

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tesisenxarxa.net) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tesisenred.net) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tesisenxarxa.net) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author



UNIVERSIDAD POLITECNICA DE CATALUÑA

DEPARTAMENTO DE INGENIERÍA DE SISTEMAS, AUTOMÁTICA E INFORMÁTICA INDUSTRIAL
PROGRAMA DE DOCTORADO EN AUTOMATICA, ROBOTICA Y VISION

**Detección y Diagnóstico de Fallos Múltiples
en Sistemas Dinámicos
Usando Análisis de Componentes Principales No
Lineal y Residuos Estructurados.**

Tesis presentada para obtener el grado de Doctor
por la Universidad Politécnica de Cataluña

Autor: Amílcar Alejandro Rincón Charris
Director: Dr. Joseba - Jokín Quevedo Casin

Barcelona 2015

AGRADECIMIENTOS

En primer lugar quiero darle las gracias a una persona muy especial en mi vida, Nidza I. Torres Colon, mi esposa. Por su apoyo incondicional y abnegada paciencia durante estos años. Sin sus consejos y aliento esta tesis no habría podido ser llevada a cabo. En segundo lugar a mis hijos Alejandro y Karina, mis amores y alegría a quienes dejé de dedicarles mi tiempo para poder culminar este proyecto. Agradezco a mis padres Ángela y Amílcar; y a mis hermanos Jabid, Ramón y en especial a mi hermana Edilia por sus palabras de apoyo constante.

Quiero agradecer a mi tutor de tesis Dr. Joseba Quevedo Casin por su apoyo incondicional y sobre todo paciencia durante el desarrollo de este trabajo. Al igual agradecer al Dr. Vicenc Puig y a la Dra. Teresa Escobet. Su guía fue fundamental para darle buen término a esta tesis.

Agradezco de manera muy especial a la Universidad Interamericana de Puerto Rico Recinto de Bayamón por el soporte económico ofrecido para poder realizar y culminar mi estudio doctoral. A mis amigos y compañeros de trabajo: Juan Martínez, Carlos Olivares, Javier Quintana, Rafael Salgado, Rafael Canales, Harry Rivera, Bert Rivera, Omar Meza, Eduardo Pérez, Eduardo Lay, Otoniel Díaz y demás, por su continuo apoyo y algunos por revisar y corregir detalladamente este documento, entregándome comentarios valiosos que pude incorporar.

Finalmente y no menos importante, agradezco también a *Dios* porque sin su voluntad nada hubiese sido posible.

RESUMEN

En los últimos años la detección y el diagnóstico de fallos en dispositivos y en procesos ha sido un campo de investigación y desarrollo que se ha podido abordar desde múltiples perspectivas: la ingeniería de control, la inteligencia artificial y la estadística entre otros. Dentro de cada una de ellas con multitud de técnicas: sistemas expertos, redes neuronales, razonamiento basado en casos, análisis de señales, observadores, ecuaciones de redundancia analítica, diagnóstico basado en consistencia, pero considerando la presencia de fallos únicos y no múltiples. En esta tesis se estudiará la detección y el diagnóstico de fallos múltiples en sistemas dinámicos utilizando análisis de componentes no lineal para la detección y los residuos estructurados para el diagnóstico. El estudio incluye el desarrollo de un modelo matemático de una mini turbina de gas SR-30 y en la implementación de algoritmos de detección y diagnóstico de fallos múltiples en dicho sistema.

La detección y el diagnóstico de fallos múltiples ha resultado ser más complejo que la detección y el diagnóstico de fallos únicos. Esto no es sólo porque aumenta el número de fallos, sino también debido a la aparición de nuevos fenómenos que deben ser considerados, tales como una combinación o la influencia mutua de los fallos, la compensación, y la explosión combinatoria de posibles escenarios de fallos. El problema de fallos múltiples es importante, ya que el supuesto de un solo fallo puede llevar a diagnósticos incorrectos cuando fallos múltiples ocurren.

ABSTRACT

In recent years the detection and diagnosis of faults in devices and processes has been a field of research and development that has been addressed from multiple perspectives: control engineering, artificial intelligence and statistics among others. Within each with a number of techniques: expert systems , neural networks, case-based reasoning , signal analysis, observers equations analytical redundancy based on consistency, but only considering the presence of multiple failures and no diagnosis. In this thesis the detection and diagnosis of multiple faults in dynamic systems using nonlinear analysis for the detection and diagnosis for structured waste components will be studied. The study includes the development of a mathematical model of a gas turbine mini SR -30 and implementation of algorithms for detection and diagnosis of multiple faults in the system.

The detection and diagnosis of multiple faults is more complex than the detection and diagnosis of single faults. This is not only because it increases the number of failures, but also because of the emergence of new phenomena that should be considered, such as a combination or interaction of faults, compensation, and the combinatorial explosion of possible failure scenarios. The problem of multiple failures is important, since the single fault assumption can lead to incorrect diagnoses when multiple faults occur.

TABLA DE CONTENIDO

AGRADECIMIENTOS	1
RESUMEN	2
ABSTRACT.....	3
LISTA DE FIGURAS.....	8
LISTA DE TABLAS	12
ABREVIATURAS Y SIMBOLOS.....	13
CAPITULO 1:.....	14
INTRODUCCION	14
1.1. Introducción	15
1.2. Motivación	17
1.3. Objetivos	18
1.4. Contribución de la tesis.....	18
1.5. Estructura de la tesis	20
1.6. Bibliografía	21
CAPITULO 2:.....	24
ESTADO DEL ARTE.....	24
2.1. Introducción.	25
2.2. Estudios Basados en Modelos Cuantitativos.....	28
2.3. Estudios Basados en Modelos Cualitativos.	29
2.4. Estudios Basados en el Análisis de Señales	33
2.5. Diagnóstico de Fallos Intermitentes.....	35
2.6. Detección y Diagnostico de Fallos Múltiples en Turbinas de Gas	36
2.6. Sumario	37
2.7. Bibliografía	38
CAPITULO 3:.....	43

