de operación es el gas natural.

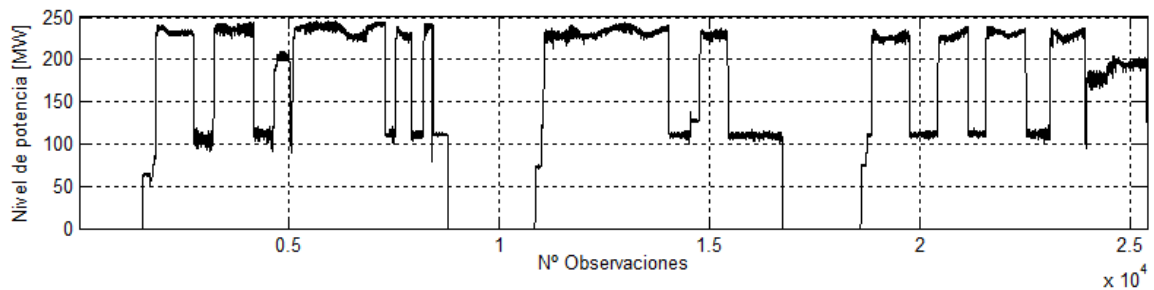


Figura 5.4: Potencia activa en la base de datos A.

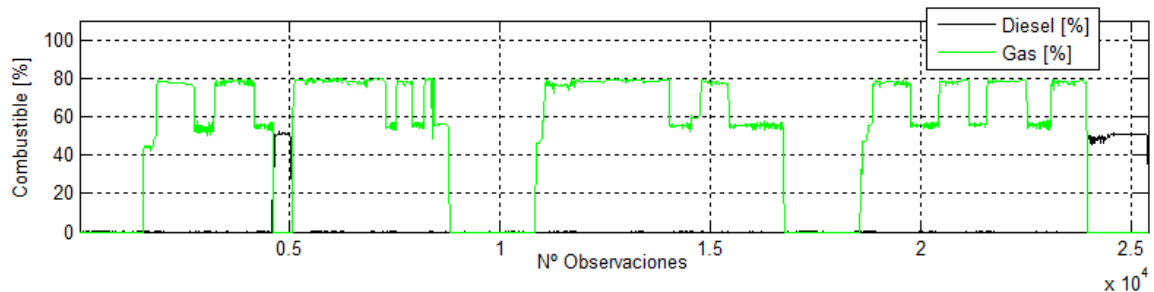


Figura 5.5: Porcentaje del flujo de combustible (diesel y gas) en la base de datos A.

En la Figura 5.6 y Figura 5.7 se encuentran las mediciones de la potencia activa y de los porcentajes de flujo de combustible en la base de datos B, respectivamente. Se puede apreciar que en este caso la turbina solo se encuentra en modo de no-operación y en operación con gas natural.

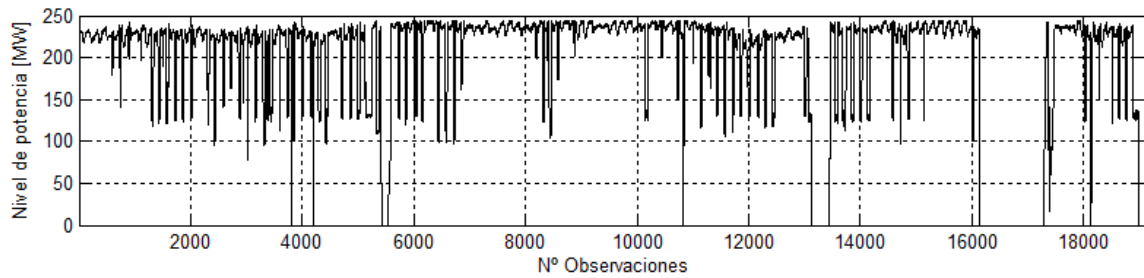


Figura 5.6: Potencia activa en la base de datos B.

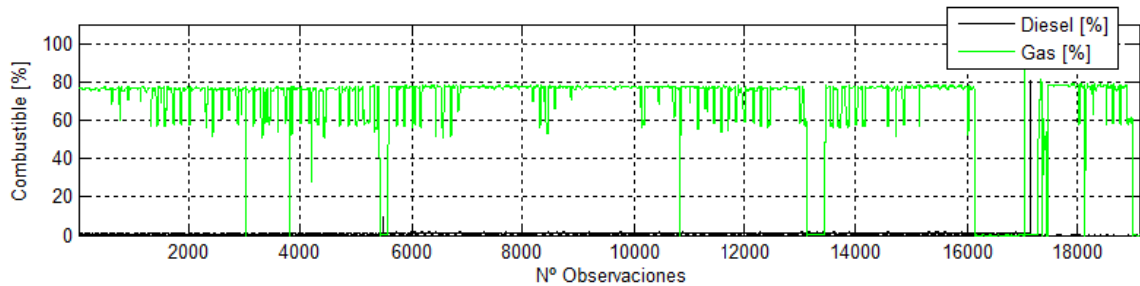


Figura 5.7: Porcentaje del flujo de combustible (diesel y gas) en la base de datos B.

5.2 GENERACIÓN DE MODELOS DE ANOMALÍAS REGISTRADAS

El sistema supervisor ha sido diseñado en base al estudio de dos tipos de anomalías que se encuentran en las bases de datos A y B respectivamente

- Anomalía #1: temperatura excedida en el álabe #20.
- Anomalía #2: suciedad en el compresor.

Estas dos anomalías son presentadas y estudiadas cada una por separado. Luego, en la Sección 5.4, el sistema supervisor realiza el estudio para conseguir que el análisis de residuos efectivamente detecte cada anomalía cuando corresponda.

La Anomalía #1 es analizada en la Sección 5.2.1 aplicando la metodología del estudio de anomalías registradas. Se especifican las variables utilizadas para modelar junto con los resultados de la detección. Por otro lado, la Anomalía #2 fue estudiada anteriormente en [1]. En dicho trabajo se evalúa con una base de datos de otra turbina de gas la herramienta de generación de modelos, y con la base de datos B se aplica la metodología del estudio de eventos para buscar aquellas variables que son más determinantes para modelar y detectar esta anomalía. En la Sección 5.2.2 se explicará esto de forma resumida mostrando los principales resultados obtenidos.

5.2.1 ANOMALÍA #1: TEMPERATURA EXCEDIDA EN ÁLABE N°20

La cámara de combustión se ubica entre el compresor y la turbina. Consiste en un recipiente donde ingresa el aire comprimido, al que se le añade el combustible, en este caso diesel o gas natural, que se quemará en forma ininterrumpida. Los gases producto de la combustión pasan a la siguiente etapa donde se expanden para hacer girar la turbina. La temperatura de salida de los gases de la cámara de combustión viene limitada por la resistencia mecánica de los álabes de la turbina. En la práctica, se utilizan los diferenciales de temperatura para verificar que no haya problemas con los álabes. Para obtener el diferencial de temperatura en los álabes, se calcula la variable *temperatura promedio de los gases* \bar{T} . Esta se obtiene eliminando de entre las 20 temperaturas de los álabes, las 3 más bajas y las 3 más altas. Con las 14 temperaturas restantes se calcula el promedio simple obteniéndose la variable \bar{T} . Luego, el diferencial de temperatura del álabe i-ésimo, considerando que tiene una temperatura T_i , se obtiene como

$$\Delta T_i = T_i - \bar{T}, \quad i = 1 \dots 20 \quad (5.1)$$

En base a lo mencionado por los operadores y por los informes de anomalías, este evento consiste en un aumento por sobre el límite permitido en el diferencial de temperatura de los gases del álabe #20 de la turbina. Para solucionar esta problemática, se debe detener la operación de la turbina y luego esperar a que los álabes de la cámara de combustión se enfríen. En la práctica esto no ocurre debido a los altos costos que implica el realizar una detención total en la turbina. Lo que efectivamente realizan los operadores es restringir la potencia, disminuyéndola en este caso a 126[MW], hasta que la temperatura se estabilice dentro de los límites permitidos.

Para estudiar este evento, se utiliza la base de datos A descrita en la Tabla 5.1. Tal como se observa en color rojo en la Figura 5.8a y Figura 5.8b, la anomalía ocurre alrededor de la observación 14500. Por otro lado, en color negro se observan los instantes en que la turbina opera de manera normal.

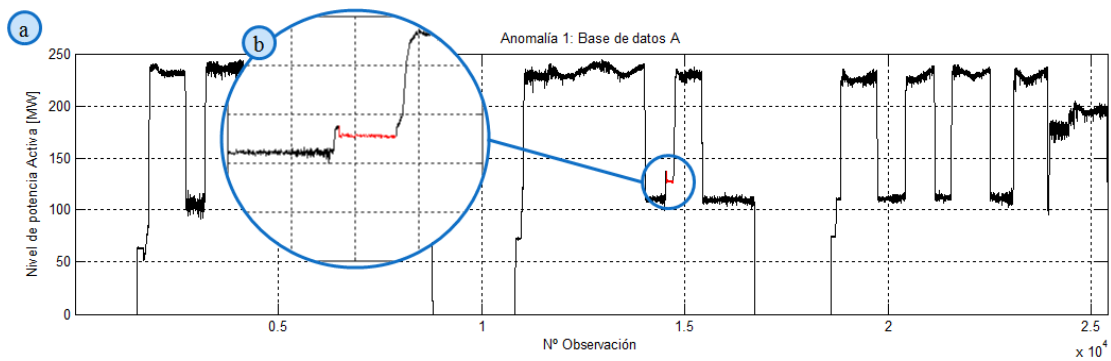


Figura 5.8: Potencia activa de la base de datos A con los instantes en que ocurre la Anomalía #1.

Debido a que la anomalía se prolonga por solo un par de horas, se determinó que un tiempo de muestreo adecuado sería de 1[*min*], disponiendo de este modo alrededor de 200 observaciones para estudiarla.

Para generar un modelo SBM de esta anomalía, la base de datos A se divide en tres subconjuntos, tal como se observa en la Figura 5.9: entrenamiento (2000-8600), validación (19400-24000) y prueba en la anomalía (11200-16500). Los instantes en que la turbina se encuentra apagada no son utilizados para generar ni para validar el modelo. Se determina que las observaciones que se ubican entre las primeras dos detenciones, y que además se encuentran previos a la anomalía, sean aquellos con los cuales se genera el modelo SBM. En los instantes entre la segunda y tercera detenciones se encuentran las observaciones en que ocurre la anomalía, y por lo tanto, estos datos se utilizan para validar y para la etapa de prueba en la anomalía. Finalmente, las observaciones ubicadas luego de la tercera detención se utilizan para validar el modelo y de ser necesario, incluirlas en el conjunto de entrenamiento. Este análisis corresponde a la Etapa 1 de la metodología de estudio de anomalías registradas.

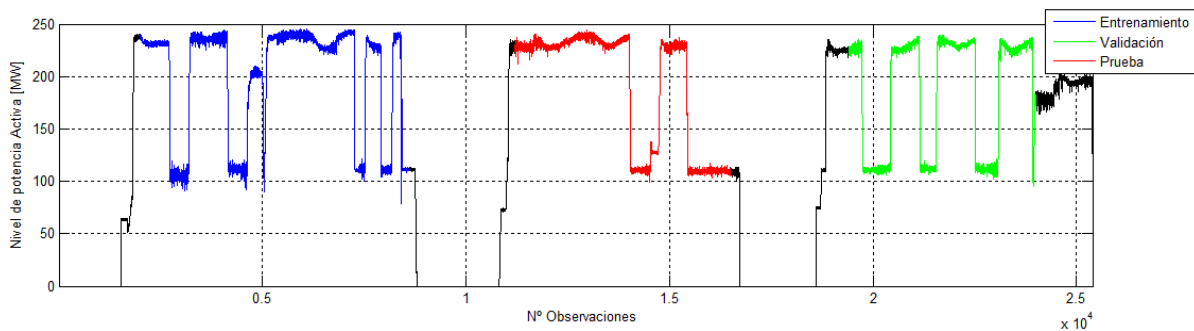


Figura 5.9: Potencia activa de la base de datos A con los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo).

La Etapa 2 corresponde a la selección preliminar de las variables para generar el modelo. Estas variables de entrada y de salida se encuentran en la Tabla 5.2. Tal como se ha mencionado en la Sección 4.2.2, esta selección se realiza en base a la información entregada por los operadores y los informes, y por el análisis de regresión PLS.

Tabla 5.2: Variables de entrada y salida del modelo preliminar.

| | VARIABLES DE ENTRADA | VARIABLES DE SALIDA |
|----------|---------------------------------|--------------------------------------|
| 1 | <i>Temperatura 1</i> | <i>Max. delta temp. gases álabes</i> |
| 2 | <i>Temperatura 2</i> | <i>Vibración X1</i> |
| 3 | <i>Señal de control</i> | <i>Vibración X2</i> |
| 4 | <i>Potencia activa</i> | <i>Vibración Y1</i> |
| 5 | <i>Temp. prom. gases álabes</i> | <i>Vibración Y2</i> |
| 6 | <i>Señal de referencia 1</i> | <i>Señal de control 1</i> |
| 7 | <i>Señal de referencia 2</i> | <i>Señal de control 2</i> |

La variable de salida “*Max. delta temp. gases álabes*” T_{max} no proviene directamente de la base de datos, sino que se calcula en base a un conjunto de variables de temperatura. Con (5.1) se obtiene el diferencial de temperatura para el álabe i -ésimo en el instante n -ésimo $\Delta T_i(n)$ y así T_{max} se calcula tal como se indica en (5.2).

$$T_{max}(n) = \max_{i=1...20} |\Delta T_i(n)| \quad (5.2)$$

Con estas variables se genera el primer modelo SBM del proceso, que corresponde a la Etapa 3 de la metodología. Desde la Figura 5.10 hasta la Figura 5.16 se encuentran los resultados obtenidos a través de este modelo, observándose las salidas estimadas en comparación a las salidas reales en los conjuntos de entrenamiento, validación y prueba en la anomalía. En Figura 5.18 y Tabla 5.3 se encuentran respectivamente, el indicador de Hotelling y la tabla de contingencia del modelo en los conjuntos de entrenamiento, validación y prueba. La tabla de contingencia será considerada un indicador del desempeño de los modelos, debido a que entrega de forma clara información del porcentaje de veces en que un modelo detecta correctamente, y el porcentaje de veces en que no lo logra.

Es importante destacar que la tabla de contingencia que estos porcentajes sólo se estiman en función del número de muestras en que el evento es o no detectado y del universo de datos disponibles para analizar. Se requiere de una cantidad considerablemente mayor de datos para caracterizar estos indicadores como *probabilidad de detección* o *probabilidad de falsa alarma*.

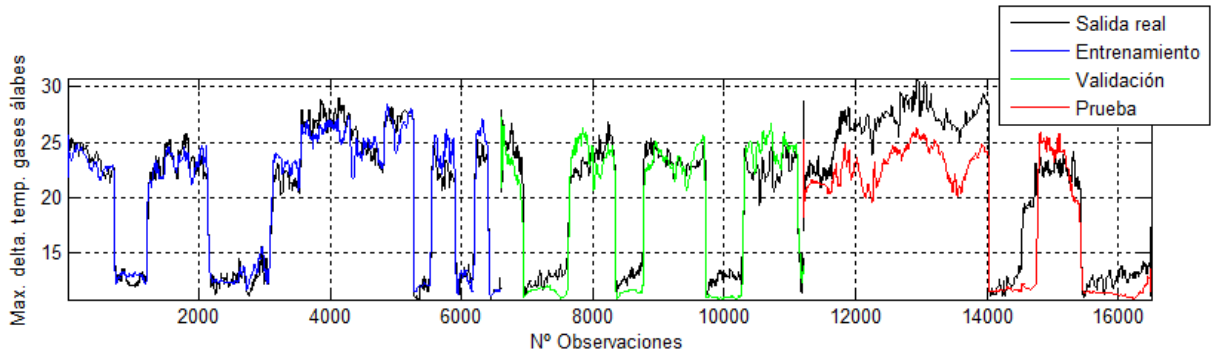


Figura 5.10: Variable de salida “*Max. delta temp. gases álabes*” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar.

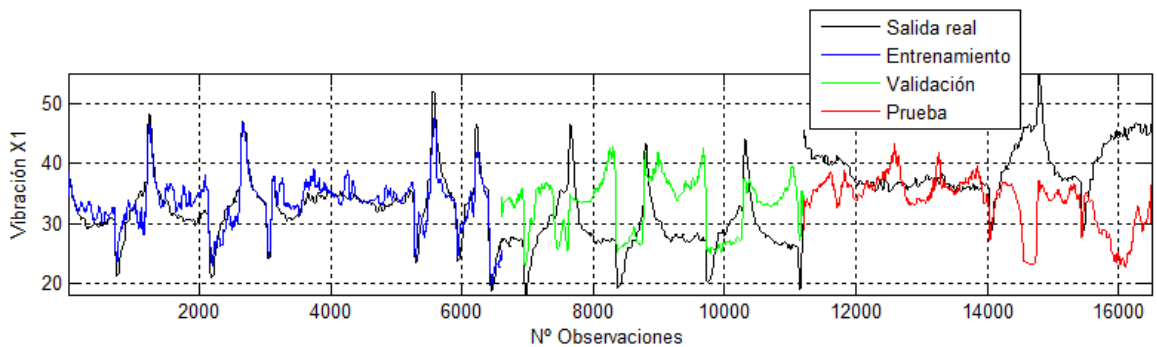


Figura 5.11: Variable de salida “*Vibración X1*” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar.

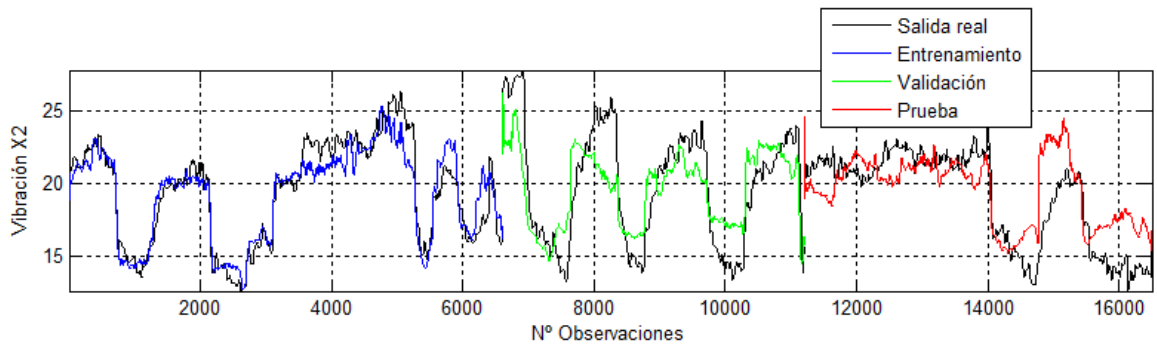


Figura 5.12: Variable de salida “Vibración X2” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar.