DESCRIPCION DEL SISTEMA.....	43
TURBINA DE GAS SR-30	43
3.1. Descripción del Sistema "MiniLab Gas Turbine"	44
3.2. Principio de funcionamiento	49
3.3. Modelo Matemático de una Turbina de Gas SR-30.....	51
3.3.1. Desarrollo de las ecuaciones del modelo.	52
3.3.2. Relaciones Fundamentales.	54
3.4. Modelo en Simulink de la Turbina de Gas SR-30	60
3.4.1. Validación del Modelo.....	63
3.5. Fallos Estudiados en la Turbina de Gas	66
3.6. Trabajo experimental	68
3.6. Sumario	69
3.7. Bibliografía	69
CAPITULO 4:.....	71
DETECCION DE FALLOS MULTIPLES EN UNA TURBINA DE GAS USANDO ANALISIS DE COMPONENTES PRINCIPALES (PCA)	71
4.1. Introducción	72
4.2. Fundamentos de PCA	73
4.2.1. Cálculo de los componentes.....	75
4.2.2. Propiedades de la matriz de covarianza Γ	79
4.2.3. Propiedades de las componentes del espacio ortogonal.....	79
4.2.4. Dimensión del espacio de componentes principales.....	80
4.3. Modelado implícito	80
4.3.1. Estandarización de datos	80
4.3.2. Subespacios de componentes principales y subespacios residual	81
4.4. Detección de fallos múltiples basados en PCA.....	84
4.4.1. Umbrales de los residuos	84
4.4.1.1. Umbral de los residuos Hotelling (T^2)	84
4.4.1.2. Umbral de los residuos SPE.....	86

4.5. Algoritmo para la detección de fallos múltiples mediante PCA.....	87
4.5.1. Calculo del PCA (fuera de línea).....	87
4.5.2. Calculo del PCA (En línea).....	87
4.6. Resultados:.....	89
4.7. Conclusiones.....	99
4.8. Bibliografía.....	101
CAPITULO 5:.....	102
DETECCION DE FALLOS MULTIPLES EN TURBINAS DE GAS USANDO ANALISIS DE COMPONENTES PRINCIPALES NO LINEAL (NLPCA).....	102
5.1. Introducción.....	103
5.2. Estado del arte del NLPCA.....	103
5.3. Análisis de Componentes Principales No Lineal (NLPCA).	104
5.4. Algoritmo de la Curva Principal.....	108
5.5. Mapeo de la Red.....	109
5.6. Demapeo de la Red.....	111
5.7. Detección de fallos múltiples usando NLPCA.....	112
5.7.1. Detección y aislamiento de fallos múltiples.....	113
5.8. Resultados:.....	115
5.11. Conclusiones.....	120
5.12. Bibliografía.....	122
CAPITULO 6:.....	123
DIAGNOSTICO DE FALLOS MULTIPLES USANDO RESIDUOS ESTRUCTURADOS.....	123
6.1. Introducción.....	124
6.2. Diagnóstico de Fallos.....	124
6.2. Redundancia analítica.....	125
6.3. Generación de residuos.....	126
6.3.1. Generación de Residuos mediante estimación de parámetros.....	127
Método de ecuación de error.....	128
Método de error de salida.....	129

6.3.2. Ecuaciones de paridad.....	130
6.3.3. Observadores.....	134
6.4. Evaluación de residuos.....	136
6.5. Localización de fallos mediante residuos estructurados	138
6.6. Metodología de Diagnóstico propuesta.....	140
6.6.1. Análisis de sensibilidad.....	140
6.6.2. Modelización del efecto de los fallos.....	142
6.6.3. Monitorización.....	143
6.7. Aplicación a fallos múltiples de una turbina de gas.....	143
6.7.1. Descripción de los fallos a diagnosticar.....	143
6.7.2. Matriz de sensibilidad	144
6.8. Resultados	150
6.8.1. Fallos Simples.....	150
6.8.2. Fallos Dobles.	152
6.8.3. Fallos Triples	155
6.9. Conclusiones	157
6.10. Bibliografía	158
CAPITULO 7:.....	159
CONCLUSIONES Y TRABAJOS FUTUROS	159
7.1. Introducción.	160
7.2. Conclusiones	160
7.3. Trabajos Futuros.	163
8. BIBLIOGRAFIA	164
APENDICES:	174

LISTA DE FIGURAS

Figura 1. 1. Fallos secuenciales	16
Figura 1. 2. Fallos simultáneos	16
Figura 2. 1. Esquemático de clasificación de los diferentes métodos de detección y diagnóstico.....	25
Figura 2. 2. Modelo basado en PCA Y CA_CMAC	34
Figura 3. 1. Sistema "Minilab" de la compañía Turbine Technologies Inc.	44
Figura 3. 2. Vista seccionada de la turbina SR-30	45
Figura 3. 3. Válvula manual de control de combustible	46
Figura 3. 4. Sistema de adquisición Compact Rio instalado.....	48
Figura 3. 5. Control automático de la válvula de combustible mediante un servomotor DC	49
Figura 3. 6. Diagrama esquemático del ciclo de Brayton	50
Figura 3. 7. Diagrama de la turbina de gas	50
Figura 3. 8. RPM versus tiempo	56
Figura 3. 9. Relación de Presiones en el compresor a diferentes RPM.....	57
Figura 3. 10. Eficiencia en el compresor a diferentes RPM.....	57
Figura 3. 11. Relación de Presiones en la Turbina a diferentes RPM.....	58
Figura 3. 12. Regresión múltiple entre la masa de aire, la relación de presión en el compresor y rpm.	59
Figura 3. 13. Regresión múltiple entre la masa de aire, la relación de presión del compresor y rpm... 59	59
Figura 3. 14. Regresión múltiple entre la masa de combustión, relación de presión en la turbina y rpm.	60
Figura 3. 15. Regresión múltiple entre la masa de combustible, la velocidad del rotor y la eficiencia de la turbina	60
Figura 3. 16. Modelo en Simulink para la turbina de gas SR30	61
Figura 3. 17. Diagrama en Simulink del compresor	61
Figura 3. 18. Diagrama en Simulink de la turbina	62
Figura 3. 19. Diagrama en Simulink de la cámara de combustión.....	62
Figura 3. 20. Diagrama en Simulink para obtener la revoluciones en el eje.....	63
Figura 3. 21. Diagrama en Simulink para obtener la potencia del compresor.	63
Figura 3. 22. Diagrama en Simulink para obtener la potencia en la turbina.....	63
Figura 3. 23. Flujo de combustible en su comportamiento real versus simulado	64
Figura 3. 24. Velocidad del eje de la turbina en su comportamiento real versus simulado	64
Figura 3. 25. Temperaturas en su comportamiento real versus simulado	65