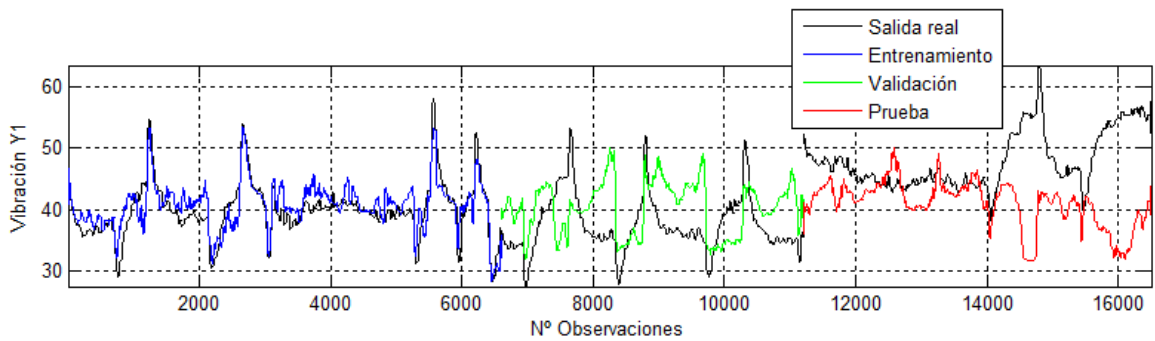


Figura 5.13: Variable de salida “Vibración Y1” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar.

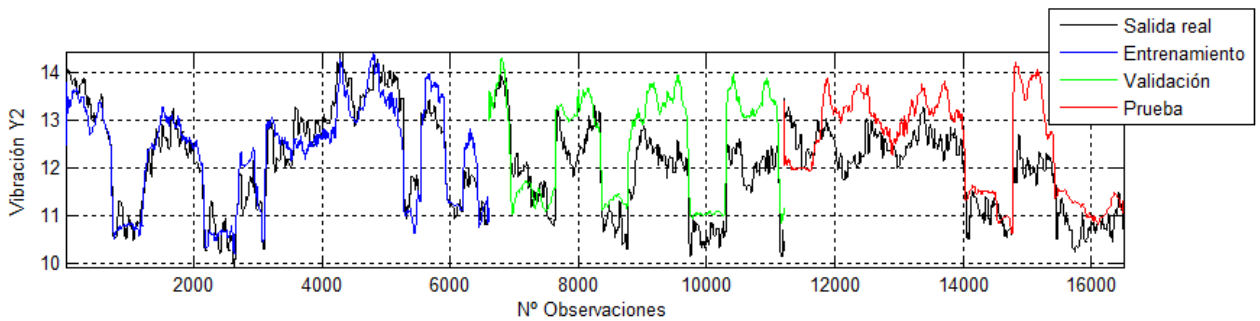


Figura 5.14: Variable de salida “Vibración Y2” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar.

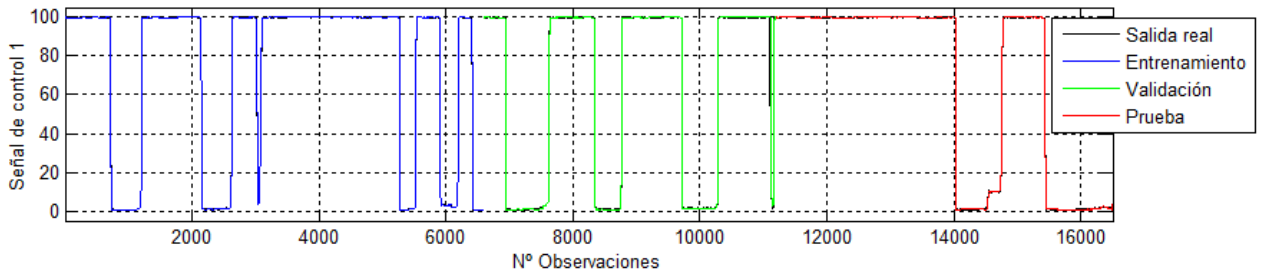


Figura 5.15: Variable de salida “Señal de control 1” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar.

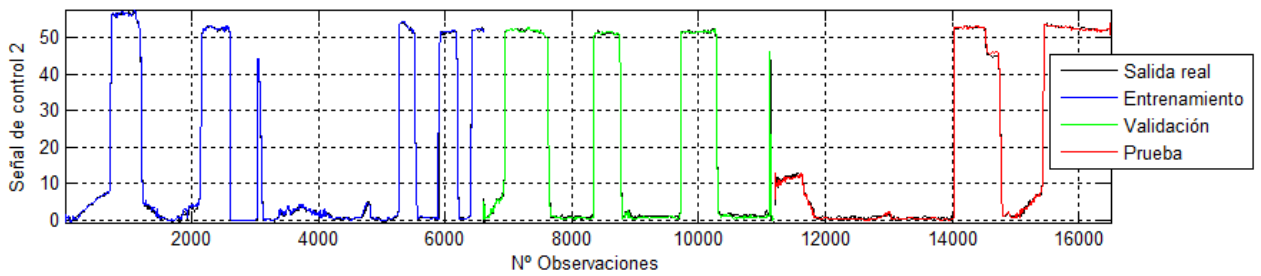


Figura 5.16: Variable de salida “Señal de control 2” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo preliminar.

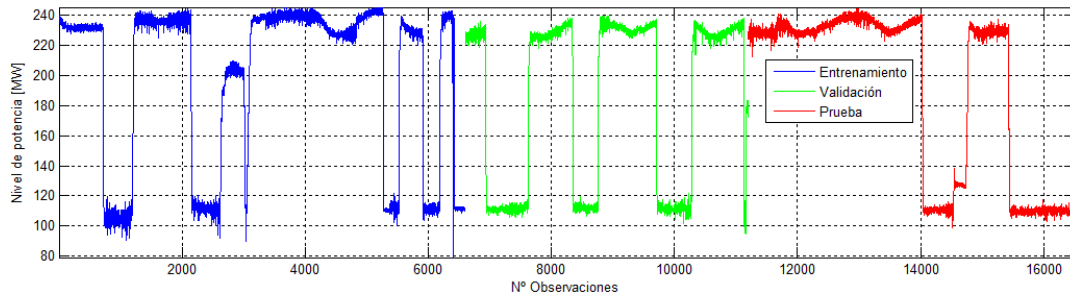


Figura 5.17: Potencia activa en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo).

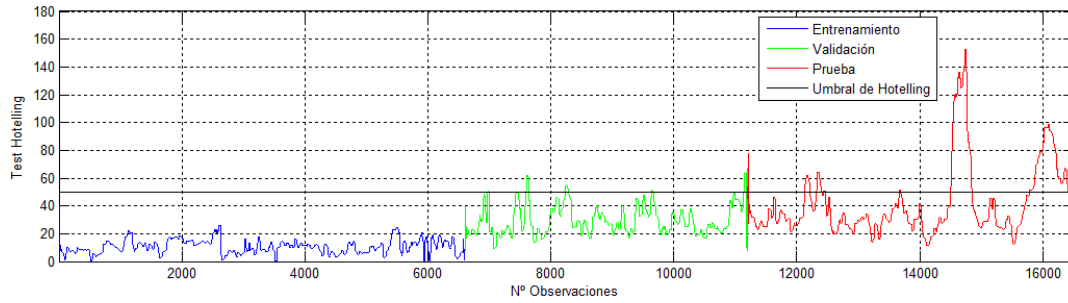


Figura 5.18: Indicador de Hotelling de los residuos con el modelo preliminar en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo).

Tabla 5.3: Tabla de contingencia del modelo preliminar en los conjuntos de entrenamiento, validación y prueba en la anomalía.

| | | Condición real del proceso | | |
|--------------------------------|--------------|----------------------------|-----------------------------------|---------------------------------|
| | | Anomalía #1 | Normal | |
| Condición estimada del proceso | Anomalía #1 | Verdaderos positivos | Falsos positivos (o falsa alarma) | Total condición estimada |
| | 225 (99.55%) | 5873 (36.07%) | 6098 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 1 (0.44%) | 10404 (63.92%) | 10405 | |
| Total condición real | | 226 | 16277 | 16503 |

A pesar que el indicador de Hotelling se encuentra sobre el umbral T_h en los instantes en que ocurre la anomalía, se observa en los residuos que los resultados no son del todo satisfactorios en la modelación y en la detección por lo que se procede a eliminar variables tal como se indica en la Etapa 5 de la metodología de estudio de eventos. En los instantes de validación, las variables “Vibración XI” y “Vibración YI” son eliminadas debido a que no aportan en la modelación. Por otro lado, las variables “Señal de control 1” y “Señal de control 2” también son eliminadas porque no aportan en la detección ya que modelan la anomalía adecuadamente y por tanto, no se produce un residuo en esos instantes. Este procedimiento se utiliza iterativamente para descartar variables hasta obtener finalmente un modelo con las variables de la Tabla 5.4.

Tabla 5.4: Variables de entrada y salida del modelo M_1 .

| | Variables de entrada | Variables de salida |
|----------|---------------------------------|--------------------------------------|
| 1 | <i>Señal de control</i> | <i>Max. delta temp. gases álabes</i> |
| 2 | <i>Potencia activa</i> | <i>Vibración X2</i> |
| 3 | <i>Temp. prom. gases álabes</i> | <i>Vibración Y2</i> |

Desde la Figura 5.19 hasta la Figura 5.21 se observan las variables de entrada del modelo M_1 , separados por una línea punteada los conjuntos de entrenamiento, validación y prueba en la anomalía. Los resultados obtenidos con el modelo generado se observan desde la Figura 5.22 hasta la Figura 5.24 con las variables reales en color negro y las variables estimadas en color azul, verde y rojo para indicar los conjuntos de entrenamiento, validación y prueba en la anomalía respectivamente. Finalmente, en la Figura 5.25 y la Tabla 5.5 se encuentran el indicador de Hotelling de los residuos y la tabla de contingencia del modelo M_1 .

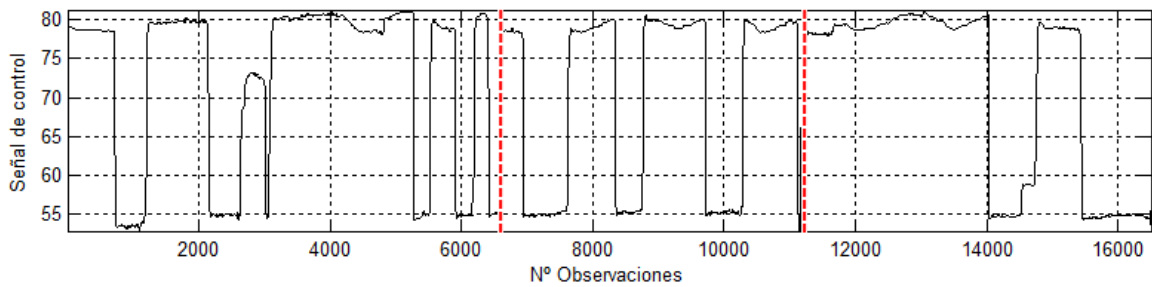


Figura 5.19: Variable de entrada “*Señal de control*” del modelo M_1 .

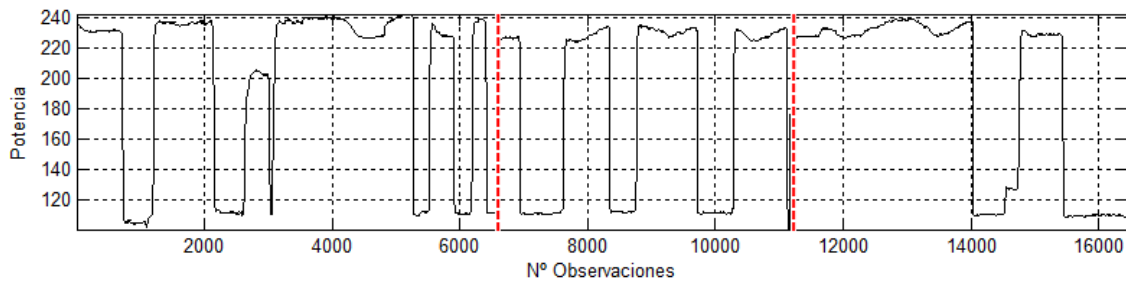


Figura 5.20: Variable de entrada “*Potencia activa*” del modelo M_1 .

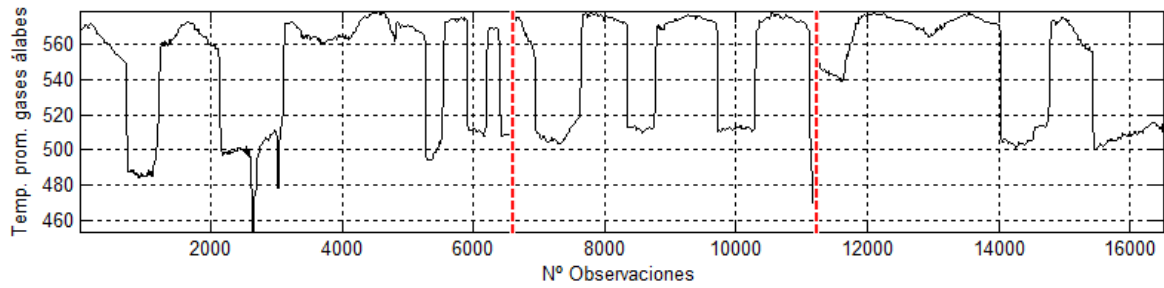


Figura 5.21: Variable de entrada “Temp. prom. gases álabes” del modelo M_1 .

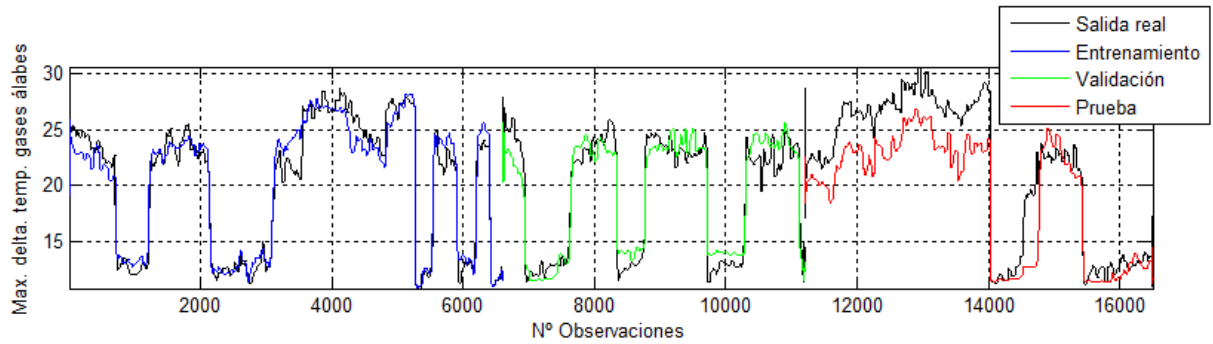


Figura 5.22: Variable de salida “Max. delta temp. gases álabes” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_1 .

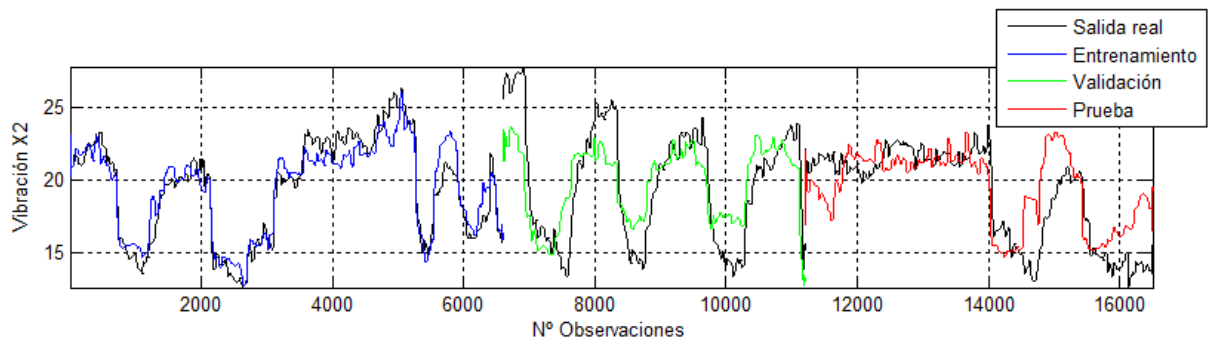


Figura 5.23: Variable de salida “Vibración X2” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_1 .

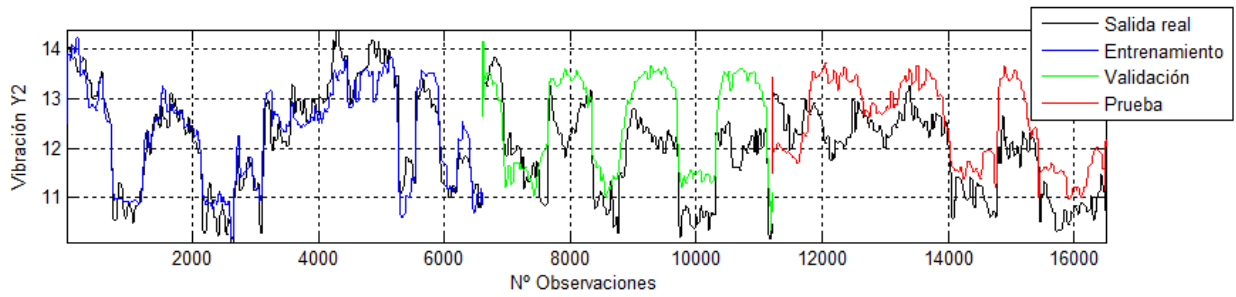


Figura 5.24: Variable de salida “Vibración Y2” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_1 .

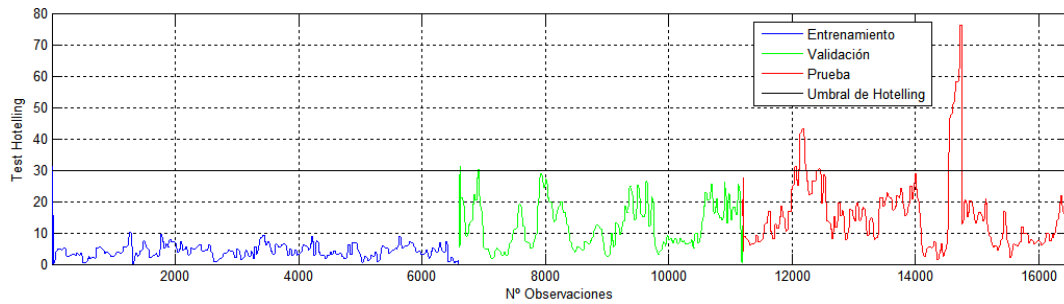


Figura 5.25: Indicador de Hotelling de los residuos con el modelo M_1 en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo).