Figura 3. 26. Presiones en su comportamiento real versus simulado.....	65
Figura 3. 27. Representación de fallos abruptos	67
Figura 3. 28. Representación de fallos graduales.....	68
Figura 3. 29. Muestra extraída de un alabe de la turbina para observar el comportamiento de la eficiencia mecánica.	68
Figura 3. 30. Desbalanceo del rotor de la turbina SR30	69
Figura 4. 1. Recta que minimiza las distancias ortogonales de los puntos	74
Figura 4. 2. Elipse que envuelve las distancias ortogonales de los puntos	75
Figura 4. 3. Esquema de detección de fallos múltiples mediante PCA.....	88
Figura 4. 4. Turbina de gas SR30 a 55000 rpm.	89
Figura 4. 5. Turbina de gas SR30 a 65000 rpm.	89
Figura 4. 6. Turbina de gas SR30 a 70000 rpm.	90
Figura 4. 7. Temperaturas en la turbina	91
Figura 4. 8. Presiones en la turbina.....	91
Figura 4. 9. Par en la turbina.....	91
Figura 4. 10. Componentes principales para 65000 rpm	93
Figura 4. 11. Criterio Hotelling para condiciones sin fallo a 65000 rpm.....	94
Figura 4. 12. Criterio de Error de Predicción al cuadrado (SPE) para condiciones sin fallo a 65000 rpm.....	94
Figura 4. 13. Criterio de Hotelling para fallo de 10% en la eficiencia del compresor.....	95
Figura 4. 14. Criterio de SPE para fallo de 10% en la eficiencia del compresor.	95
Figura 4. 15. Criterio de Hotelling T2 para fallos secuenciales.....	95
Figura 4. 16. Error de predicción al cuadrado para dos fallos secuenciales 5%WC y 5%ET a 55000 rpm.....	96
Figura 4. 17. Hotelling T2 aplicado al sistema con fallos simultáneos 5EC 10ET a 65000 rpm.....	96
Figura 4. 18. SPE aplicado al sistema con fallos simultáneos 5EC 10ET a 65000 rpm.	97
Figura 4. 19. Hotelling para fallos simultáneos de 2% EC y 3%ET.....	97
Figura 4. 20. SPE fallos simultáneos 2%EC y 3%ET.....	97
Figura 4. 21. Hotelling fallos secuenciales 5%WC y 5%ET a 55000 rpm.	98
Figura 4. 22. SPE fallos secuenciales 5%WC y 5%ET.....	98
Figura 4. 23. Hotelling T2 fallos incipientes simultáneos 5EC 5ET 2000s	98
Figura 4. 24. SPE fallos incipientes simultáneos 5EC 5ET 2000s.....	99
Figura 5. 1. El PCA minimiza la suma de los cuadrados de la desviación ortogonal usando una línea recta.....	105

Figura 5. 2. El NLPCA minimiza la suma de los cuadrados de la desviación ortogonal utilizando una curva.....	105
Figura 5. 3. PCA versus NLPCA para los mismos datos.....	105
Figura 5. 4. Reducción de dimensionalidad de los datos	106
Figura 5. 5. Uso de una red neuronal para la extracción de los componentes principales no lineales	107
Figura 5. 6. Proyección de los puntos a la curva.....	109
Figura 5. 7. Red neuronal para el mapeo	110
Figura 5. 8. Red neuronal para demapeo	110
Figura 5. 9. Residuos usando PCA	114
Figura 5. 10. Residuos usando NLPCA	114
Figura 5. 11. Comparación de los residuos usando PCA versus NLPCA.....	114
Figura 5. 12. NLPCA aplicando la curva principal a 55000 rpm.	115
Figura 5. 13. NLPCA aplicando la curva principal a 65000 rpm.	116
Figura 5. 14. SPE para un fallo de 10% ET	116
Figura 5. 15. SPE para fallos múltiples simultaneo de 10%ET y 8%EC.....	117
Figura 5. 16. SPE para fallos múltiples secuenciales de 10%ET y 8%EC	117
Figura 5. 17. SPE para fallos múltiples secuenciales de 10%ET y 8%EC	118
Figura 5. 18. SPE para fallos múltiples secuenciales de 10%ET, 8%EC y 5%WT	118
Figura 5. 19. Cambio en T02 cuando existen 3 fallos secuenciales.....	119
Figura 5. 20. Variación de las rpm cuando existen tres fallos secuenciales	119
Figura 5. 21. Validación de la red neuronal de mapeo.....	120
Figura 6. 1. Esquema básico de detección de fallos usando redundancia analítica	126
Figura 6. 2. Estimación de parámetros ecuación del error	129
Figura 6. 3. Estimación de Parámetros. Error de salida.	129
Figura 6. 4. Grafica de control	137
Figura 6. 5. Conjunto de residuos estructurados a) Dedicado b) Generalizado	139
Figura 6. 6. Fallo de 5% en la eficiencia del compresor	151
Figura 6. 7. Fallo de 3% en la eficiencia de la turbina.....	151
Figura 6. 8. Fallo de 5% en WF	152
Figura 6. 9. Fallo de 7% de PRT.....	152
Figura 6. 10. Dos Fallos simultáneos del 2% WC y 3% WT.....	153
Figura 6. 11. Dos Fallos simultáneos del 5% WC y 6% EC	153
Figura 6. 12. Dos Fallos simultáneos 5% WC y 8% ET	154
Figura 6. 13. Dos Fallos simultáneos 7% WT y 5% EC	154
Figura 6. 14. Dos Fallos secuenciales 5%EC 8%EC y 5%ET 8%ET.....	155

Figura 6. 15. Tres Fallos simultáneos 5%EC, 7%ET y 9%WF	155
Figura 6. 16. Tres Fallos simultáneos 5%EC, 7%WT y 3%ET	156
Figura 6. 17. Tres Fallos simultáneos 5%EC, 8%ET y 2WF.....	156
Figura 6. 18. Tres Fallos secuenciales 5%EC 8%EC, 5%ET 8%ET y 3%WT 5%WT	157

LISTA DE TABLAS

Tabla 3. 1 Especificaciones de los sensores instalados en la turbina de gas SR-30.....	45
Tabla 3. 2. Propiedades para combustible Jet	47
Tabla 3. 3. Condiciones de operación de la turbina SR 30	47
Tabla 3. 4. Módulos Compact Rio instalados para la conexión de sensores.....	48
Tabla 3. 5. Variables de estado del modelo de la turbina	53
Tabla 3. 6. Constantes calculadas para los parámetros de la turbina	58
Tabla 3. 7. Fallos estudiados en los sensores de la turbina de gas	66
Tabla 4. 1. Variables medidas en la turbina de gas SR30	90
Tabla 4. 2. Fallos analizados	99
Tabla 4. 3. Fallos múltiples analizados	99
Tabla 6. 1. Descripción de los fallos en componentes de la turbina	143
Tabla 6. 2. Variables medidas en la turbina	144
Tabla 6. 3. Matriz de Residuos para 55000 rpm	147
Tabla 6. 4. Matriz de Residuos para 60000 rpm	148
Tabla 6. 5. Matriz de Residuos para 65000 rpm	148
Tabla 6. 6. Matriz de Sensibilidad para 55000 rpm	149
Tabla 6. 7. Matriz de Sensibilidad para 60000 rpm	149
Tabla 6. 8. Matriz de Sensibilidad para 65000 rpm	150

ABREVIATURAS Y SIMBOLOS

Tw	Tiempo de espera
a(k)	Valor de activación
PCA	Análisis de componentes principales
P01	Presión estática de entrada del compresor
P02	Presión de estancamiento de salida del compresor
P03	Presión de la cámara de combustión
P04	Presión de estancamiento de la salida de la turbina
P05	Presión de estancamiento de la salida de la tobera
T01	Temperatura estática de entrada del compresor
T02	Temperatura de estancamiento de salida del compresor
T03	Temperatura de estancamiento de la entrada de la turbina
T04	Temperatura de estancamiento de la salida de la turbina
T05	Temperatura de estancamiento de la salida de la tobera
Fw	Flujo de combustible
Tn	Torque de la turbina
N	Velocidad de rotación del eje de la turbina en revoluciones por minuto (RPM)
σ_I	Coefficiente de pérdida de presión en el ducto de entrada
σ_{comb}	Coefficiente de pérdida de presión de combustión
η_{comb}	Eficiencia de la combustión
\dot{m}	Flujo másico
U	Energía interna
C_v	Capacidad calorífica a volumen constante
C_p	Capacidad calorífica a presión constante
h	Entalpia
Q	Calor
W	Trabajo
η_{mech}	Eficiencia mecánica
LHV	Poder calorífico del combustible
F	Fallo
MC	Masa de aire en el compresor
EC	Eficiencia del compresor
MT	Obstrucción de la masa de aire y combustible en la turbina
ET	Eficiencia en la turbina
WF	Obstrucción en la válvula de combustible

CAPITULO 1:

INTRODUCCION

Contenido

- 1.1. Introducción
- 1.2. Motivación
- 1.3. Objetivos
- 1.4. Contribución de la tesis
- 1.5. Estructura de la tesis
- 1.6. Bibliografía

1.1. Introducción

Es una realidad que los sistemas modernos se vuelven cada vez más complejos y que los algoritmos de control que se implementan son cada vez más sofisticados. De tal manera que las características de fiabilidad, disponibilidad y seguridad adquieren cada vez mayor importancia. Dichas características son importantes no solo en aquellos sistemas cuya seguridad es crítica, tales como plantas nucleares, plantas químicas y aeronaves, sino en cualquier tipo de proceso de fabricación automatizado [Nie-10], [Kos-09a], [Cas-04].

Para los sistemas en los que la seguridad es crítica, las consecuencias de los fallos pueden ser extremadamente serias en términos de vidas humanas, impacto medioambiental y pérdidas económicas; por lo que existe una necesidad creciente en la supervisión en línea y en el diagnóstico de fallos con el objetivo de incrementar la fiabilidad, teniendo en cuenta que los síntomas que presentan fallos que se están desarrollando pueden ayudar a evitar fallos irreversibles como caídas del sistema y catástrofes. Para aquellos sistemas donde la seguridad no es crítica, las técnicas de diagnóstico de fallos en línea se pueden utilizar para mejorar la eficiencia, mantenibilidad, disponibilidad y fiabilidad de la planta. Los métodos modernos de diagnóstico de fallos pueden aportar información del estado del sistema que permiten implementar un mejor plan de mantenimiento [Gar-11].

Numerosos trabajos están dedicados al análisis de los métodos de detección y diagnóstico de fallos únicos y muy pocos con enfoque hacia los fallos múltiples, sean simultáneos o secuenciales. Algunos de los más populares incluye el diagnóstico extendido de matrices [Kos-04a], [Kos-04b], modelos basados en teoría de conjuntos [Reg-83], [Reg-85], consistencia basada en razonamiento [Rei-87], [Ham-92], [Lig-04], gráficos de lógica causal [Fus-96], [Lig-97], [Lig-04], y muchos otros [Tza-89a], [Tza-89b], [Dav-92], [Kor-04]. En el área de detección y aislamiento de fallos, emergente de la teoría clásica de control automático, estos problemas son discutidos en [Fra-90], [Fra-96] y [Kor-04].

A continuación se presentan algunas definiciones pertinentes a esta tesis:

Fallo: Desviación no permitida de, al menos, una propiedad característica o parámetro de un sistema de su condición estable, usual o estándar. Un fallo es la aparición de un modo de fallo. [Bla-00].

Fallo abrupto: fallo cuyo efecto aparece repentinamente (por ejemplo, modelado mediante un escalón).

Fallo incipiente: fallo cuyo efecto aparece progresivamente (Por ejemplo modelado mediante rampa).

Detección de fallos: determinación de la presencia de fallos en el sistema así como el instante de su aparición.

Diagnóstico de fallos: determinación del tipo, tamaño, localización e instante de aparición de un fallo. Incluye la detección, el aislamiento y la estimación del fallo.

Aislamiento de fallos: determinación del tipo, localización e instante de detección de un fallo. Se realiza después de la etapa de detección.

Fallos secuenciales: dos o más fallos son secuenciales cuando ocurren en instantes de tiempos diferentes y entre fallos no se soluciona el problema o si los fallos ocurren en intervalos de tiempo más largos que el necesario para formular el diagnóstico [Kos-09a]. Ver figura 1.1.

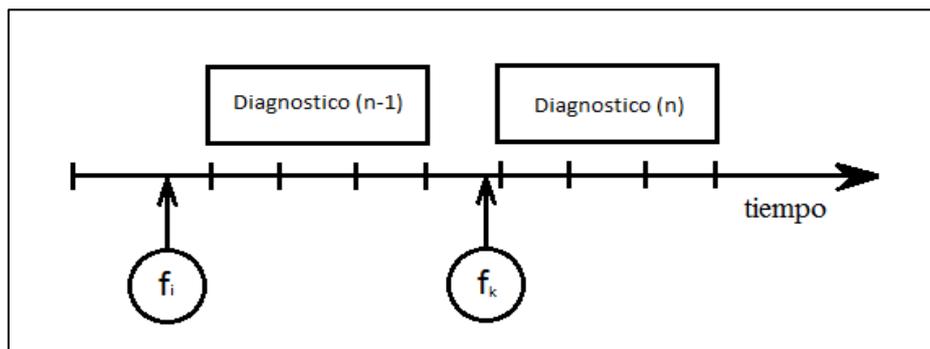


Figura 1. 1. Fallos secuenciales

Fallos simultáneos: dos o más fallos se consideran simultáneos cuando ocurren en el mismo instante de tiempo o si los fallos ocurren en intervalos de tiempo más cortos que el necesario para formular el diagnóstico [Kos-09a]. Ver figura 1.2.