Tabla 5.5: Tabla de contingencia del modelo M_1 en los conjuntos de entrenamiento, validación y prueba en la anomalía.

| | | Condición real del proceso | | |
|--------------------------------|--------------|----------------------------|-----------------------------------|---------------------------------|
| | | Anomalía #1 | Normal | |
| Condición estimada del proceso | Anomalía #1 | Verdaderos positivos | Falsos positivos (o falsa alarma) | Total condición estimada |
| | 208 (91.90%) | 213 (1.30%) | 421 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 18 (8.00%) | 16064 (98.69%) | 16082 | |
| Total condición real | | 226 | 16277 | 16503 |

Tal como ocurre con el modelo preliminar, la Anomalía #1 también es detectada con el modelo M_1 al utilizar como indicador el índice de Hotelling. El umbral T_h se ha reajustado en base a los datos de validación y de prueba en $T_h = 30$. Adicionalmente se observa con la Tabla 5.3 y la Tabla 5.5 que el número de falsas alarmas disminuye considerablemente de un 36.07% a un 1.30%. Por otro lado, esto ocasiona una disminución en 8% en la capacidad de detección. El tiempo de cómputo también es menor debido a que la cantidad de variables de entrada y salida utilizadas para modelar ha disminuido a 6.

Finalmente, en la Tabla 5.6 se encuentra otro parámetro de evaluación: el error cuadrático medio de cada variable de salida calculado como en la Sección 2.5.1. Este indicador muestra que la “Vibración X2” mejora considerablemente en el modelo final. Con respecto a las variables “Max. delta temp. gases álabes” y “Vibración Y2” no es posible obtener conclusiones sobre si existe una mejora sustancial o no.

Tabla 5.6: Error cuadrático medio en los modelos preliminar y M_1 en los conjuntos de entrenamiento y validación

| | Modelo | Max. delta. T° gases | Vibración X2 | Vibración Y2 |
|----------------------|---------------|---------------------------------|---------------------|---------------------|
| Entrenamiento | Preliminar | 1.4262 | 1.4393 | 0.1660 |
| | M_1 | 1.5627 | 1.2902 | 0.1342 |
| Validación | Preliminar | 2.9298 | 6.3290 | 0.7906 |
| | M_1 | 2.3220 | 6.2095 | 0.8819 |

5.2.2 ANOMALÍA #2: SUCIEDAD DEL COMPRESOR

La Anomalía #2 corresponde a una acumulación de suciedad en zonas internas del compresor de la turbina a gas de la central termoeléctrica de ciclo combinado San Isidro. Como consecuencia, se produce una disminución en la eficiencia de operación, es decir, para una misma cantidad de combustible la potencia es cada vez menor. Esta anomalía se soluciona manteniendo apagada la turbina para conseguir a limpiar la zona. Según la experiencia de los operadores, para el mismo nivel de combustible esta limpieza provoca un aumento en la potencia generada y en la presión del compresor, y una disminución en la temperatura de descarga del compresor.

Para el estudio de este evento se utiliza la base de datos B. Esta anomalía se puede prolongar por meses, ya que su presencia no impide el funcionamiento de la turbina, solo que opere a una menor eficiencia. De este modo, se determinó un tiempo de muestreo adecuado de 10[*min*]. Tal como se observa en la Figura 5.26, la anomalía se registra en dos intervalos de observaciones: desde el comienzo de la base de datos hasta 5400, y entre las observaciones 11500 y 13100 aproximadamente.

A continuación se muestran los resultados finales de la metodología para el estudio de eventos en la anomalía #2, ya que el procedimiento detallado se encuentra en [1]. El modelo M_2 que aquí se genera, es posteriormente utilizado en la Sección 5.4 en el sistema supervisor.

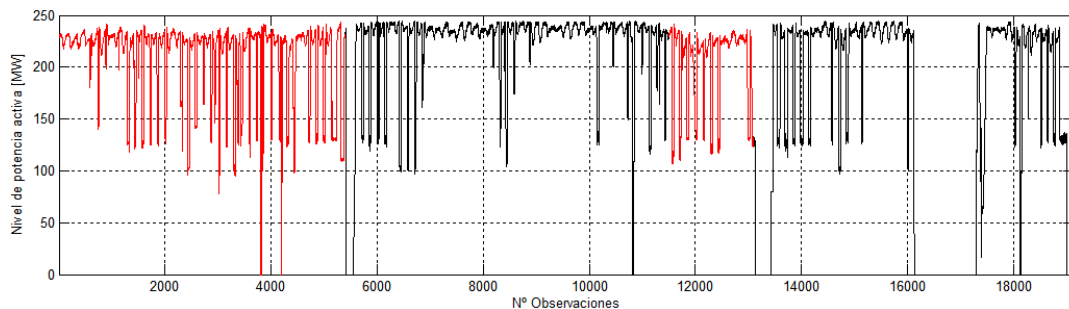


Figura 5.26: Potencia activa de la base de datos B con los instantes en que ocurre la Anomalía #2.

La base de datos B contiene a observaciones que abarcan desde 20-05-2011 hasta 03-10-2011. Se divide en tres subconjuntos, tal como se observa en la Figura 5.27: entrenamiento (5900-7489, 17620-18115, 18145-18200), validación (7490-10800, 13500-16000, 18250-18950) y prueba en la anomalía (1-3800, 3830-5400, 11500-13100). Los instantes en que la turbina se encuentra apagada no son utilizados para generar ni para validar el modelo. Debido a que se realizaron modificaciones en algunas componentes de la turbina, los datos de entrenamiento 5900-7489 fueron insuficientes para modelar las observaciones posteriores a 16000. Esta es la razón de que se incorporaran los intervalos de observaciones 17620-18115 y 18145-18200 al conjunto de entrenamiento. A diferencia de la base de datos A, en este caso la turbina solo opera con gas natural. En caso que la turbina comience a funcionar con petróleo diesel, se recomienda agregar observaciones al modelo para que esta condición de operación sea considerada.

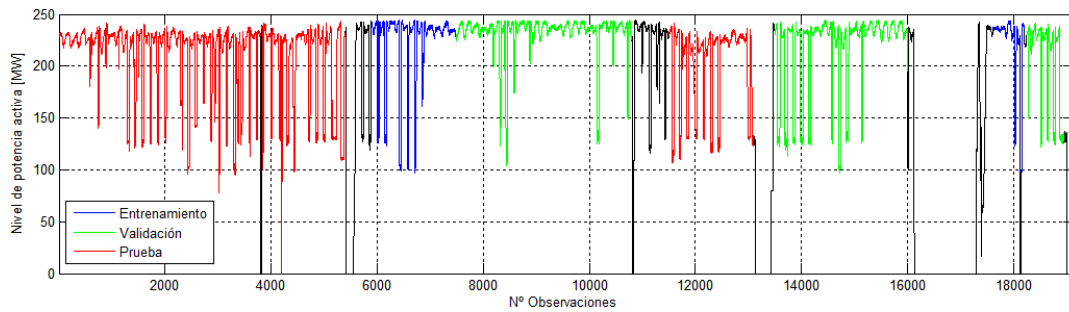


Figura 5.27: Potencia activa de la base de datos B con los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo).

Luego de estudiar las variables con los métodos de selección a través del conocimiento de operadores y con regresión PLS; y por medio de los métodos de eliminación de la Etapa 5 de la metodología de estudios, se obtienen las variables de entrada y de salida de la Tabla 5.7.

Tabla 5.7: Variables de entrada y salida del modelo M_2 .

| | Variables de entrada | Variables de salida |
|----------|-----------------------------|----------------------------|
| 1 | <i>Flujo de combustible</i> | <i>Presión</i> |
| 2 | <i>Señal de control</i> | <i>Potencia activa</i> |
| 3 | <i>Temperatura 1</i> | <i>Temperatura 2</i> |
| 4 | <i>Posición Válvula 1</i> | |
| 5 | <i>Posición Válvula 2</i> | |

Desde la Figura 5.28 hasta la Figura 5.32 se observan las variables de entrada del modelo, separados por una línea punteada los conjuntos de entrenamiento, validación y prueba en la anomalía. Los resultados obtenidos con el modelo generado se observan desde la Figura 5.33 hasta la Figura 5.35 a través de los residuos entre las variables reales y estimadas en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo). En la Figura 5.36 y la Figura 5.37 se encuentran la potencia activa y el indicador de Hotelling de los residuos respectivamente. Finalmente, en Tabla 5.8 se encuentra la tabla de contingencia del modelo M_2 en los conjuntos de entrenamiento, validación y prueba.

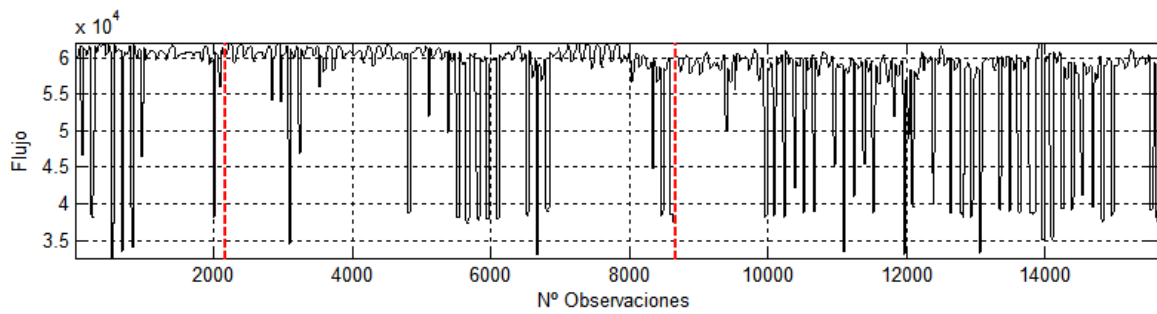


Figura 5.28: Variable de entrada “*Flujo de combustible*” del modelo M_2 .

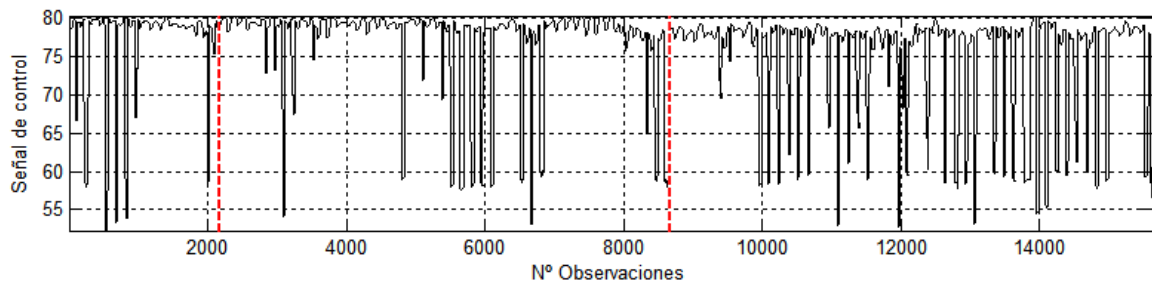


Figura 5.29: Variable de entrada “*Señal de control*” del modelo M_2 .

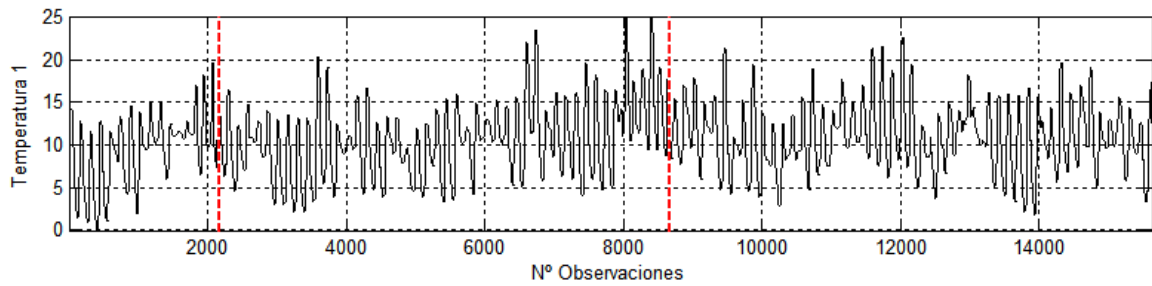


Figura 5.30: Variable de entrada “*Temperatura 1*” del modelo M_2 .

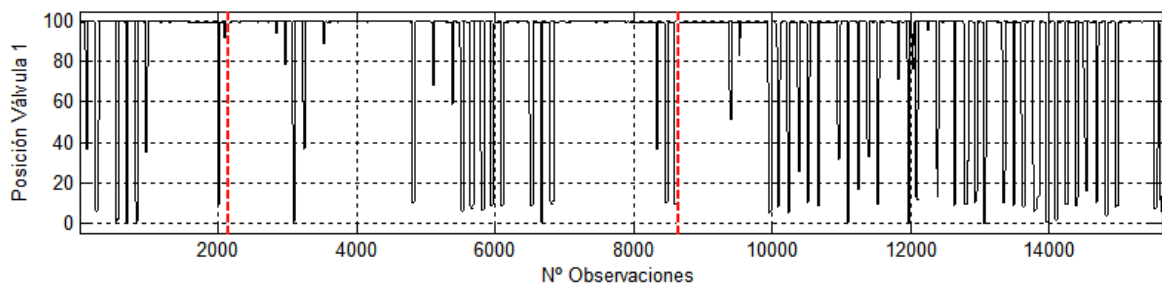


Figura 5.31: Variable de entrada “*Posición Válvula 1*” del modelo M_2 .

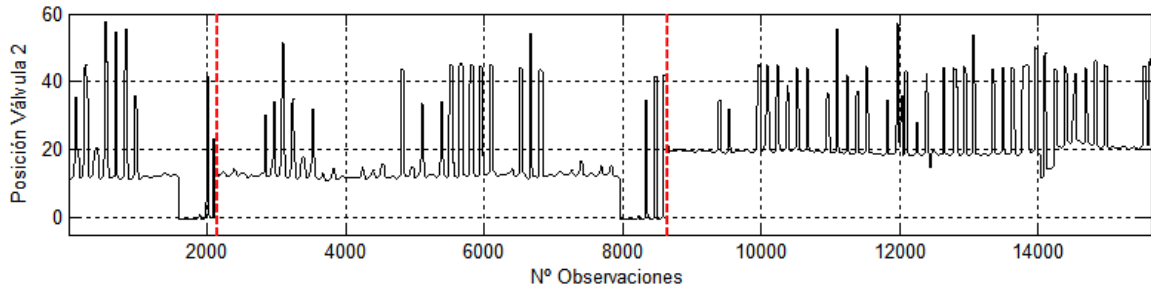


Figura 5.32: Variable de entrada “*Posición Válvula 2*” del modelo M_2 .

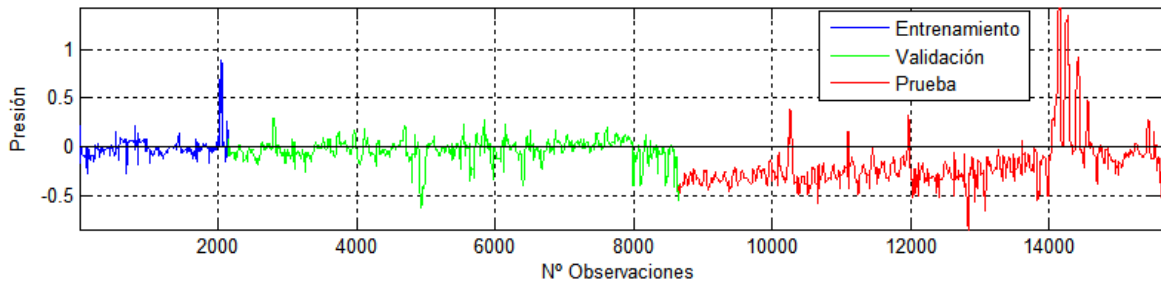


Figura 5.33: Variable de salida “*Presión*” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_2 .

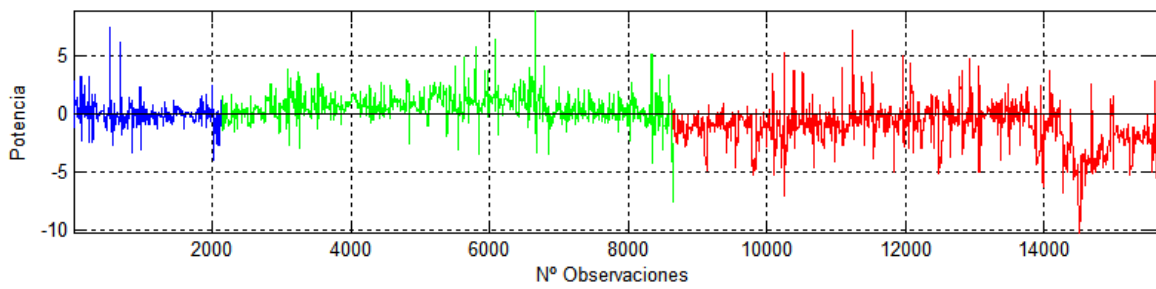


Figura 5.34: Variable de salida “*Potencia activa*” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_2 .

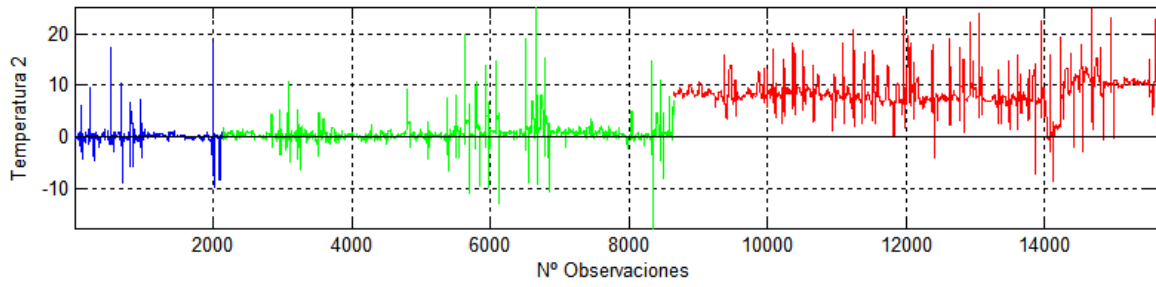


Figura 5.35: Variable de salida “*Temperatura 2*” real (negro) y estimada en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) del modelo M_2 .

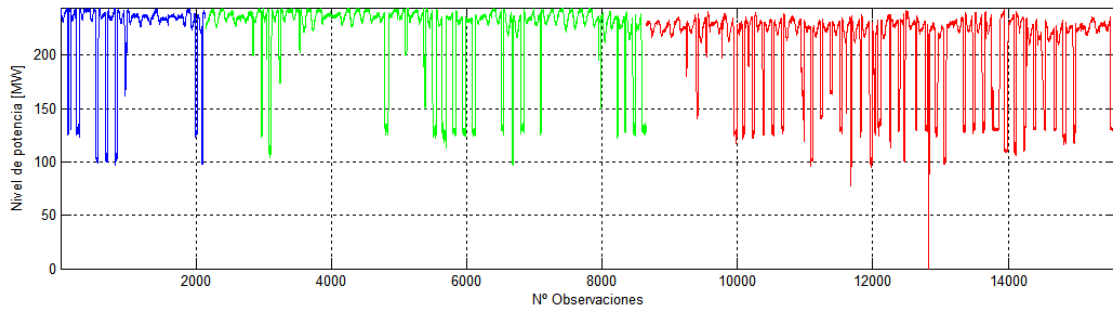


Figura 5.36: Potencia activa en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo).

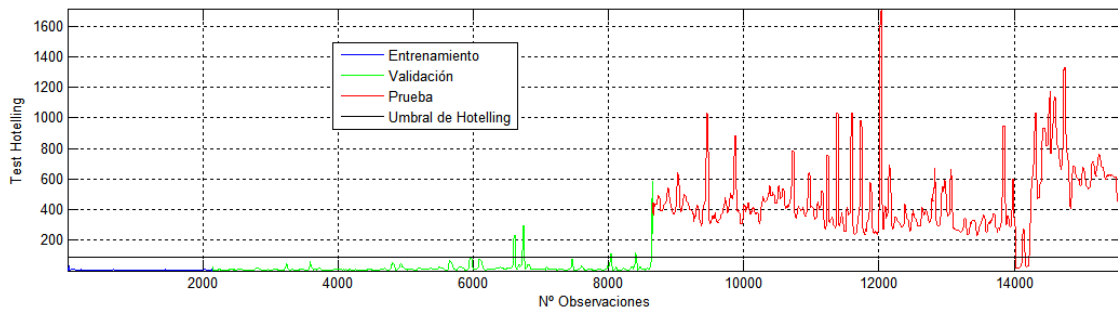


Figura 5.37: Indicador de Hotelling de los residuos con el modelo M_2 en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo).

Tabla 5.8: Tabla de contingencia del modelo M_2 en los conjuntos de entrenamiento, validación y prueba en la anomalía

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|-------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalía #2 | Normal | |
| Condición estimada del proceso | Anomalía #2 | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | | 6872 (98.56%) | 647 (7.47%) | 7519 |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | | 100 (1.43%) | 8008 (92.52%) | 8108 |
| Total condición real | | 6972 | 6972 | 8655 |

Las observaciones del conjunto de entrenamiento comprendidas entre 17620 y 18200 han sido agregadas al conjunto de entrenamiento debido a que los operadores del proceso advirtieron que luego de la última detención de la máquina, se hicieron cambios en algunos componentes de la turbina. Al no incluir estos instantes, el indicador de Hotelling aumentó considerablemente a pesar que el sistema había sido limpiado recientemente. Este tipo de consideraciones deben tomarse en cuenta a todo nivel en el sistema supervisor, es decir, con el modelo clasificador y con los modelos de anomalías registradas. Se observa que la capacidad de detección alcanza un 98.56% del total de datos en que el proceso se encuentra en presencia de la Anomalía #2. Además, solo hay un 7.47% de falsas alarmas.

5.3 MODELO CLASIFICADOR

Con las anomalías ya definidas y estudiadas, se procede a la generación del modelo clasificador que se utiliza en la Fase I del sistema supervisor y permite determinar si el proceso se encuentra en una condición de operación normal o no, y generar una alarma en caso que sea necesario. Hay que tener presente que los objetivos primordiales del modelo clasificador son: que sea capaz de detectar todas las anomalías que están siendo estudiadas y disminuir el número de falsos positivos para así evitar la Fase II y Fase III del sistema supervisor que utilizarían recursos computacionales innecesarios.

Tal como fue explicado en la metodología para el estudio de eventos, la primera etapa consiste en determinar el conjunto de entrenamiento para generar el modelo y el conjunto de prueba en la anomalía para reajustarlo. Ya que el proceso en estudio cuenta con dos bases de datos disímiles entre sí, tanto en el tiempo de muestreo, como en las componentes físicas de la turbina, se determina que los conjuntos contengan muestras de ambas. En la Figura 5.38 y Figura 5.39 se observa la potencia activa en los conjuntos de entrenamiento y prueba de las bases de datos A y B respectivamente.

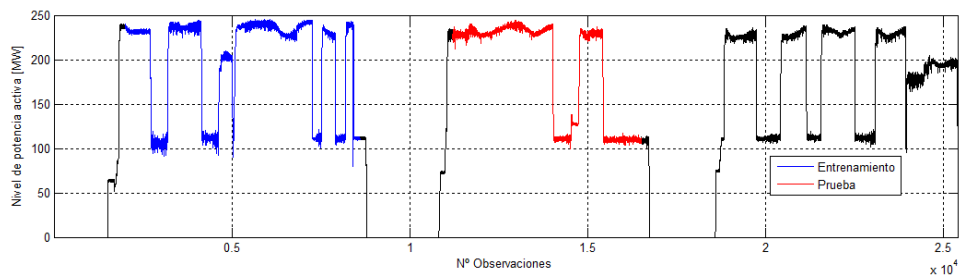


Figura 5.38: Potencia activa de la base de datos A con los conjuntos de entrenamiento (azul) y prueba en la anomalía (rojo)

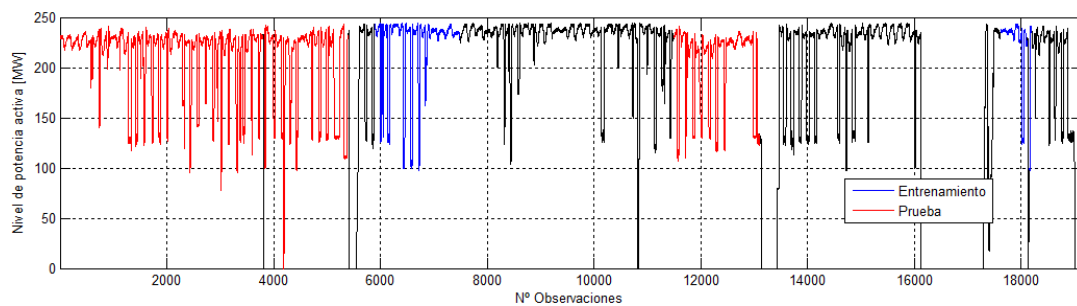


Figura 5.39: Potencia activa de la base de datos B con los conjuntos de entrenamiento (azul) y prueba en la anomalía (rojo).

La siguiente etapa consiste en determinar las variables de entrada del modelo clasificador. A este conjunto pertenecen todas las variables de entrada y de salida de las anomalías que están siendo estudiadas. Por lo tanto, en el proceso en estudio las variables de entrada del modelo son las siguientes:

Tabla 5.9: Variables de entrada del modelo clasificador.

| | Nombre | | Nombre |
|---|--------------------------------------|----|---------------------------------|
| 1 | <i>Max. delta temp. gases álabes</i> | 7 | <i>Temperatura 1</i> |
| 2 | <i>Presión</i> | 8 | <i>Temperatura 2</i> |
| 3 | <i>Flujo de combustible</i> | 9 | <i>Posición Válvula 1</i> |
| 4 | <i>Vibración X2</i> | 10 | <i>Posición Válvula 2</i> |
| 5 | <i>Vibración Y2</i> | 11 | <i>Potencia activa</i> |
| 6 | <i>Señal de control</i> | 12 | <i>Temp. prom. gases álabes</i> |

Con las variables de entrada ya definidas, se procede a la obtención de las variables de salida con la determinación de la agrupación a la que pertenece cada observación y a través del modelo SBM. En la Figura 5.40 se observan las muestras del conjunto de entrenamiento en las primeras dos componentes principales, los centros de cada agrupación y la agrupación a la que pertenece cada una de las observaciones. Esta última corresponde a las variables de salida con las que se genera el modelo SBM.

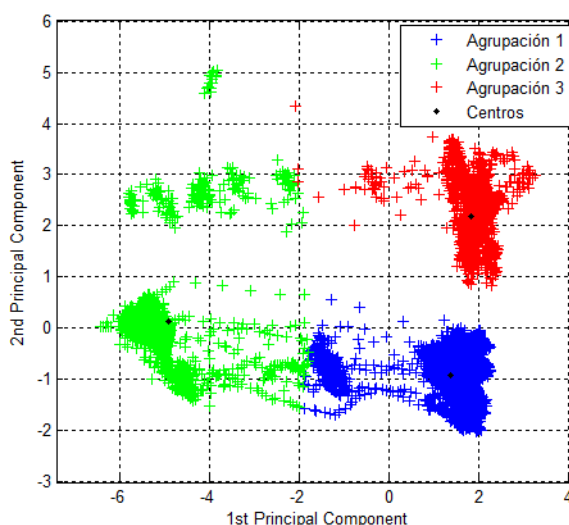


Figura 5.40: Las tres agrupaciones y centros (negro) en el conjunto de entrenamiento.

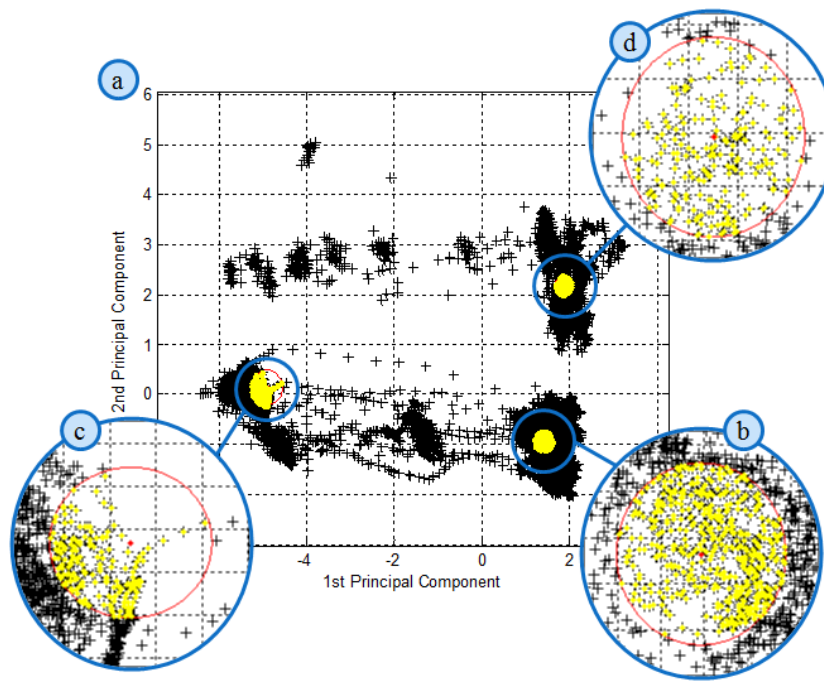


Figura 5.41: Observaciones seleccionadas en cada agrupación para obtener las matrices D_i y D_o .

Para obtener las matrices D_i y D_o se escoge un 10% de las observaciones más cercanas a cada centro tal como se explica en la Sección 3.2.1. En la Figura 5.41 se observan las tres agrupaciones encontradas con los centros de cada una (rojo) y las observaciones seleccionadas (amarillo). Con las matrices D_i y D_o características del modelo SBM, se obtienen las variables de salida estimadas. A continuación, se obtiene el umbral de decisión de cada agrupación que determina tal como se explica en la Sección 3.2.3 según (4.4) si el sistema supervisor genera o no una alarma.

En la Figura 5.42, Figura 5.43 y Figura 5.44 se observan las salidas obtenidas al clasificar cada observación del conjunto de entrenamiento en alguna de las tres agrupaciones encontradas de acuerdo a la cercanía que tiene a cada centro (negro), la salida estimada con el modelo basado en similitud (azul) y el umbral de decisión (verde).

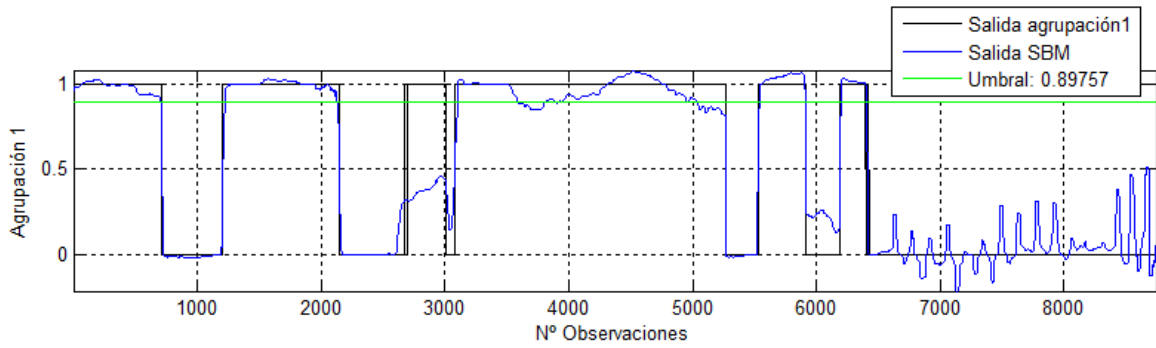


Figura 5.42: Variable de salida “Agrupación 1” en el conjunto de entrenamiento (negro), salida estimada con SBM (azul) y umbral de decisión (verde).

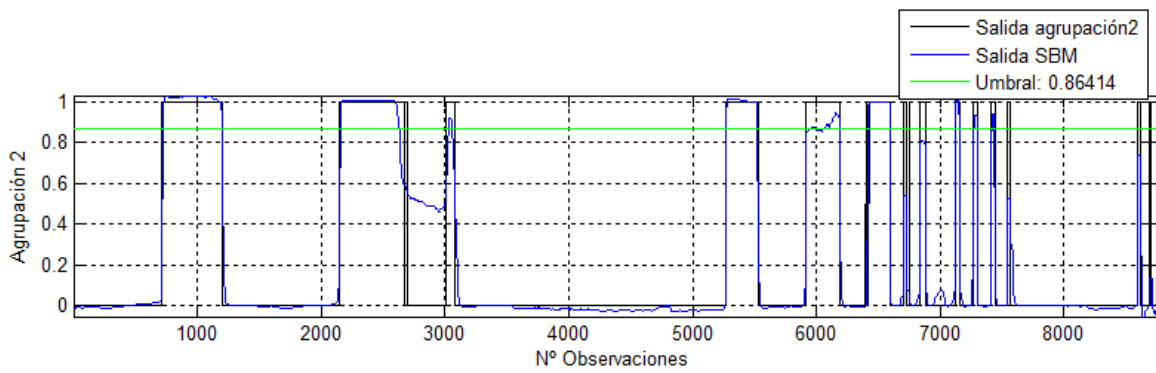


Figura 5.43: Variable de salida “Agrupación 2” en el conjunto de entrenamiento (negro), salida estimada con SBM (azul) y umbral de decisión (verde).

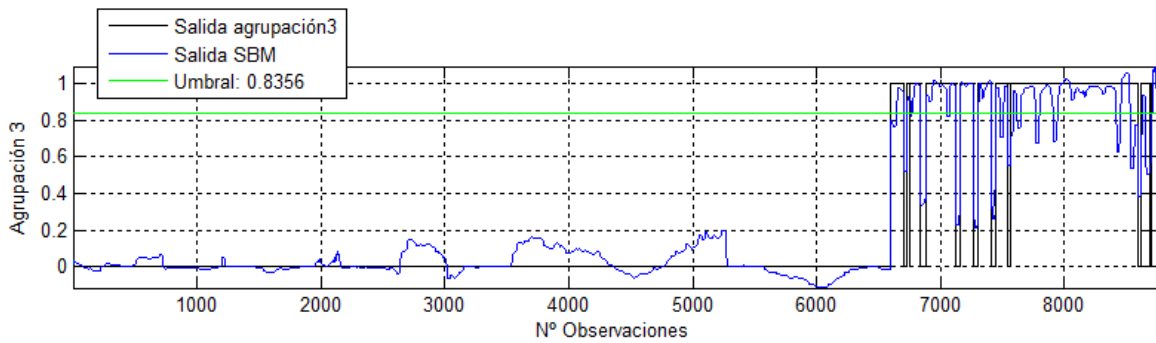


Figura 5.44: Variable de salida “Agrupación 3” en el conjunto de entrenamiento (negro), salida estimada con SBM (azul) y umbral de decisión (verde).

En la Figura 5.45 y Figura 5.46 se observa la potencia activa en el conjunto de entrenamiento y la variable binaria “alarma” que define si el sistema supervisor genera una señal de alerta o no. Se esperaría que en el conjunto de entrenamiento la señal de alarma fuese igual a 0 en todos los instantes debido a que estas observaciones corresponden a una condición normal del proceso. Sin embargo, esto no sucede en gran parte debido a que cierto porcentaje de estas alertas se produce en los cambios de operación. Estas falsas alarmas podrían evitarse en el futuro si se genera un detector de cambios de condición de operación.

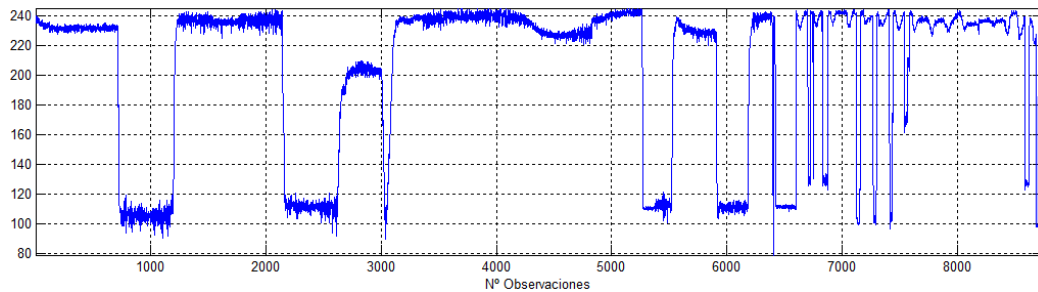


Figura 5.45: Potencia activa en el conjunto de entrenamiento.

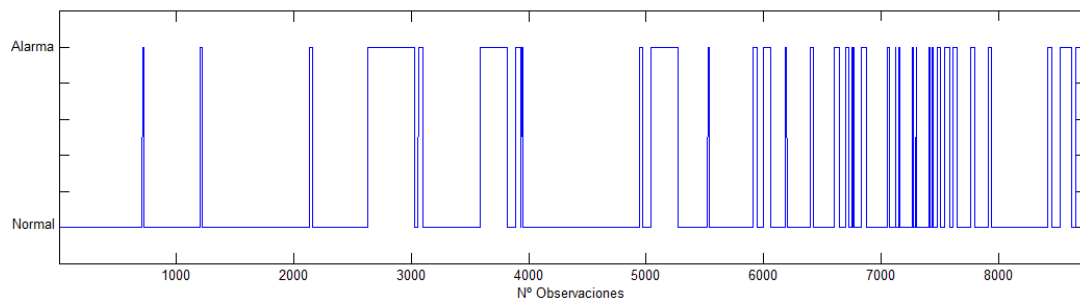


Figura 5.46: Señal de alarma en el conjunto de entrenamiento.

Con el modelo clasificador ya generado, se obtiene en el conjunto de prueba las salidas estimadas en base a los centros de las agrupaciones y a través del modelo SBM. En la Figura 5.47, Figura 5.48 y Figura 5.49 se observan las muestras de la base de datos A con la salida obtenida a través de las agrupaciones (negro), la salida estimada con el modelo SBM en el conjunto de entrenamiento (azul) y de prueba (rojo). Además, se observa el umbral en cada agrupación (verde).

Tal como se observa en la Figura 5.51, la Anomalía #1 es detectada con el modelo clasificador, ubicándose alrededor de la observación 10000. Se ha aplicado una mediana móvil con largo $k = 30$ a la señal de alarma para filtrar los cambios de operación que están siendo considerados como falsas alarmas. Se ha escogido ese largo debido a que gran parte de los cambios de operación en esta base de datos se producen con 30 observaciones o menos.

Otro hecho importante para mencionar es que el modelo clasificador detecta como anomalía las muestras alrededor de 3000 que corresponden a la turbina utilizando combustible diesel. La explicación es que esta operación no fue incluida al determinar las agrupaciones de la base de datos, debido a la poca cantidad de observaciones de la operación con diesel.