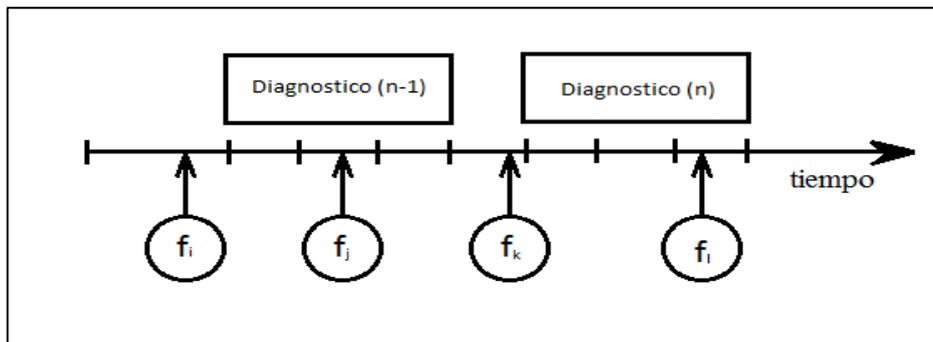


Figura 1. 2. Fallos simultáneos

1.2. Motivación

El problema de detección y diagnóstico de fallos en sistemas dinámicos generalmente se enfoca en los fallos únicos, pero muy pocas veces o casi nunca se consideran los fallos múltiples, ni simultáneos, ni secuenciales [Kos-09a]. Razón por la cual es un área de investigación en la cual existen varias interrogantes que resolver. Existen sistemas donde diversos componentes se alimentan de la misma energía o fuente y si ésta falla, todos o parte de los componentes pueden fallar en el sistema, siendo este un fallo múltiple y muy común en los sistemas actuales por su naturaleza modular.

Otra razón por la cual es importante estudiar los fallos múltiples; es que hay sistemas que cuando se quieren analizar por primera vez, y desde hace tiempo se han producido simultáneamente o no, fallos que permanecen latentes, cuando esto ocurre, no se puede aplicar métodos que consideren fallos únicos porque existen o pueden existir múltiples fallos latentes que se deseen diagnosticar.

Adicionalmente, si se diagnostica un sistema cada cierto tiempo (una vez al día), puede ocurrir que se hayan producido en ese lapso de tiempo un desencadenamiento de fallos no simultáneos, secuencialmente, pero cuando se quiera analizar el sistema haya más de un fallo. Aquí tampoco sirven los métodos tradicionales de diagnosticar fallos únicos.

Para su fase de diseño y prueba las compañías fabricantes de turbinas de avión muestran un gran interés en buscar técnicas que le ayuden a mejorar la fiabilidad de sus productos. El desarrollo de algoritmos que le permitan determinar la ubicación de sensores para conseguir un determinado grado de diagnosticabilidad y capacidad de reposición y tolerancia a fallos múltiples sería un gran reto para el estudio en mención [Ben-00], [Gas-02].

El tema de diagnóstico de fallos múltiples es un tema de mucha dificultad como lo han podido expresar y analizar estudios recientes [Kos-04b], [Kos-04a], [Rei-87], [Ham-92], [Lig-04], [Fus-96], [Lig-97], [Lig-04]. Esto no es sólo porque aumenta el número de fallos, sino también debido a la aparición de nuevos fenómenos que deben ser considerados, tales como una combinación o la influencia mutua de los fallos, la compensación, y la explosión combinatoria de posibles escenarios de fallos ya que estos podrían ser de diferente magnitud. Los fallos múltiples en sistemas dinámicos son difíciles de detectar porque ellos pueden enmascarar o compensar otros efectos. El problema de los fallos múltiples es importante debido a que la presunción de un fallo único puede ser incorrecta o presentarse diagnósticos erróneos cuando fallos múltiples ocurren [Lig-08]. El diagnóstico de fallos múltiples es un problema de gran dificultad en sistemas dinámicos porque el efecto de los fallos puede alterar el resultado de la firma de fallos [Dai-07].

1.3. Objetivos

El objetivo general de esta tesis es:

- Desarrollar una metodología de detección y diagnóstico de fallos múltiples en sistemas dinámicos complejos. Para la detección de fallos múltiples en componentes y sensores se utilizarán las técnicas estadísticas del análisis de componentes principales lineal y no lineal; adicional para el diagnóstico se exploran las técnicas de residuos estructurados para el caso de fallos en componentes y sensores.

Los objetivos específicos para esta tesis son:

- Desarrollar un modelo matemático no lineal y discreto de una turbina de gas SR-30, para probar las técnicas propuestas de detección y diagnóstico de fallos múltiples en un sistema dinámico complejo.
- Plantear una metodología de detección de fallos únicos o múltiples en componentes o sensores para un sistema dinámico complejo usando análisis de componentes principales lineal (PCA).
- Plantear una metodología de detección de fallos únicos o múltiples en componentes o sensores para un sistema dinámico complejo usando análisis de componentes principales no lineal (NLPCA).
- Plantear una metodología de diagnóstico de fallos únicos o múltiples en el proceso para un sistema dinámico complejo usando residuos estructurados.
- Probar de manera teórica y experimental la implementación de la metodología de detección y diagnóstico de fallos múltiples, mediante el uso de Simulink, para el modelo de la turbina de gas SR-30. La turbina real fue usada para calibrar el modelo en Simulink y para probar fallos en sensores. Para el caso de fallos en componentes no se usó el sistema real debido a que era extremadamente peligroso y no se disponía de las medidas de seguridad adecuadas.