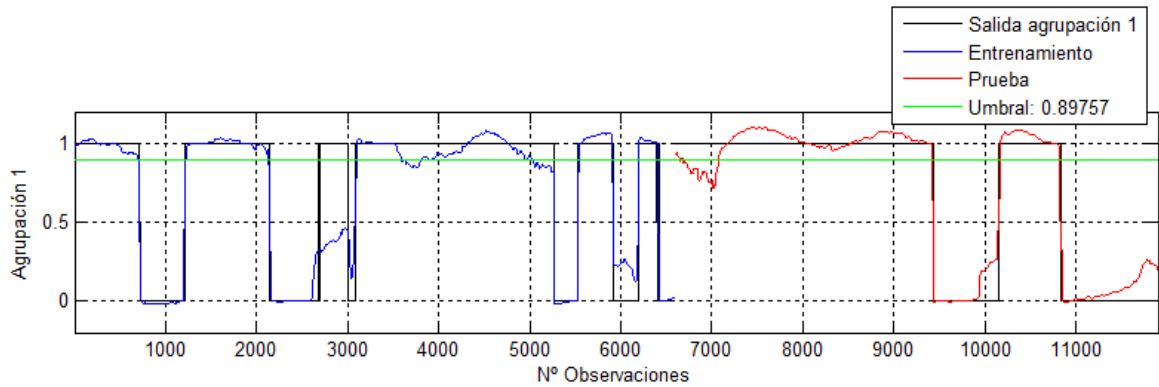


Figura 5.47: Variable de salida “Agrupación 1” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde).

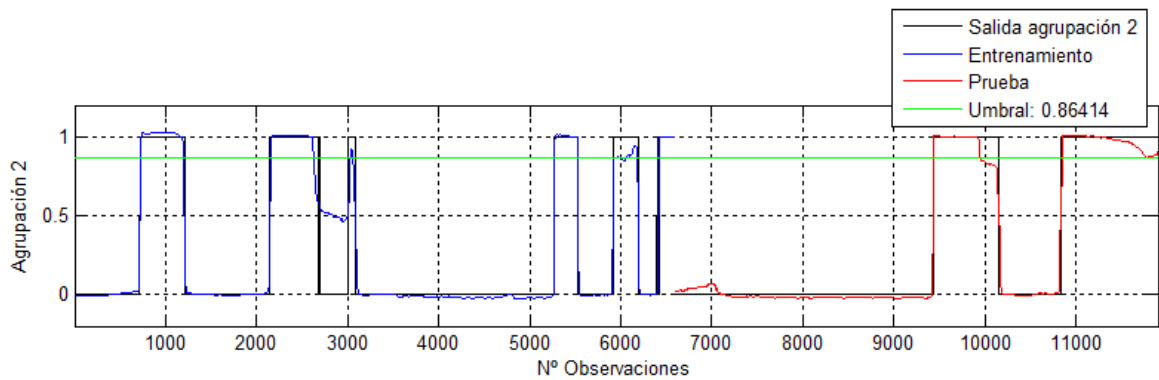


Figura 5.48: Variable de salida “Agrupación 2” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde).

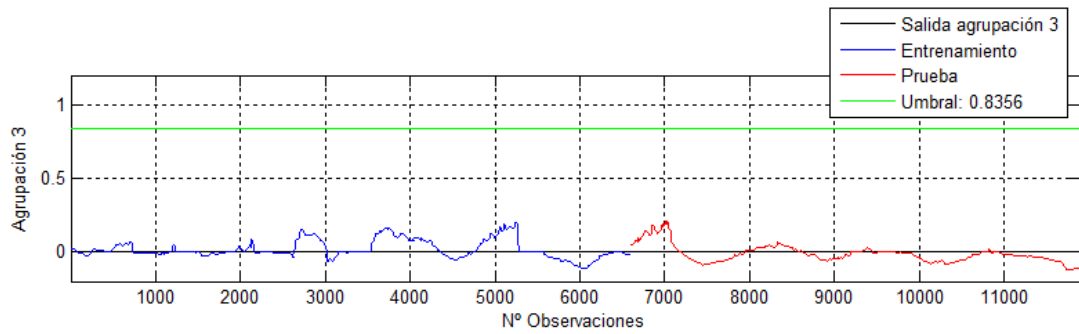


Figura 5.49: Variable de salida “Agrupación 3” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde).

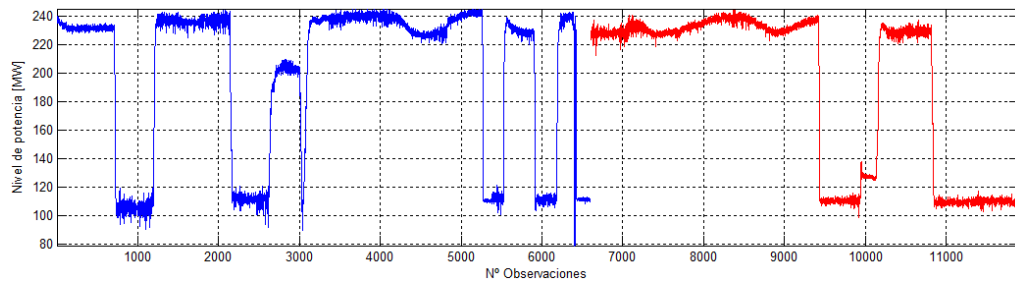


Figura 5.50: Potencia activa en la base de datos A con los conjuntos de entrenamiento (azul) y de prueba en la anomalía (rojo).

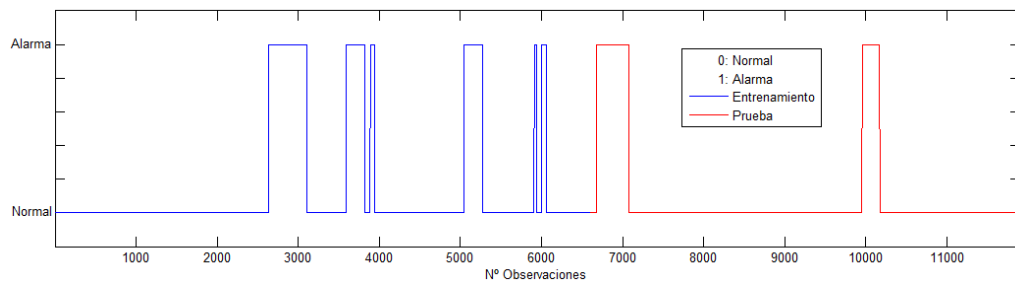


Figura 5.51: Señal de alarma en la base de datos A con los conjuntos de entrenamiento (azul) y de prueba en la anomalía (rojo).

Como se ha mencionado anteriormente, uno de los objetivos del modelo clasificador utilizado en la Fase I del sistema supervisor, es que sea capaz de detectar todas las anomalías en estudio (en esta aplicación, la Anomalía #1 y Anomalía #2). A continuación, en la Figura 5.52, Figura 5.53 y Figura 5.54 se observan los resultados en la base de datos B con las tres agrupaciones encontradas. Al igual que en la Anomalía #1, se observan la salida obtenida a través de las agrupaciones (negro), la salida estimada con el modelo SBM en los conjuntos de entrenamiento (azul) y de prueba (rojo) y el umbral de decisión.

En la Figura 5.56 se encuentra la señal de alerta de los conjuntos de entrenamiento (azul) y prueba (rojo) de la base de datos B. La Anomalía #2 ocurre a lo largo de todo el conjunto de prueba, por lo que se esperaría que este gráfico tuviera valor 0 en todo el conjunto de entrenamiento y 1 en el conjunto de prueba. Se ha utilizado además una mediana móvil de largo $k = 5$, menor que en el caso anterior, debido a que las bases de datos difieren en el tiempo de muestreo y en las características físicas de la turbina (los cambios en sus componentes entre los años 2008 y 2011).

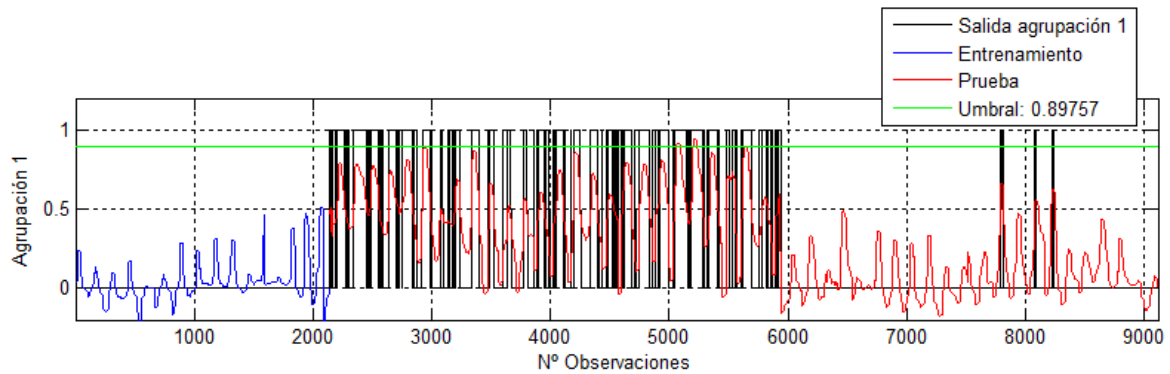


Figura 5.52: Variable de salida “Agrupación 1” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde).

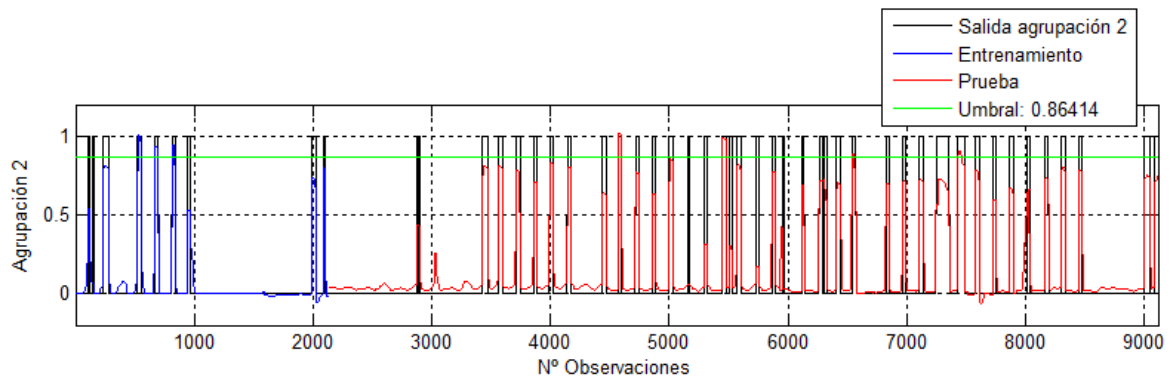


Figura 5.53: Variable de salida “Agrupación 2” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde).

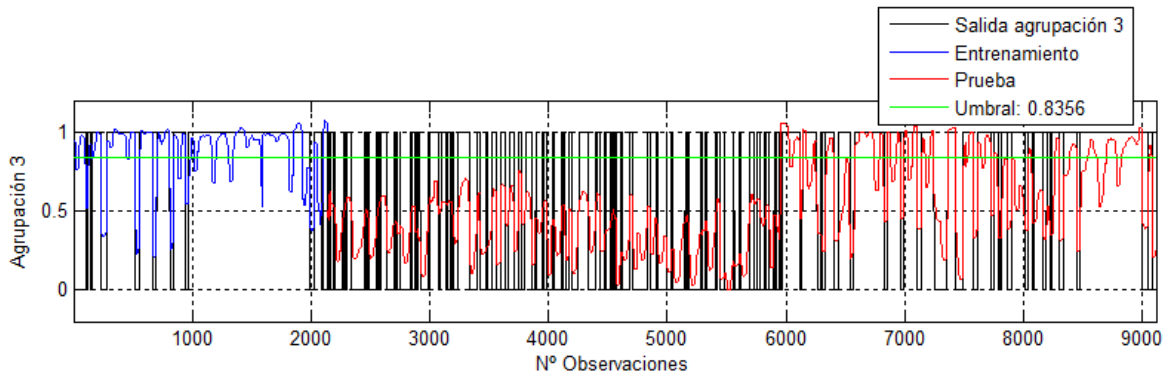


Figura 5.54: Variable de salida “Agrupación 3” (negro); salida estimada con SBM en el conjunto de entrenamiento (azul) y conjunto de prueba (rojo); umbral de decisión (verde).

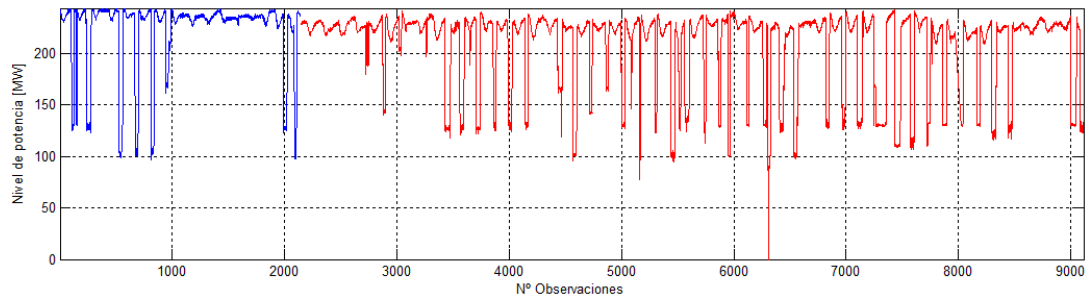


Figura 5.55: Potencia activa en la base de datos B con los conjuntos de entrenamiento (azul) y de prueba en la anomalía (rojo).

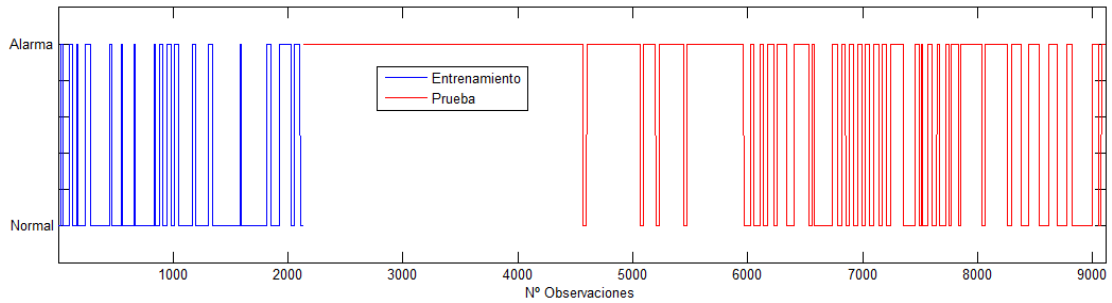


Figura 5.56: Señal de alarma en la base de datos B con los conjuntos de entrenamiento (azul) y de prueba en la anomalía (rojo).

A continuación, en la Tabla 5.10 y Tabla 5.11 se encuentran las tablas de contingencia del estado del sistema con el modelo clasificador en la Anomalía #1 y Anomalía #2, respectivamente.

Tabla 5.10: Tabla de contingencia de la Anomalía #1 en los conjuntos de entrenamiento y prueba de la base de datos A.

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|--------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalías | Normal | |
| Condición estimada del proceso | Anomalías | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | 179 (98.35%) | 1678 (14.32%) | 1857 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 3 (1.65%) | 10042 (85.68%) | 10045 | |
| Total condición real | | 182 | 11720 | 11902 |

Tabla 5.11: Tabla de contingencia de la Anomalía #2 en los conjuntos de entrenamiento y prueba de la base de datos B.

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|---------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalías | Normal | |
| Condición estimada del proceso | Anomalías | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | 5391 (77.33%) | 516 (24.09%) | 5907 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 1581 (22.67%) | 1626 (75.91%) | 3207 | |
| Total condición real | | 6972 | 2142 | 9114 |

En la Anomalía #1, se observa que el 98.35% de observaciones con la anomalía son detectados correctamente, teniéndose un índice de error cercano al 1%. Además, en la condición normal, aproximadamente el 85% es clasificado como una operación normal, mientras que se produce un 14% de falsas alarmas. En la Anomalía #2, la tasa de detección es del 77% del total de observaciones que presentan anomalía, mientras que el nivel de error alcanza el 22%. Por otro lado, cerca del 76% es correctamente clasificado en una condición normal, mientras que en el 24% restante se producen falsas alarmas.

5.4 SISTEMA SUPERVISOR

El sistema supervisor contempla tres fases y tiene por objetivo detectar a través de los modelos generados, anomalías que ya hayan sido registradas anteriormente en un proceso. El sistema supervisor se ha generado en base al estudio de dos anomalías: temperatura excedida en álabe N°20 (Sección 5.2.1) y suciedad en el compresor (Sección 5.2.2) ambas afectando al rendimiento de la turbina con respecto a la potencia activa y el nivel de combustible utilizado. Con las dos anomalías ya estudiadas, se procede a la generación del modelo clasificador (Sección 5.3).

En este punto, se plantea la siguiente interrogante, *¿el modelo M_1 de la Anomalía #1 es capaz de detectar además la Anomalía #2, y viceversa?* En la Sección 5.4.1 se resuelve esta duda observando los indicadores de Hotelling de los dos modelos y realizando los reajustes que sean necesarios. Finalmente, en la Sección 5.4.2 y Sección 5.4.3 se muestran los resultados del sistema supervisor en las bases de datos A y B respectivamente.

5.4.1 REAJUSTES EN LOS MODELOS DEL SISTEMA SUPERVISOR

A continuación, en la Figura 5.58 y Figura 5.60 se observan los indicadores de Hotelling al utilizar los modelos M_1 y M_2 en las bases de datos B y A (Anomalía #2 y Anomalía #1) respectivamente, en los conjuntos de entrenamiento, validación y prueba en la anomalía.

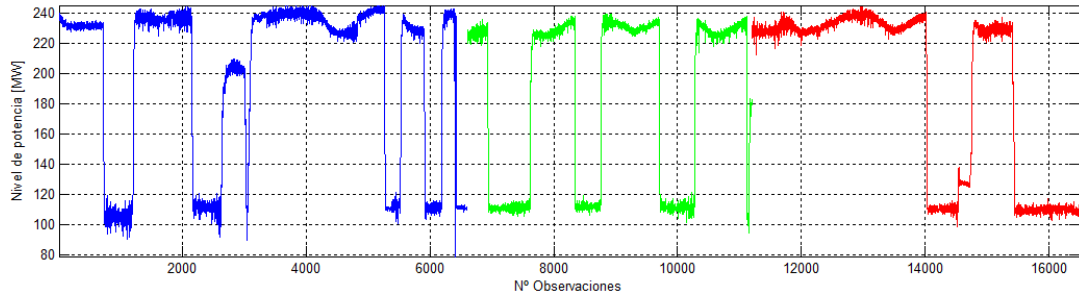


Figura 5.57: Potencia activa en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) en base de datos A.

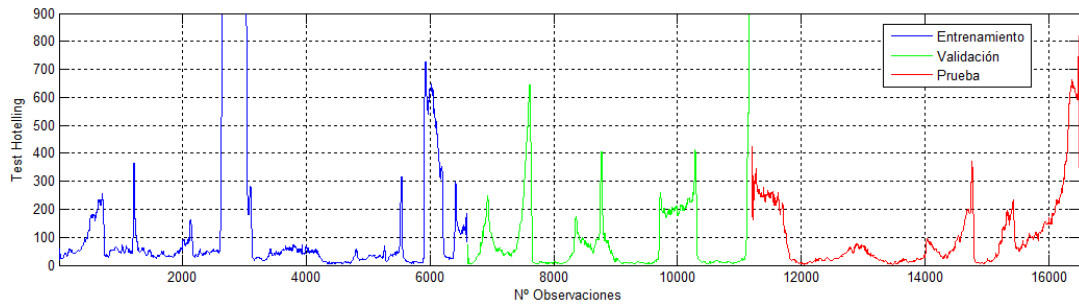


Figura 5.58: Indicador de Hotelling de los residuos con el modelo M_2 en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) en base de datos A (Anomalía #1).

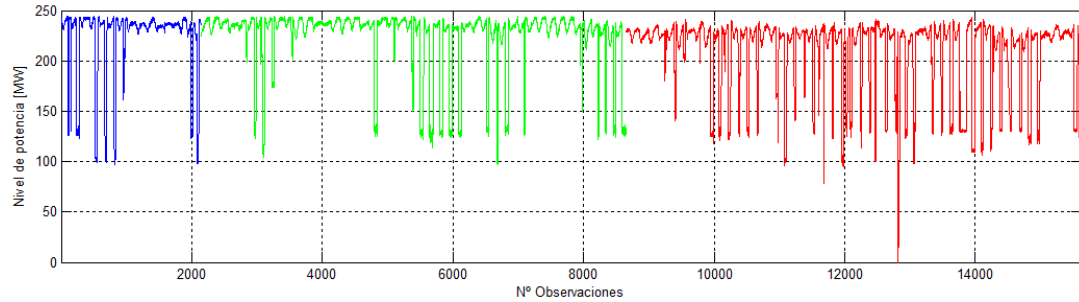


Figura 5.59: Potencia activa en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) en base de datos B.



Figura 5.60: Indicador de Hotelling de los residuos con el modelo M_1 en los conjuntos de entrenamiento (azul), validación (verde) y prueba en la anomalía (rojo) en base de datos B (Anomalía #2).

Con respecto a la Figura 5.58 de la base de datos A, en las observaciones cercanas a 3000 y a 11000 el índice de Hotelling alcanza un valor de 13000. La razón de esto es que la turbina se encuentra operando con combustible diesel, mientras que el modelo M_2 ha sido generado con la base de datos B, la cual no cuenta con observaciones de la turbina en este tipo de operación. En cuanto a la detección de la Anomalía #1, se observa que esta no puede ser detectada utilizando el modelo M_2 ya que cualitativamente no hay diferencias entre los tres conjuntos. En la Figura 5.60 correspondiente a los resultados en la base de datos B, se observa que el indicador de Hotelling solo genera alarma en algunas observaciones del conjunto de prueba y no en todas como debiese ser. Además, en los conjuntos de entrenamiento y validación, también se producen señales de alerta de la anomalía. Se concluye, del mismo modo que en el caso anterior, que la Anomalía #2 no es posible detectarla utilizando el modelo M_1 .

Con estos resultados, se determina agrupar (tal como ocurrió con el modelo clasificador) en un solo conjunto las observaciones utilizadas para generar y validar los modelos M_1 y M_2 . Esto se realiza con el objetivo de incorporar información de la base de datos A al modelo M_2 , e información de la base de datos B al modelo M_1 . Así, los conjuntos de entrenamiento, validación y prueba de los modelos M_1 y M_2 corresponden a la unión de los conjuntos observados en la Figura 5.9 y Figura 5.27. Con esto, la diferencia entre ambos modelos no radica en las observaciones utilizadas para generarlos, sino en las variables de entrada y salida.

Además, se genera el modelo M_3 que representa los instantes en que el proceso se encuentra durante la Anomalía #2. En los conjuntos de entrenamiento y validación del modelo M_3 (Figura 5.61), el sistema se encuentra en una condición anómala, mientras que en el conjunto de prueba el sistema se encuentra en una condición normal. Debido a que esta anomalía ha sido registrada solo en la base de datos B, entonces los conjuntos de entrenamiento y validación utilizarán información de esta solamente.

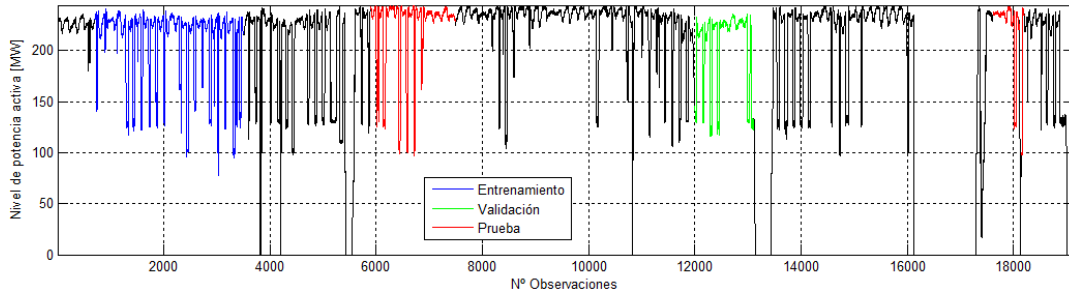


Figura 5.61: Potencia activa en los conjuntos de entrenamiento (azul) y validación (verde) en la base de datos B para el modelo M_3 .

Con los conjuntos ya determinados, se procede a la generación de los modelos M_1 , M_2 y M_3 a través de la herramienta explicada en la Sección 3.1. A continuación se observan los indicadores de Hotelling en los conjuntos de entrenamiento, validación y prueba en la anomalía de los modelos M_1 , M_2 y M_3 .

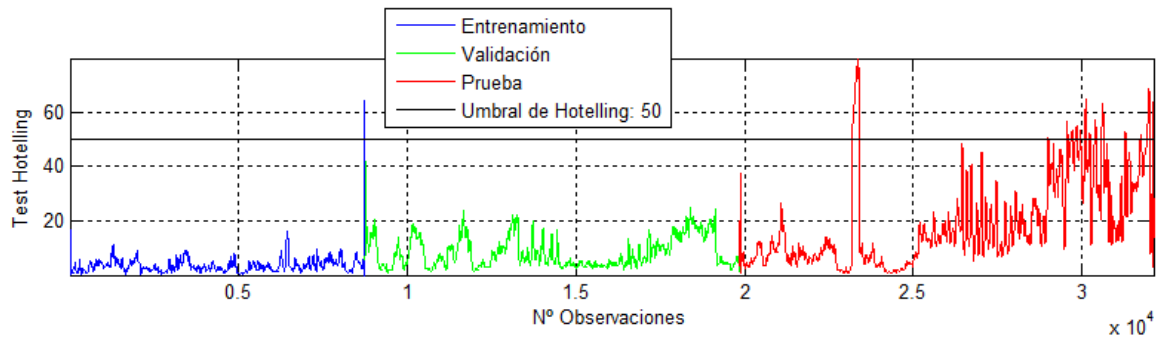


Figura 5.62: Indicador de Hotelling de los residuos con el modelo M_1 en los conjuntos de entrenamiento (azul), validación (verde) y prueba (rojo).

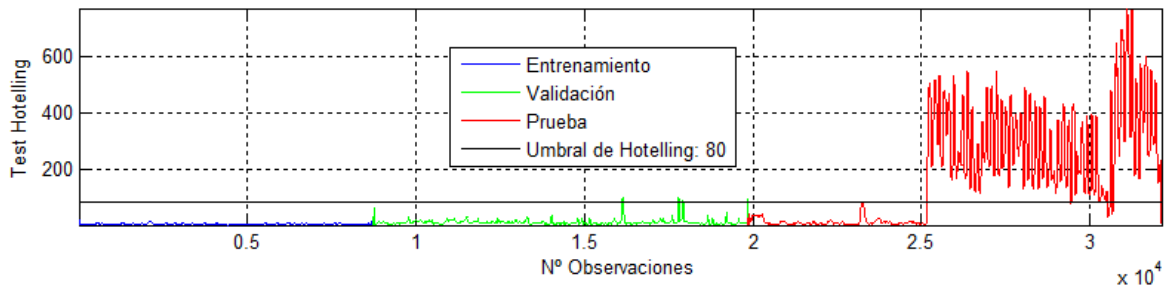


Figura 5.63: Indicador de Hotelling de los residuos con el modelo M_2 en los conjuntos de entrenamiento (azul), validación (verde) y prueba (rojo).

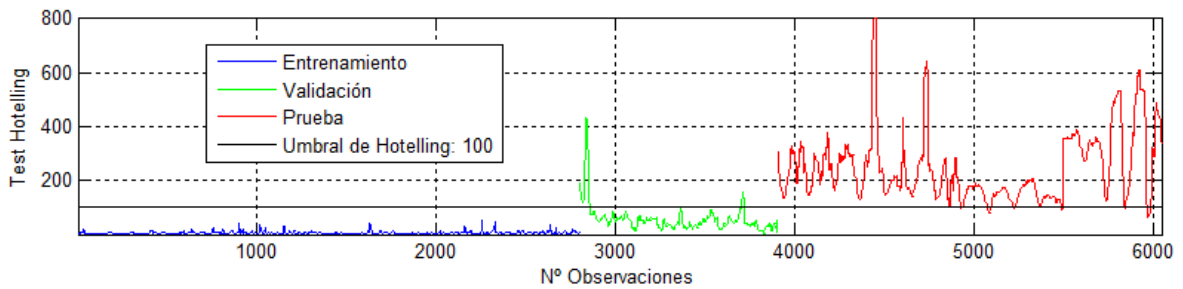


Figura 5.64: Indicador de Hotelling de los residuos con el modelo M_3 en los conjuntos de entrenamiento (azul), validación (verde) y prueba (rojo).

Se observa en este caso que el modelo M_1 detecta efectivamente la Anomalía #1 (alrededor de 24000) y además logra que los datos de la base de datos B se mantengan en un gran porcentaje bajo el umbral de Hotelling. Por otro lado, en el modelo M_2 el indicador de Hotelling se mantiene bajo el umbral en todos los instantes de tiempo (incluso cuando ocurre la Anomalía #1), y solo lo supera en presencia de la Anomalía #2. Con respecto al modelo M_3 , se observa a través de los conjuntos de entrenamiento y validación que cumple el objetivo de modelar en los instantes en que ocurre la anomalía, y cuando el proceso se encuentra en una condición normal, este indicador supera el umbral.

Como el modelo clasificador ya considera observaciones de las dos bases de datos, no es necesario realizar modificaciones como las efectuadas con los modelos de anomalías registradas. Con los modelos ya reajustados, se procede a utilizar las dos bases de datos para obtener las señales discretas que indican la condición en la que se encuentra el proceso en cada instante. Cuando la turbina no se encuentra en operación, se genera una señal de “Apagada”. Se obtienen las señales de alarma del modelo clasificador y de los tres modelos de anomalías registradas. Además, se genera la señal de estados del sistema supervisor. Esta señal indica la condición final en la que se encuentra el proceso en base a los resultados de los modelos. Los estados en los que se puede encontrar el sistema son los indicados en la Tabla 5.12.

Tabla 5.12: Tabla de estados del sistema supervisor

| | M_1 | M_2 | M_3 | Estado | | M_1 | M_2 | M_3 | Estado |
|-----------|-------|-------|-------|--------------------|-----------|-------|-------|-------|-------------------------|
| 1 | 0 | 0 | 0 | <i>Normal</i> | 15 | 1 | 1 | 2 | <i>Desconocido</i> |
| 2 | 0 | 0 | 1 | <i>Normal</i> | 16 | 1 | 2 | 0 | <i>Desconocido</i> |
| 3 | 0 | 0 | 2 | <i>Desconocido</i> | 17 | 1 | 2 | 1 | <i>Desconocido</i> |
| 4 | 0 | 1 | 0 | <i>Normal</i> | 18 | 1 | 2 | 2 | <i>Anomalía #2</i> |
| 5 | 0 | 1 | 1 | <i>Normal</i> | 19 | 2 | 0 | 0 | <i>Anomalía #1</i> |
| 6 | 0 | 1 | 2 | <i>Desconocido</i> | 20 | 2 | 0 | 1 | <i>Anomalía #1</i> |
| 7 | 0 | 2 | 0 | <i>Desconocido</i> | 21 | 2 | 0 | 2 | <i>Desconocido</i> |
| 8 | 0 | 2 | 1 | <i>Desconocido</i> | 22 | 2 | 1 | 0 | <i>Anomalía #1</i> |
| 9 | 0 | 2 | 2 | <i>Anomalía #2</i> | 23 | 2 | 1 | 1 | <i>Anomalía #1</i> |
| 10 | 1 | 0 | 0 | <i>Normal</i> | 24 | 2 | 1 | 2 | <i>Desconocido</i> |
| 11 | 1 | 0 | 1 | <i>Normal</i> | 25 | 2 | 2 | 0 | <i>Desconocido</i> |
| 12 | 1 | 0 | 2 | <i>Desconocido</i> | 26 | 2 | 2 | 1 | <i>Desconocido</i> |
| 13 | 1 | 1 | 0 | <i>Normal</i> | 27 | 2 | 2 | 2 | <i>Anomalía #1 y #2</i> |
| 14 | 1 | 1 | 1 | <i>Normal</i> | | | | | |

Donde el significado de “0”, “1” y “2” para los modelos M_1 , M_2 y M_3 puede ser “*Normal (Clasificador)*” si el modelo clasificador genera una señal “*normal*” o “*apagada*”; “*Normal*” si los modelos de anomalías generan una señal “*Normal*”; o “*Anomalía*” si los modelos detectan alguna condición anómala.

Tabla 5.13: Significado de las señales de alarma en cada modelo

| | M_1 | M_2 | M_3 |
|----------|------------------------------|------------------------------|------------------------------|
| 0 | <i>Normal (Clasificador)</i> | <i>Normal (Clasificador)</i> | <i>Normal (Clasificador)</i> |
| 1 | <i>Normal</i> | <i>Normal</i> | <i>Normal</i> |
| 2 | <i>Anomalía #1</i> | <i>Anomalía #2</i> | <i>Anomalía #2</i> |

5.4.2 RESULTADOS EN LA BASE DE DATOS A

A continuación se observa en la Figura 5.65 la potencia activa en la base de datos A con los datos de la Anomalía #1 coloreados en rojo. Las señales de alarma del modelo clasificador y de los modelos M_1 , M_2 y M_3 se observan en la Figura 5.66, Figura 5.67, Figura 5.68 y Figura 5.69. La señal de estados del sistema supervisor se encuentra en la Figura 5.70.

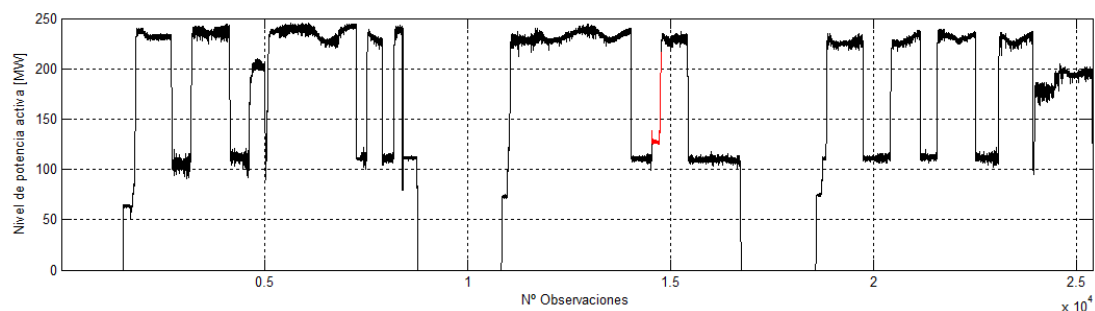


Figura 5.65: Potencia activa en la base de datos A en presencia de la Anomalía #1.

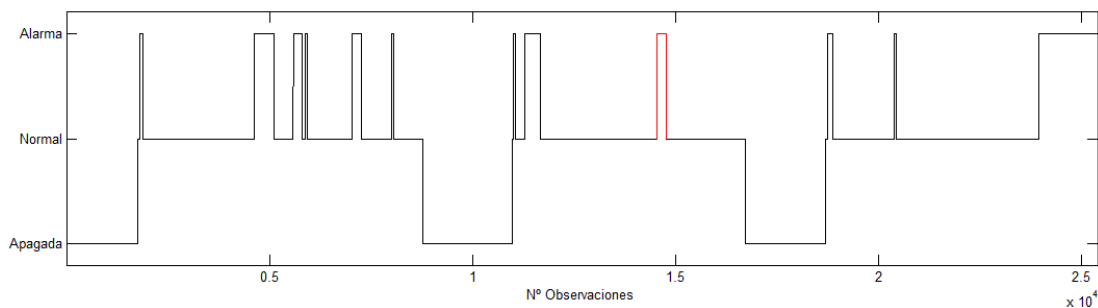


Figura 5.66: Señal de alarma del modelo clasificador en la base de datos A.

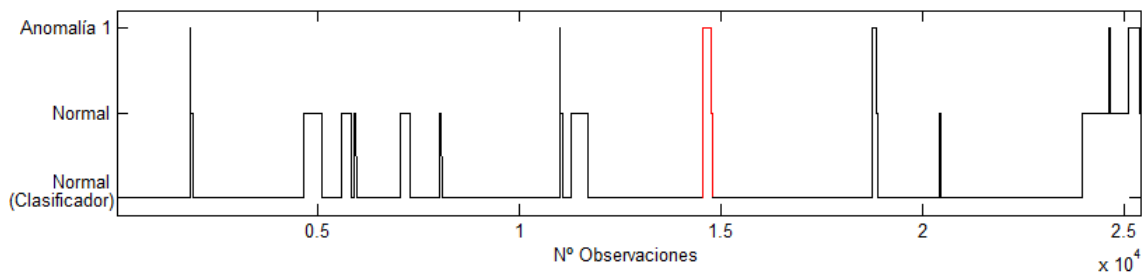


Figura 5.67: Señal de alarma del modelo M_1 en la base de datos A.

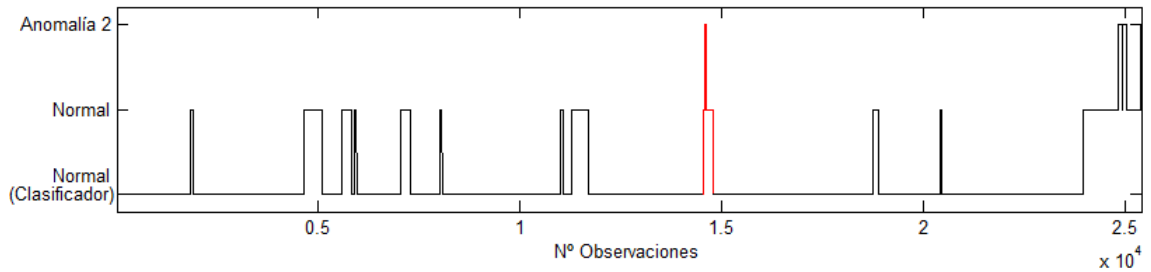


Figura 5.68: Señal de alarma del modelo M_2 en la base de datos A.