1.4. Contribución de la tesis

Las contribuciones de esta tesis pueden resumirse en los siguientes puntos principales:

- Recopilación de problemáticas y estado del arte de sistemas de detección y diagnóstico de fallos múltiples, que puede servir como referencia y punto de partida para futuras investigaciones en esta línea.
- Un nuevo esquema para un sistema de detección y diagnóstico de fallas múltiples que combina modelos, el análisis estadístico basado en el análisis de componentes principales no

lineal y los residuos estructurados [Gar-11], [Kem-06], [Mor-07]. A diferencia de la gran mayoría de estudios referentes a la detección y diagnóstico de fallos múltiples hallados en la literatura actual, la metodología de detección de fallos basados en análisis de componentes principales no lineal permite obtener resultados de detección y diagnóstico. En adición, no se necesita de un modelo del sistema. Este desarrollo realiza detección y diagnóstico de una forma práctica, sencilla y permite ser implementado sin mayores supuestos sobre el proceso en estudio.

- Una extensión del método tradicional de diagnóstico basado en modelos usando residuos estructurados, pero aplicado a fallos múltiples. Se propone la utilización de la información proporcionada por la sensibilidad del residuo frente al fallo para el diseño del diagnosticador según la teoría de Gertler [Ger-98], [Ger-10]. Esta sensibilidad proporciona una información cuantitativa del efecto del fallo sobre el residuo y una información cualitativa sobre su sentido de variación (signo). La utilización de esta información en la etapa de diagnóstico permitirá separar fallos que aun presentando la misma firma de fallos teórica binaria, presentan sensibilidades cualitativamente o cuantitativamente diferentes. Esta metodología permite diagnosticar y conocer la magnitud aproximada de los fallos detectados de una manera sencilla y puede ser implementado de manera fácil sobre el proceso en estudio.

1.5. Estructura de la tesis

Esta tesis se estructura de 8 capítulos:

En el capítulo 2 se presenta la revisión de literatura, en la cual se detalla de manera general las metodologías y técnicas usadas actualmente para la detección y diagnóstico de fallos múltiples en tres grandes grupos: los basados en modelos cuantitativos, modelos cualitativos y basados en señales. Al final del capítulo se detallan los últimos estudios en detección y diagnóstico de fallos en turbinas de gas.

En el capítulo 3 se muestra la definición del equipo experimental, la modelación matemática de una turbina de gas SR-30, las ecuaciones que la caracterizan, el modelo desarrollado en Simulink y los fallos modelados para la verificación de la teoría propuesta en esta tesis.

El capítulo 4 se plantea la propuesta del método del análisis de componentes principales (PCA) para la detección de fallos múltiples, propiedades de la matriz de covarianza, dimensión del espacio de componentes principales, modelado implícito, umbral de residuos usando el criterio de Hotelling y el espacio de residuales. En la parte final de este capítulo se muestran los resultados obtenidos al aplicar la metodología PCA para fallos múltiples.

En el capítulo 5 se plantea la propuesta del método del análisis de componentes principales no lineal (NLPCA) para la detección de fallos múltiples, umbral de residuos usando el criterio de Hotelling y el espacio de residuales. Finalmente se compara el método del análisis de componentes principales (PCA) y el método del análisis de componentes principales no lineales (NLPCA) para la detección y el diagnóstico de fallos múltiples en sensores y componentes. En la parte final de este capítulo se muestran los resultados obtenidos al aplicar la metodología NLPCA para fallos múltiples.

En el capítulo 6 se presenta el diseño de residuos estructurados para el diagnóstico de fallos múltiples en componentes. Análisis de sensibilidad, modelización del efecto de los fallos, monitorización y matriz de sensibilidad aplicada al caso de fallos múltiples.

Al final de cada capítulo se muestran los resultados de diagnóstico aplicados a una turbina de gas SR-30 y finalmente en los capítulos 7 y 8 se presentan las conclusiones, trabajos futuros y la bibliografía.

1.6. Bibliografía

- [Ben-00] Benoff, D. (2000). Reliability Centered Maintenance Gives a new Meaning to the Phrase: If it can't Broke don't fix it, Business & Commercial Aviation Magazine.
- [Bla-02] Blázquez, L. (2002). Diagnóstico Automático de Fallos para Sistemas Dinámicos No Lineales. Publicación del Departamento de Ingeniería Eléctrica y Electrónica de la Universidad de León.
- [Bla-03] Blanke, M., Kinnaert, M. y Lunze, J.(2003). Diagnosis and Fault Tolerant Control. Springer.
- [Cam-07] Camacho, O. (2007). Fault Diagnostic base on multivariate statistical techniques. Revista técnica de la Universidad de Zulia. Vol. 30, No 3, pp. 253-262.
- [Cas-04] Casavola, A., Famularo, D. y Franze, G. (2004). Robust Multiple-fault detection and isolation: A gradient flow approach. Proceeding of the 2004 American Control Conference.
- [Dai-07a] Daigle, M. Koutsoukos, X. y Biswas, G. (2007). A Quantitative Approach to Multiple Isolation in Continuous Systems. Institute for Software Integrated Systems (ISIS). AAAI'07 Proceedings of the 22nd national conference on Artificial intelligence - Volume 1.
- [Dai-07b] Daigle, M. Koutsoukos, X. y Biswas, G. (2007). Multiple Diagnoses in Complex Physical Systems. Institute for Software Integrated Systems (ISIS).
- [Dav-92] Davis, R. y Hamscher, W. Model-based reasoning: Troubleshooting, in Hamscher, W., Console, L. y Kleer, J. (Eds.). (1992). Readings in Model-Based Diagnosis, Morgan Kaufmann Publishers, San Mateo, CA, pp. 3–24..
- [Fra-90] Frank, P. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy – A survey and some new results, Automática 26(3): pp.459–474.
- [Fra-96] Frank, P. (1996) Analytical and qualitative model-based fault diagnosis – A survey and some new results, European Journal of Control 2(1): pp. 6–28.
- [Gar-11] García, D., Fuente, M. (2011). Estudio Comparativo de detección de fallos basadas en Análisis de Componentes Principales. Elsevier.

- [Gas-02] Gastineau, Z. (2002). Propulsion Technology Planning for Engine Health Management, Controls and Engine Health Management, Turbine Engine Division Propulsion Directorate, Air Force Research Laboratory.
- [Ger-98] Gertler, J. (1998). Fault Detection and Diagnosis in Engineering Systems, Marcel Dekker, Inc. New York- Basel- Hong Kong.
- [Ger-10] Gertler, J., Romera, J., Puig, V. y Quevedo, J. (2010). Leak Detection and Isolation in water distribution network using Principal Component Analysis and Structured Residuals. Conference on Control and Fault Tolerant Systems.
- [Ham-92] Hamscher, W., Console, L. y Kleer, J. (1992) . Readings in Model-Based Diagnosis, Morgan Kaufmann, San Mateo, CA.
- [Ise-06] Isermann, R. (2006). Fault Diagnosis Systems. An Introduction from Fault Detection to Fault tolerance. Springer, Verlag. Germany.
- [Kor-04] Korbicz, J., Koscielny, J., Kowalczyk, Z. y Cholewa, W. (2004). Fault Diagnosis. Models, Artificial Intelligence, Applications, Springer-Verlag, Berlin.
- [Kos-04a] Koscielny, J. (2004). Methodology of process diagnosis, in Korbicz, J., Koscielny, J., Kowalczyk, Z. y Cholewa. Fault Diagnosis. Models, Artificial Intelligence, Applications, Springer-Verlag, Berlin, pp. 57–114. 2004.
- [Kos-04b] Koscielny, J. (2004). Models in process diagnosis, in Korbicz, J., Koscielny, J., Kowalczyk, Z. y Cholewa., Fault Diagnosis. Models, Artificial Intelligence, Applications, Springer-Verlag, Berlin, pp. 29–43.
- [Kem-06] Kempowsky, T., Subias, A., Aguilar, J. (2006) Process situation assessment: from a fuzzy partition to a finite state machine. Engineering Applications of Artificial Intelligence. Vol.19, N°5, pp.461-477.
- [Lig-08] Ligeza, J., Koscielny, J. (2008). A New Approach to Multiple Diagnosis: A Combination of Diagnostic matrices, Graphs, Algebraic and Rule-based Models. The case of two layer models. Int. Journal Appl. Math. Comput. Sci., Vol. 18, No.4, pp. 465-476.
- [Mor-07] Mora, J., Barrera, V. y Carrillo, G. (2007). Fault Location in Power Distribution Systems Using a Learning Algorithm for Multivariable Data Analysis. IEEE Transactions on Power Delivery, VOL. 22, No. 3.

- [Nie-10] Nieto, J., Castañón, L. y Morales, R. (2010). Multiple Fault Diagnosis in Electrical Power Systems with Dynamic Load Changes Using Probabilistic Neural Networks. *Computación y Sistemas*. Vol. 14 No. 1.
- [Pui-06] Puigjaner, L., Ollero, P., de Prada, C. y Jiménez, L. (2006). Estrategias de modelado, simulación y optimización de procesos químicos. Editorial Síntesis.
- [Reg-83] Reggia, J., Nau, D. y Wang, P. (1983) Diagnostic expert system based on a set covering model, *International Journal on Man-Machine Studies* 19(5): pp. 437–460.
- [Reg-85] Reggia, J., Nau, D. y Wang, P. (1985). A formal model of diagnostic inference. Problem formulation and decomposition, *Information Sciences* 37(1–3): pp.227–256.
- [Rei-87] Reiter, R. (1987). A theory of diagnosis from first principles, *Artificial Intelligence* 32(1): pp. 57–95.
- [Tza-89a] Tzafestas, S. (1989). System fault diagnosis using the knowledge-based methodology, in Tzafestas S. *Knowledge-Based System Diagnosis, Supervision and Control*, Plenum Press, New York, NY/London.
- [Tza-89b] Tzafestas, S. (1989). *Knowledge-Based System Diagnosis, Supervision and Control*, Plenum Press, New York, NY/London.

CAPITULO 2:

ESTADO DEL ARTE

Contenido

- 2.1. Introducción
- 2.2. Estudios Basados en Modelos Cuantitativos
- 2.3. Estudios Basados en Modelos Cualitativos
- 2.4. Estudios Basados en Análisis de Señales
- 2.5. Detección y Diagnostico de Fallos Múltiples en Turbinas de Gas
- 2.6. Sumario
- 2.7. Bibliografía

2.1. Introducción.

Existen una gran cantidad de trabajos de investigación relacionados con la detección y diagnóstico de fallos únicos y solo una pequeña cantidad relacionados con fallos múltiples. La mayoría de los métodos utilizados tanto como para fallos únicos y/o múltiples son analíticos, basados en inteligencia artificial (IA) o métodos estadísticos [Was-96], [Kos-03], [Kos-09], [Cas-04], [Xua-09], [Sam-07], [Dai-06], [Dai-07], [Lig-08], [Gar-11], [Zhu-10], [Nie-10].

Venkatasubramanian clasifica los métodos de detección y diagnóstico de fallos en tres grupos según lo muestra la figura 2.1. [Ven-03]:

- Basados en Modelos Cuantitativos.
- Basados en Modelos Cualitativos.
- Basados en Análisis de Señales.

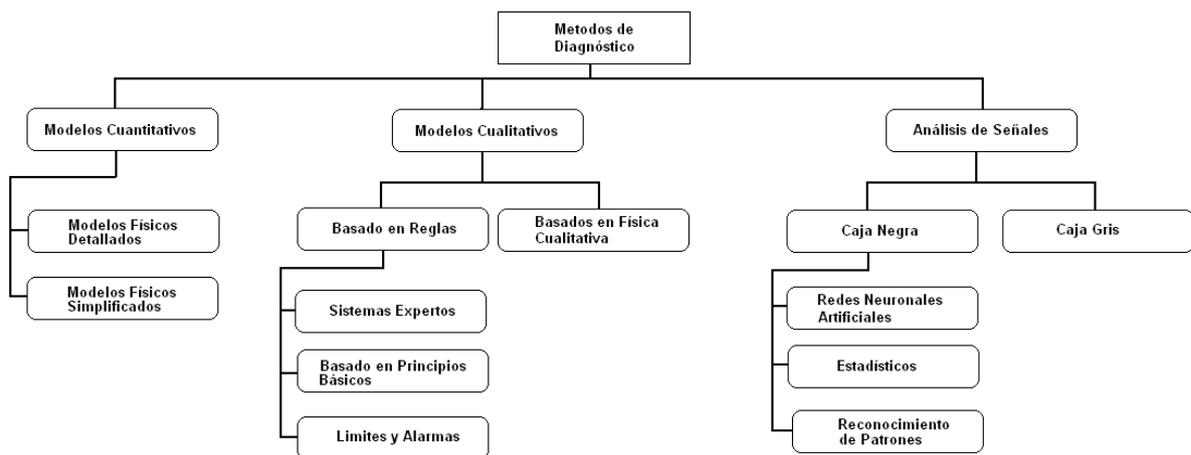


Figura 2. 1. Esquemático de clasificación de los diferentes métodos de detección y diagnóstico

Los métodos basados en modelos cuantitativos usados para la detección y diagnóstico de fallos se basan en modelos matemáticos del sistema. La aparición de uno o varios fallos es detectado por las discrepancias entre el comportamiento observado y el que es predicho por el modelo. Estos enfoques hacen uso de la estimación del estado, técnicas de identificación de parámetros y relaciones de paridad para generar residuos. Sin embargo, a menudo es difícil y se requiere mucho tiempo para desarrollar modelos matemáticos precisos que caractericen a todos los fenómenos físicos que ocurren en los procesos industriales.