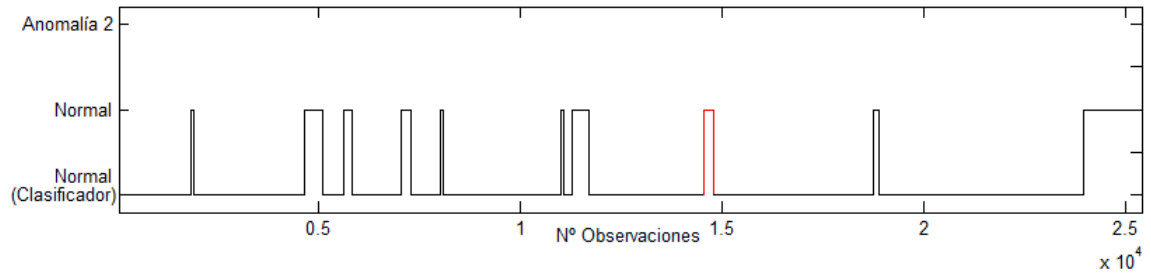


Figura 5.69: Señal de alarma del modelo M_3 en la base de datos A.

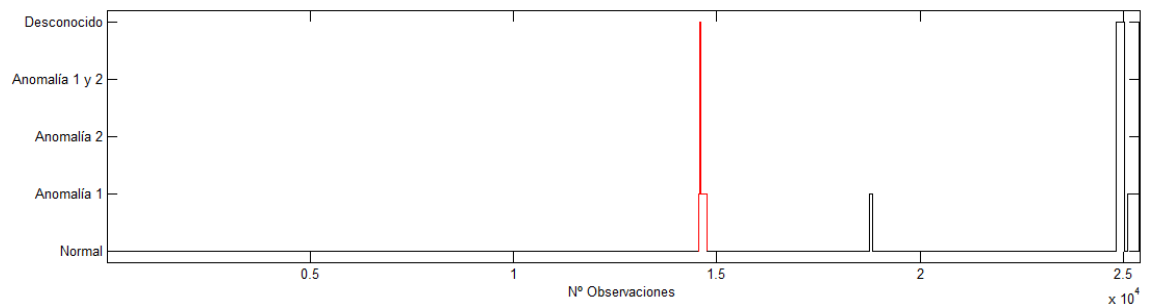


Figura 5.70: Señal de estado del proceso en la base de datos A.

En la Figura 5.66 se observa que el modelo clasificador detecta la anomalía, pero al mismo tiempo, aparece una serie de observaciones con una señal de alarma errónea. Hay que recalcar nuevamente que estos errores se producen en gran parte en los cambios en los niveles de potencia y en los instantes en que la turbina opera con combustible diesel (alrededor de 5000 y de 25000). En la Figura 5.67 los niveles de alarmas erróneas disminuyen considerablemente en comparación a las obtenidas con el modelo clasificador. Además se cumple el objetivo de detectar la Anomalía #1.

Por otro lado, de la Figura 5.68 y Figura 5.69 se observa que en los instantes en que ocurre la anomalía, los modelos M_2 y M_3 no generan una señal de alarma. Esto demuestra que esta anomalía efectivamente corresponde a un problema en la temperatura de los álabes y no en la suciedad del compresor. Los errores en la señal de alarma del modelo M_2 que indican que en el proceso hay suciedad en el compresor (cuando en teoría no debiera detectarse), son corregidos con los resultados obtenidos con el modelo M_3 que muestran que el proceso efectivamente se encuentra en estado normal. La señal de estado del proceso de la Figura 5.70 muestra que la anomalía es correctamente detectada y que solo se producen falsas alarmas en dos ocasiones.

A continuación se encuentran las tablas de contingencia de los modelos generados al ser aplicados en la base de datos A. En la Tabla 5.14 se encuentra la tabla del modelo clasificador. Se ha determinado que las observaciones que han sido etiquetadas con “Apagada” no sean consideradas en estos porcentajes. De este modo, la tabla solo relaciona aquellos instantes que han sido considerados en los estados normal/alarma. Las tablas de contingencia de los tres modelos de anomalías se encuentran en la Tabla 5.15, Tabla 5.16 y Tabla 5.17. En estos casos, se muestran los resultados teniendo en cuenta los estados *Normal/Alarma* de Anomalía #1 en el modelo M_1 y *Normal/Alarma* de Anomalía #2 en los modelos M_2 y M_3 . Finalmente, la Tabla 5.18 entrega el porcentaje de detección de la señal de estado en la base de datos A.

Tabla 5.14: Tabla de contingencia del modelo clasificador en la base de datos A.

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|--------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalías | Normal | |
| Condición estimada del proceso | Anomalías | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | 218 (96.46%) | 3167 (16.44%) | 3385 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 8 (3.53%) | 16086 (83.55%) | 16094 | |
| Total condición real | | 226 | 19253 | 19479 |

Tabla 5.15: Tabla de contingencia del modelo M_1 en la base de datos A.

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|--------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalía #1 | Normal | |
| Condición estimada del proceso | Anomalía #1 | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | 203 (89.82%) | 424 (2.2%) | 627 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 23 (10.17%) | 18829 (97.7%) | 18852 | |
| Total condición real | | 226 | 19253 | 19479 |

Tabla 5.16: Tabla de contingencia del modelo M_2 en la base de datos A.

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|-------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalía #2 | Normal | |
| Condición estimada del proceso | Anomalía #2 | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | 0 | 245 (1.25%) | 245 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 0 | 19234 (98.74%) | 19234 | |
| Total condición real | | 0 | 19479 | 19479 |

Tabla 5.17: Tabla de contingencia del modelo M_3 en la base de datos A.

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|-------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalía #2 | Normal | |
| Condición estimada del proceso | Anomalía #2 | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | 0 | 0 (0.00%) | 0 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 0 | 19479 (100%) | 19479 | |
| Total condición real | | 0 | 19479 | 19479 |

Tabla 5.18: Porcentajes de detección de la señal de estado en la base de datos A.

| Estado | Observaciones estimadas | Total |
|-------------|-------------------------|-------|
| Normal | 24576 (97.64%) | 25168 |
| Anomalía #1 | 166 (73.45%) | 226 |

Se observa que cerca de un 90% de la Anomalía #1 es detectado, mientras que hay un 2% de falsas alarmas. A pesar que los conjuntos son diferentes a los que se utilizan en la Sección 5.2.1, se observa que los porcentajes no difieren significativamente. La diferencia radica en el total de datos del proceso en estado normal. Sin embargo, la relación entre falsos positivos y verdaderos negativos, se mantiene. Por otro lado, las falsas alarmas que genera el modelo M_2 son corregidas por el modelo M_3 el cual obtuvo un 100% de verdaderos negativos. Aquellas observaciones que eran detectadas como Anomalía #1, pero que al mismo tiempo se producía una contradicción entre los modelos M_2 y M_3 (uno indicara un estado normal, mientras que el otro un estado de anomalía), ahora son catalogados con el estado “Desconocido”. Esto tiene como consecuencia, una disminución considerable en la capacidad de detección del sistema supervisor de un 89.82% al 73.45%. En la práctica, se recomienda que este estado genere de todos modos una señal que alerte a los operadores de una condición anormal en el proceso.

5.4.3 RESULTADOS EN LA BASE DE DATOS B

En la Figura 5.71 se observa la potencia activa en la base de datos B, indicando en color rojo las observaciones en donde se ha registrado la suciedad en el compresor que afecta el rendimiento de la turbina. En la Figura 5.72 se muestran los resultados del modelo clasificador. Los resultados de los modelos M_1 , M_2 y M_3 se encuentran en la Figura 5.73, Figura 5.74 y Figura 5.75 respectivamente. Finalmente, en la Figura 5.76 y Figura 5.77 se encuentra la señal de estado del proceso en la base de datos B.

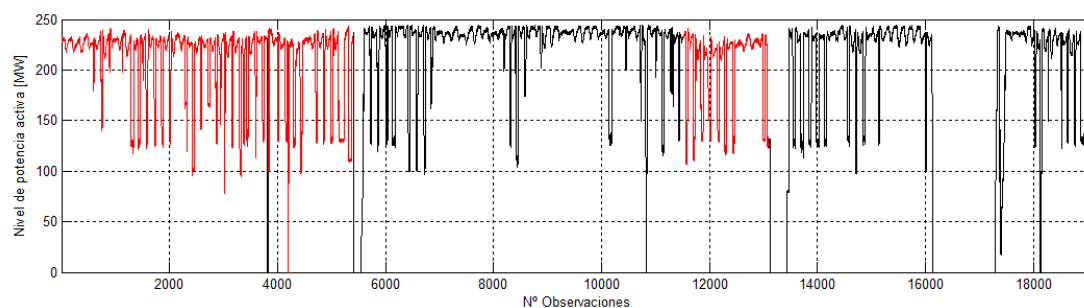


Figura 5.71: Potencia activa en la base de datos B en presencia de la Anomalia #2.

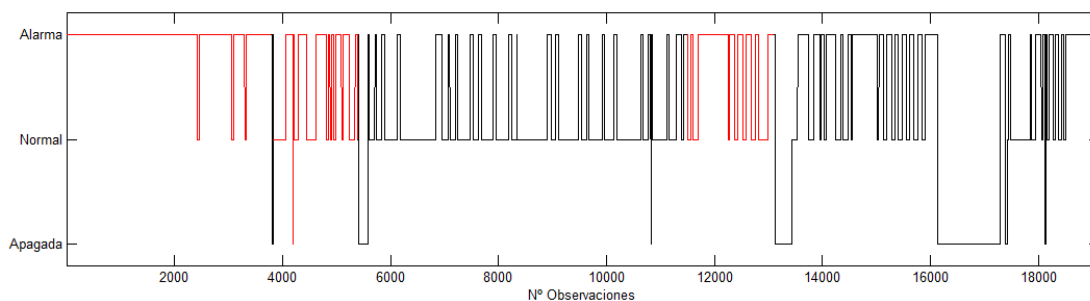


Figura 5.72: Señal de alarma del modelo clasificador en la base de datos B.

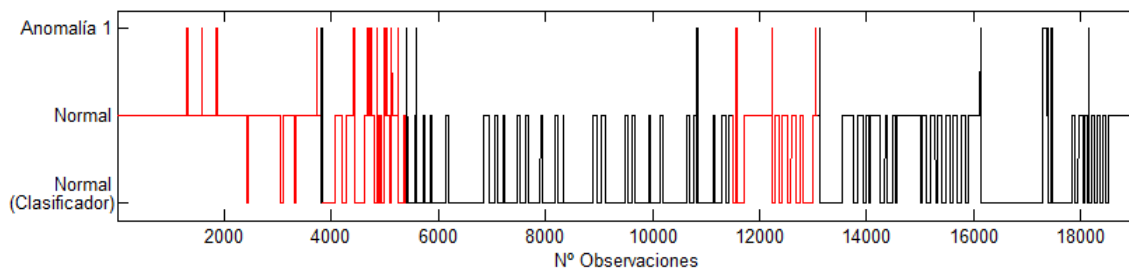


Figura 5.73: Señal de alarma del modelo M_1 en la base de datos B.

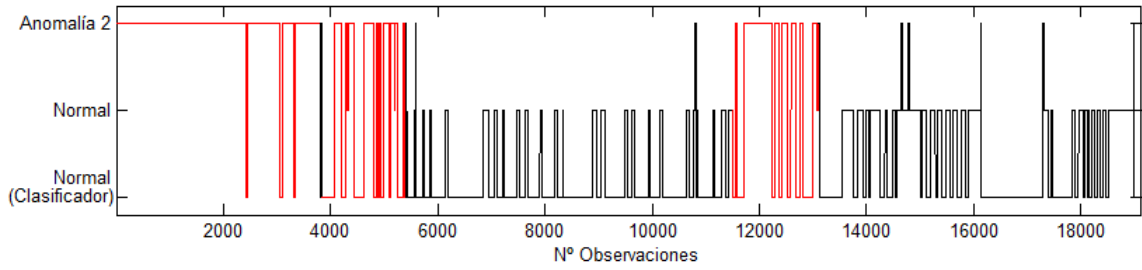


Figura 5.74: Señal de alarma del modelo M_2 en la base de datos B.

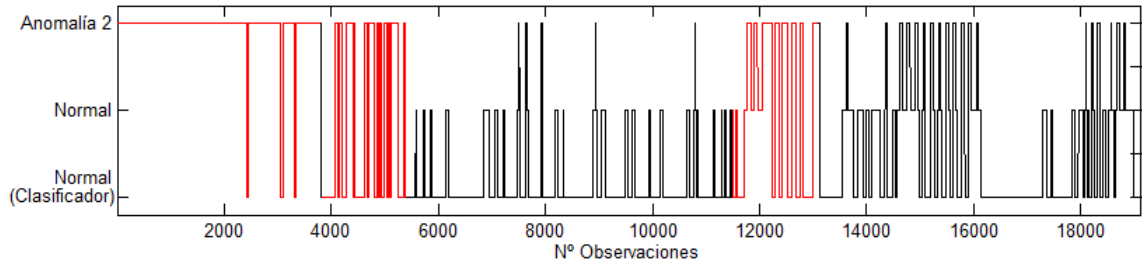


Figura 5.75: Señal de alarma del modelo M_3 en la base de datos B.

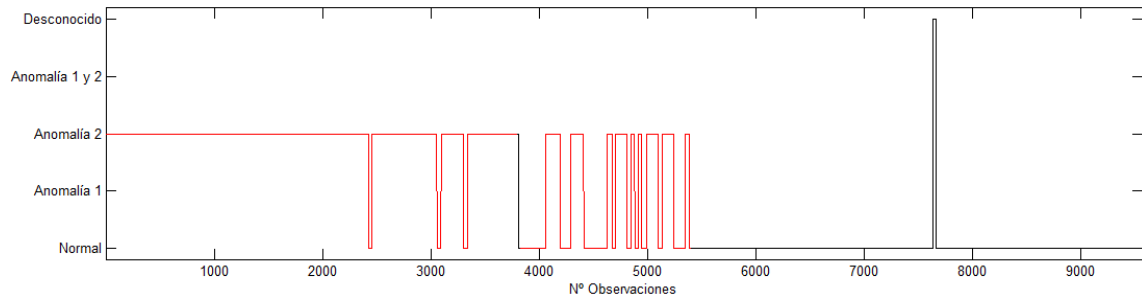


Figura 5.76: Señal de estado del proceso en la base de datos B (observaciones 1-9567).

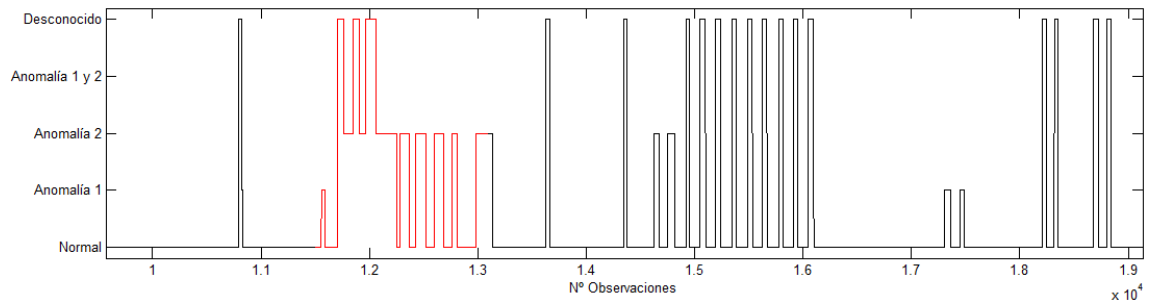


Figura 5.77: Señal de estado del proceso en la base de datos B (observaciones 9568-19134).

A diferencia del caso anterior, en la base de datos B se observa que el modelo clasificador establece que un porcentaje no menor de observaciones se encuentra en estado normal, cuando debiesen ser consideradas como alarmas. Esto tiene como consecuencia que los modelos de anomalías no sean utilizados en estos instantes, y por lo tanto la señal de estado del sistema supervisor concluye erróneamente que son observaciones en estado normal. Otro porcentaje importante de observaciones es considerado en estado “Desconocido” debido a que el modelo M_2 indica presencia de anomalía, mientras que el modelo M_3 no, y viceversa. A priori, si el estado es desconocido se debiese decretar una alarma como precaución.

A continuación, para poder medir el desempeño de los modelos en la base de datos B, se encuentran las tablas de contingencia de estas observaciones en el modelo clasificador en la Tabla 5.19 y en los modelos de anomalías registradas M_1 , M_2 y M_3 en la Tabla 5.20, Tabla 5.21 y Tabla 5.22, respectivamente. Se consideran los mismos supuestos que en la Sección 5.4.2 en la base de datos A.

Tabla 5.19: Tabla de contingencia del modelo clasificador en la base de datos B.

| | | Condición real del proceso | | |
|--------------------------------------|-----------|----------------------------|--------------------------------------|---|
| | | Anomalías | Normal | |
| Condición estimada del proceso | Anomalías | Verdaderos positivos | Falsos positivos (o falsa alarma) | Total condición estimada |
| | | 5382 (77.19%) | 3697 (35.83%) | |
| | Normal | Falsos negativos | Verdaderos negativos | |
| | | 1590 (22.80%) | 6620 (64.16%) | 8210 |
| Total condición real | | 6972 | 10317 | 17289 |

Tabla 5.20: Tabla de contingencia del modelo M_1 en la base de datos B.

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|-------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalía #1 | Normal | |
| Condición estimada del proceso | Anomalía #1 | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | 0 | 319 (1.84%) | 319 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 0 | 16970 (98.15%) | 16970 | |
| Total condición real | | 0 | 17289 | 17289 |

Tabla 5.21: Tabla de contingencia del modelo M_2 en la base de datos B.

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|---------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalía #2 | Normal | |
| Condición estimada del proceso | Anomalía #2 | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | 5319 (76.29%) | 133 (1.28%) | 5452 | |
| Condición estimada del proceso | Normal | Falsos negativos | Verdaderos negativos | |
| | 1653 (23.70%) | 10184 (98.71%) | 11837 | |
| Total condición real | | 6972 | 10317 | 17289 |

Tabla 5.22: Tabla de contingencia del modelo M_3 en la base de datos B.

| | | Condición real del proceso | | Total condición estimada |
|--------------------------------|-------------|----------------------------|-----------------------------------|--------------------------|
| | | Anomalía #2 | Normal | |
| Condición estimada del proceso | Anomalía #2 | Verdaderos positivos | Falsos positivos (o falsa alarma) | |
| | | 5127 (73.53%) | 898 (8.70%) | 6025 |
| | Normal | Falsos negativos | Verdaderos negativos | |
| | | 1845 (26.46%) | 9419 (91.29%) | 11264 |
| Total condición real | | 6972 | 10317 | 17289 |

Tabla 5.23: Porcentajes de detección de la señal de estado en la base de datos B.

| Estado | Observaciones estimadas | Total |
|-------------|-------------------------|-------|
| Normal | 11226 (92.30%) | 12162 |
| Anomalía #2 | 5186 (74.38%) | 6972 |

Con respecto al modelo clasificador, se observa que hay un porcentaje considerable del 22.8% de observaciones que son consideradas en estado normal, cuando en la realidad corresponden a anomalías. Esto significa que al menos ese porcentaje no logra ser detectado por el sistema supervisor. Al incorporar los modelos M_2 y M_3 se observa que no se producen cambios significativos en los porcentajes de detección. Con respecto al porcentaje de falsas alarmas, este cae de manera importante desde un 35.8% a un valor cercano al 10%. El modelo M_1 genera un 2% de falsas alarmas. Tal como se observa en Tabla 5.23, la señal de estado de la Anomalía #2 entrega un porcentaje de detección del 74.38%. Así como en la Anomalía #1, esto ocurre principalmente debido al nuevo estado del sistema “Desconocido”.

Tal como se mencionó anteriormente, es importante recalcar que estos indicadores porcentuales se calculan en función del número de muestras en las que estos eventos son o no detectados y con respecto al universo de datos disponibles. Así, se requiere de una cantidad considerablemente mayor de datos para caracterizar estos indicadores como *probabilidad de detección* o *probabilidad de falsa alarma*.

CAPÍTULO 6: CONCLUSIONES

Con el aumento en la complejidad de los procesos, la industria moderna requiere de sistemas de monitoreo que sean capaces de proteger los equipos y de disminuir los daños y los costos que ocasionan aquellas perturbaciones que no pueden ser reguladas por los controladores. Es por estas motivaciones que se propuso como objetivo general el diseño de un sistema supervisor en línea para detección y aislamiento de fallas utilizando modelos basados en similitud.

Para el diseño del sistema supervisor se planteó una estructura que en una primera etapa detectara cualquiera de las anomalías en estudio, y que en una segunda etapa indicara de cual se trata. Para lograr tal objetivo, se generaron dos tipos de modelos que fueron utilizados en la primera y segunda etapa respectivamente: modelo clasificador SBM y modelo de anomalía SBM. El modelo clasificador se basó en la utilización de las componentes principales para asignar a cada vector de datos alguna agrupación. La señal de alarma de detección de anomalías se producía si el vector de datos escapaba de los límites (umbrales) de todas las agrupaciones. Con el modelo de anomalías se obtenían las salidas estimadas y se comparaban con las salidas reales para obtener los residuos, y con estos, el indicador de Hotelling. La condición para generar una alarma es que este indicador supere el umbral de Hotelling. Finalmente, se genera una señal del estado del proceso. Además, de la generación de los modelos, era necesario disponer de una metodología de estudio de eventos con el objetivo de seleccionar y descartar las variables que estarían en tales modelos.

El sistema supervisor se utilizó en un proceso consistente en una turbina a gas de una central de ciclo combinado y que presentaba dos eventos. La Anomalía #1 que consistía en un aumento de la temperatura de un álabe por sobre el límite permitido, lo que obligaba a disminuir los niveles de potencia. La Anomalía #2 consistía en una suciedad excesiva en el compresor que afectaba directamente a los niveles de potencia para un mismo nivel de combustible.

La efectividad del sistema supervisor se midió a través de indicadores que entregaban los porcentajes de detección y los porcentajes de falsas alarmas. Estos indicadores están directamente relacionados con el número de muestras en las que estos eventos son o no detectados y con el tamaño del conjunto de datos disponibles. De este modo, es necesario de una cantidad considerablemente mayor de datos para caracterizar estos indicadores como *probabilidad de detección* o *probabilidad de falsa alarma*.

Al aplicar la metodología de estudio de eventos en la detección de la Anomalía #1 se obtiene un porcentaje de detección que alcanza el 91.9% y un porcentaje de falsas alarmas del 1.3%. La metodología de eventos se encarga de encontrar un conjunto óptimo local de variables de entrada y de salida. Es por esto que los porcentajes de detección y de falsas alarmas podrían eventualmente mejorar si se encuentra el conjunto de variables óptimo global para la anomalía. Otro punto que se concluye de estos resultados es la necesidad de disponer de bases de datos en donde la anomalía se produzca en al menos dos ocasiones y por instantes más prolongados. De este modo, sería posible validar no solo la capacidad de detección de este modelo, sino también la capacidad de modelar dicha anomalía. También se puede observar que la detección ocurrió cuando la Anomalía #1 se inicia, y no antes. A pesar de esto, la importancia del modelo radica en que se obtienen las variables que además de la temperatura, se ven influenciadas por esta anomalía. Esto ayudaría a distinguir entre diferentes tipos de fallas que tienen aparentemente la misma consecuencia. Con respecto a la Anomalía #2, esta ya había sido estudiada en detalle en [1] encontrándose las variables de entrada y de salida del modelo. Los indicadores porcentuales fueron calculados obteniéndose un porcentaje de detección del 98.6% y un 7.5% de falsas alarmas.

El modelo clasificador que se diseñó logra que la Anomalía #1 y la Anomalía #2 tengan tasas de detección de 96.5% y 77.2%, respectivamente. Para aumentar estas tasas de detección y disminuir el porcentaje de falsas alarmas es que se proponen algunas recomendaciones para el trabajo futuro con este tipo de modelos. Los centros de las agrupaciones, así como también la pertenencia de las observaciones a cada agrupación, fueron obtenidas de manera automática. Sin embargo, el número de agrupaciones se determinó de forma manual y solo en las primeras dos componentes principales. Así, se recomienda aumentar el número de componentes e indagar en métodos que determinen el número de agrupaciones. También se puede concluir que es necesario un mayor número de anomalías en estudio, con el objetivo de perfeccionar la metodología de estudio de eventos utilizada para seleccionar las variables de entrada del modelo clasificador.

Debido a los constantes cambios físicos que tuvo la turbina entre los años 2008 y 2011, es que se hizo necesario observar el comportamiento del modelo M_1 en la base de datos B y el del modelo M_2 en la base de datos A. Los resultados mostraron la necesidad de incorporar observaciones de las bases de datos A y B a los modelos M_2 y M_1 respectivamente. Así, es posible concluir que los cambios que se realizan en el proceso pueden invalidar absolutamente un modelo, requiriéndose de nuevas observaciones para generarlo, o realizar un nuevo estudio de variables a través de la metodología de eventos. Además, se generó M_3 que corresponde a un modelo del proceso en presencia de la Anomalía #2. De esta forma hay dos indicadores de Hotelling para detectar esta anomalía.

Con los modelos ya actualizados fue posible simular el sistema supervisor que se planteó en los objetivos, en las bases de datos A y B. Los resultados de los modelos M_1 y M_2 del sistema supervisor muestran que las anomalías #1 y #2 tienen tasas de detección del 96.46% y 76.29%. Sin embargo, estos porcentajes decrecen al 73.45% y 74.38% con la señal de estados del sistema supervisor. Esta diferencia se explica por el porcentaje de observaciones que ahora son calificadas con el estado “*Desconocido*”. De este modo, se recomienda que este estado sea considerado como una advertencia de que hay algún evento anómalo.

Como trabajo futuro, además de las recomendaciones que ya se han descrito, se requiere de la implementación de este esquema de supervisión en una plataforma que reciba los datos, genere y actualice los modelos en línea, y que entregue una señal del estado del proceso. De este modo se evalúan los tiempos de cómputo reales en un proceso y la interacción de los operarios con el sistema supervisor. La empresa CONTAC INGENIEROS LTDA se encuentra actualmente implementando estos algoritmos en una plataforma en línea para aplicarlos en una turbina a gas perteneciente a ENDESA CHILE. Esto corresponde a la segunda fase de este proyecto. También es necesario incorporar el estudio de otras anomalías al sistema supervisor para seguir evaluando las herramientas y las metodologías que han sido diseñadas. Tanto en este trabajo como en [1], los modelos que se generaron eran estáticos. Sin embargo, tal como se mencionó anteriormente, la herramienta de generación de modelos SBM puede ser adaptada para obtener modelos dinámicos del proceso. Debido a que esto no fue desarrollado a lo largo de este proyecto, es que se sugiere como trabajo futuro realizar estudios en un caso dinámico. Finalmente, tal como ocurrió con las señales de la Anomalía #2, se requiere generar herramientas que eliminen las oscilaciones que se producen en las señales de alarma entre los estados “*Anomalía*” y “*Normal*”.

REFERENCIAS

1. A. León, "Detección de anomalías en procesos industriales usando modelos basados en similitud", Enero 2011.
2. B. Bagheri, H. Ahmadi, R. Labbafi, "Application of data mining and feature extraction on intelligent fault diagnosis by Artificial Neural Network and k-nearest neighbor", IEEE ICEM, pp. 1-7, Septiembre 2010.
3. B. Ayhan, M. Chow, M. Song, "Multiple Discriminant Analysis and Neural-Network-Based Monolith and Partition Fault-Detection Schemes for Broken Rotor Bar in Induction Motors", IEEE T. Industrial Electronics, vol. 53, no. 4, pp. 1298-1308, Agosto 2006.
4. B. Trajin, J. Regnier, J. Faucher, "Detection of bearing faults in asynchronous motors using Luenberger speed observer", IEEE Conference Industrial Electronics, pp. 3073-3078, Noviembre 2008.
5. C. Carincotte, S. Derrode, S. Bourennane, "Unsupervised change detection on SAR images using fuzzy hidden Markov chains", IEEE T. Geoscience and Remote Sensing, vol. 44, no. 2, pp. 432- 441, Febrero 2006.
6. C. Huang, T. Chen, C. Huang, "Robust control of induction motor with a neural-network load torque estimator and a neural-network identification", IEEE T. Industrial Electronics, vol. 46, no. 5, pp. 990-998, Octubre 1999.
7. D. Cayrac, D. Dubois, H. Prade, "Handling uncertainty with possibility theory and fuzzy sets in a satellite fault diagnosis application", IEEE T. Fuzzy Systems, vol. 4, no. 3, pp. 251-269, Agosto 1996.
8. D. Gazzana, M. Oliveira, A. Bretas, A. Bettioli, A. Lerm, M. Goncalves, "An expert system for substation fault detection in thermoelectric generation plants" International Symposium Modern Electric Power Systems, Septiembre 2010.
9. D. Trinh, H. Chafouk, "Fault detection and isolation using Kalman filter bank for a wind turbine generator", Mediterranean Conference on Control & Automation, pp. 144-149, Junio 2011.
10. D. Wilson, G. Irwin, "PLS modelling and fault detection on the Tennessee Eastman benchmark", American Control Conference, vol. 6, pp. 3975-3979, 1999.
11. F. Tobar, L. Yacher, R. Paredes, M. Orchard, "Anomaly detection in power generation plants using similarity-based modeling and multivariate analysis", American Control Conference, pp. 1940-1945, Junio-Julio 2011.
12. G. Beale J. Kim, "Fisher discriminant analysis and the T^2 statistic for process fault detection and classification", IEEE Conference Industrial Electronics Society, vol. 3, pp. 1995-2000, Noviembre 2002.
13. G. Gonzalez, R. Paut, A. Cipriano, D. Miranda, G. Ceballos, "Fault detection and isolation using concatenated wavelet transform variances and discriminant analysis", IEEE T. Signal Processing, vol. 54, no. 5, pp. 1727-1736, Mayo 2006.

14. H. Wang, L. Ping, Z. Yuan, "Understanding PCA fault detection results by using expectation analysis method", IEEE Conference on Decision and Control, vol. 4, pp. 4377- 4382, Diciembre 2002.
15. J. F. Araya, "Estimación de parámetros para su aplicación a detección y diagnóstico de fallas", Pontificia Universidad Católica De Chile, Escuela De Ingeniería, Septiembre 2004.
16. J. Fan, R. Guo, S. Chang, J. Lee, "Abnormal trend detection of sequence-disordered data using EWMA method" IEEE/SEMI ASMC, pp. 169-174, Noviembre 1996.
17. J. Lee, B. Kang, K. Shin, S. Kang, "Online process monitoring scheme for fault detection based on Independent Component Analysis (ICA) and Local Outlier Factor (LOF)", International Conference on Computers and Industrial Engineering (CIE), Julio 2010.
18. I. Guney, E. Kilic, O. Ozgonenel, M. Ulutas, E. Karadeniz, "Fault detection in induction motors with independent component analysis (ICA)", IEEE Bucharest PowerTech Conference, Junio-Julio 2009.
19. L. Chiang, E. Russell, R. Braatz, Fault detection and diagnosis in industrial systems, Great Britain: Springer, pp. 3-11, 2001.
20. L. Chiang, E. Russell, R. Braatz, Fault detection and diagnosis in industrial systems, Great Britain: Springer, pp. 21-25; 35-38, 2001.
21. L. Gong, D. Schonfeld, "Space Kernel Analysis", IEEE ICASSP, pp. 1577-1580, Abril 2009.
22. M. Fuente, D. Garcia-Alvarez, G. Sainz-Palmero, T. Villegas, "Fault detection and identification method based on multivariate statistical techniques" IEEE Conference on ETFA, Septiembre 2009.
23. M. Orchard, A. M. Cipriano, A. Z. Cipriano, M. Viale†, A. Vigliocco, "A Model Based Fault Detection and Diagnosis System for Rolling Mill Equipment", Department of Electrical Engineering, Pontificia Universidad Católica de Chile.
24. M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, D. Teneketzis, "Failure diagnosis using discrete event models", IEEE Conference on Decision and Control, vol. 3, pp. 3110-3116, Diciembre 1994.
25. N. Kabbaj, M. Ramzi, B. Dahhou, H. Youlal, G. Enea, "Fault detection and isolation in a greenhouse using parity relations" IEEE Conference ETFA, vol. 2, pp. 747-752, Septiembre 2003.
26. P. Eykhoff, System Identificacion, John Wiley, London, 1974.
27. P. Medina, "Detección y aislamiento de fallas en cámaras de combustión de la turbina a gas de la central Nueva Renca", Abril 2007.
28. Q. He, J. Wang, "Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes", IEEE T. Semiconductor Manufacturing, vol. 20, no. 4, pp. 345-354, Noviembre 2007.
29. R. Clark, D. Fosth, V. Walton, "Detecting Instrument Malfunctions in Control Systems" IEEE T. Aerospace and Electronic Systems, vol. AES-11, no. 4, pp. 465-473, Julio 1975.

30. R. Duda, P. Hart, D. Stork, "Pattern Classification", New York: John Wiley & Sons, 2001.
31. R. Haber, H. Unbehauen, Structure identification of nonlinear dynamic systems – a survey on input/output approaches. *Automatica*, 26(4), pp. 651-677, 1990.
32. R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*, Springer, 2006, pp. 5-19.
33. R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*, Springer, 2006, pp. 61-284.
34. R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*, Springer, 2006, pp. 286-343.
35. R. Isermann, "On fuzzy logic applications for automatic control, supervision, and fault diagnosis", *IEEE T. Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 28, no. 2, pp. 221-235, Marzo 1998.
36. R. Patton, P. Frank, R. Clarke, *Fault diagnosis in dynamic systems: theory and application*, Prentice-Hall, 2000.
37. S. El-Shal, A. Morris, "A Fuzzy Expert System for Fault Detection in Statistical Process Control of Industrial Processes", *IEEE T. Systems, Man and Cybernetics, Part C: Applications And Reviews*, vol. 30, no. 2, pp. 281-289, Mayo 2000.
38. S. Huang, K. Tan, "Fault Detection and Diagnosis Based on Modeling and Estimation Methods", *IEEE T. Neural Networks*, vol. 20, no. 5, pp. 872-881, Mayo 2009.
39. S. Ibaraki, S. Suryanarayanan, M. Tomizuka, "Design of Luenberger state observers using fixed-structure H_∞ optimization and its application to fault detection in lane-keeping control of automated vehicles", *IEEE/ASME T. Mechatronics*, vol. 10, no. 1, pp. 34-42, Febrero 2005.
40. S. Wegerich, "Similarity based modeling of time synchronous averaged vibration signals for machinery health monitoring", *IEEE Aerospace Conference*, vol. 6, pp. 3654- 3662, Marzo 2004.
41. V. Chandola, A. Banerjee, V. Kumar, "Anomaly Detection: A Survey", Department of Computer Science and Engineering University of Minnesota, Agosto 2009.
42. V. Todorovic, P. Tadic, Z. Djurovic, "Expert system for fault detection and isolation of coal-shortage in thermal power plants" *Conference on Control and Fault-Tolerant Systems*, pp. 666-671, Octubre 2010.
43. V. Tseng, C. Kao, "A Novel Similarity-Based Fuzzy Clustering Algorithm by Integrating PCM and Mountain Method", *IEEE T. Fuzzy Systems*, vol. 15, no. 6, pp. 1188-1196, Diciembre 2007.
44. W. Lee, D. Grosh, F. Tillman, C. Lie, "Fault Tree Analysis, Methods, and Applications - A Review", *IEEE T. Reliability*, vol. R-34, no. 3, pp. 194-203, Agosto 1985.

45. X. Hong, Y Yuan, X. Hu, "Prediction of Moving Objects' K-Nearest Neighbor Based on Fuzzy-Rough Sets Theory", International Conference on Fuzzy Systems and Knowledge Discovery, vol. 1, pp. 407-411, Agosto 2007.
46. X. Shi, J. Liang, L. Ye, B. Hu, "A method of fault diagnosis based on PCA and Bayes classification", World Congress on Intelligent Control and Automation (WCICA), pp. 5628-5631, Julio 2010.
47. X. Zhang, W. Yan, X. Zhao, H. Shao, "Nonlinear On-line Process Monitoring and Fault Detection Based on Kernel ICA", ICIA, pp. 222-227, Diciembre 2006.
48. Y. Wang, M. Zhang, Y. Guo, "Fault detection and data restoration based on PCA for sensors of autonomous underwater vehicle", International Conference on Mechatronics and Automation, pp. 4801-4805, Agosto 2009.
49. Y. Zhang, Y. Chen, Q. Meng, T. Wang, "Fault Detection of Networked Control Systems Based on Kalman Filter", International Conference on Innovative Computing Information and Control, pp. 348, Junio 2008.
50. Z. Niu, J. Liu, Y. Niu, Y. Pan, "A reformative PCA-based fault detection method suitable for power plant process", International Conference on Machine Learning and Cybernetics, vol. 4, pp. 2133-2138, Agosto 2005.
51. Endesa Chile. Central térmica generadora de ciclo combinado, San Isidro.
< http://www.endesa.cl/Endesa_Chile/action.asp?id=11128 > [consulta: 02 Mayo 2012].
52. Endesa Educa. Centrales térmicas de ciclo combinado.
< http://www.endesaeduca.com/Endesa_educa/recursos-interactivos/produccion-de-electricidad/ix.-las-centrales-termicas-de-ciclo-combinado > [consulta: 02 Mayo 2012].
53. Wikipedia. Ejemplo del algoritmo de agrupación de datos por k-medias.
< http://en.wikipedia.org/wiki/K-means_clustering > [consulta: 02 Mayo 2012].
54. D. Cornejo, R. Marambio, Pontificia Universidad Católica de Chile. Escuela de Ingeniería. Departamento de Ingeniería Eléctrica. Esquema de evaluación de opciones reales en mercados eléctricos
< http://web.ing.puc.cl/~power/alumno06/RealOptions/cent_term.html > [consulta: 02 Mayo 2012].