Las fortalezas de los métodos de detección y diagnóstico de fallos basados en modelos cuantitativos incluyen:

- Los modelos se basan en fundamentos básicos de física o de ingeniería.

- Proporcionan los estimadores más precisos de salida cuando están bien formulados.
- Los modelos detallados basados en los principios fundamentales pueden modelar condiciones normales como operaciones bajo fallo; por lo tanto, los fallos pueden distinguirse fácilmente de las condiciones normales.
- Los transitorios en un sistema dinámico sólo pueden ser modelados con modelos físicos detallados.

Las debilidades de los métodos de detección y diagnóstico de fallos basados en modelos cuantitativos incluyen:

- Pueden ser complejos y computacionalmente intensivos.
- Hay que tener en cuenta la incertidumbre de los procesos y la existencia de partes no modelables de forma determinista (partes estocásticas).
- El esfuerzo requerido para desarrollar un modelo es significativo.
- Estos modelos generalmente requieren muchos insumos para describir el sistema, algunos de los valores que pueden no ser fácilmente disponibles.
- Al depender de extensas entradas de datos del usuario al modelo se le crean oportunidades de error de los datos de entrada que pueden tener un gran impacto en los resultados.

Los métodos de detección de fallos basados en modelos físicos de acuerdo con la literatura revisada se muestran como poco probables para ser los métodos elegidos en un futuro próximo, debido a las deficiencias enumeradas anteriormente, y los modelos físicos simplificados continuarán siendo útiles en las aplicaciones de detección y diagnóstico de fallos.

Los métodos basados en modelos cualitativos usados para la detección y diagnóstico de fallos se fundamentan en usar el razonamiento simbólico que generalmente combina diferentes tipos de conocimientos con la teoría de grafos para analizar las relaciones entre las variables de un sistema. Las principales ventajas de estos métodos son:

- Son muy adecuadas para los ambientes ricos en datos y procesos no críticos.
- Estos métodos son fáciles de desarrollar y aplicar.
- Su razonamiento es transparente y proporcionan la capacidad de razonar, incluso en condiciones de incertidumbre.
- Poseen la capacidad de dar explicaciones de los diagnósticos propuestos por la metodología basada en las relaciones de causa-efecto.
- Algunos métodos proporcionan la capacidad de realizar detección y diagnóstico de fallos sin un conocimiento preciso del sistema y valores numéricos exactos de entradas y parámetros.

Las principales desventajas de los métodos basados en modelos cualitativos son:

- Estos métodos son específicos para un sistema o proceso.
- Aunque estos métodos son fáciles de desarrollar, es difícil asegurar que todas las reglas son siempre aplicables especialmente cuando los sistemas son complejos.
- A medida que las nuevas condiciones se añaden para ampliar las existentes o adaptarse a las circunstancias especiales, la simplicidad del modelo se pierde.
- Estos modelos dependen en gran medida de la experiencia y los conocimientos de los desarrolladores.

Enfoques basados en conocimiento, tales como los sistemas expertos pueden ser considerados como enfoques alternativos o complementarios cuando los modelos analíticos no están disponibles.

Los métodos basados en el análisis de señales y usados para la detección de fallos requieren de una gran cantidad de datos históricos del proceso. Hay varias maneras en que estos datos pueden ser transformados y presentados como parte del conocimiento previo de un sistema. Estas transformaciones se conocen como la extracción de características y pueden ser cualitativas como las utilizadas por los sistemas expertos y los métodos cualitativos de análisis de tendencias o cuantitativos como los utilizados en las redes neuronales, Análisis de Componentes Principales, Mínimos Cuadrados Parciales y Reconocimiento Estadístico de Patrones.

Las principales ventajas de los métodos basados en modelos cualitativos son:

- Estos métodos se adaptan bien a los sistemas donde los modelos analíticos están poco desarrollados o son insuficientes para explicar el comportamiento observado.
- Son adecuados cuando existen abundantes datos de entrenamiento.
- Los modelos de caja negra son fáciles de desarrollar y no requieren de un conocimiento de la física del sistema que se está modelando.
- Los requisitos computacionales varían pero en general son manejables.
- Existe una gran cantidad de información disponible y bien documentada sobre los métodos matemáticos subyacentes.

Las debilidades de los procesos basados en los métodos de análisis de las señales incluyen:

- Los modelos de caja gris requieren en primera instancia una comprensión completa del sistema y experiencia estadística.
- La mayoría de los modelos no se pueden utilizar cuando se desea extrapolar más allá del rango de los datos de entrenamiento.

- Requieren una gran cantidad de datos de entrenamiento, así como del funcionamiento normal y del funcionamiento en condiciones de fallo.
- Los modelos son específicos para el sistema para el que se han desarrollado y rara vez se puede utilizar en otros sistemas.

Muy recientemente, la necesidad de desarrollar métodos más poderosos ha sido reconocida, y las técnicas híbridas que combinan varios métodos de razonamiento comienzan a utilizarse [Ven-03].

2.2. Estudios Basados en Modelos Cuantitativos

A continuación se presentan algunos de los estudios más recientes en el área de detección y diagnóstico de fallos múltiples basados en modelos cuantitativos.

Koscielny y Bartys proponen en el 2009 un método de descomposición dinámica para el aislamiento de fallos múltiples en sistemas a gran escala [Kos-09a], [Kos-09b]. El algoritmo desarrollado se caracteriza por la reducción significativa de la potencia de cálculo computacional necesaria para el aislamiento de fallos múltiples. En términos generales establecen que los fallos no son aislables si las firmas de fallo son iguales.

El aislamiento de fallos múltiples utilizado en este estudio implica la utilización de un gran conjunto de estados del sistema. En este estudio se llegó a la conclusión de que el aislamiento de fallos múltiples no es fácil de implementar en la práctica debido a la baja eficiencia y una demanda considerable de potencia de cálculo y espacio de memoria. Adicionalmente es importante saber si la búsqueda del espacio de estado del sistema es absolutamente necesaria para la realización del diagnóstico de fallos múltiples .

Koscielny ha desarrollado métodos de reducción del número de posibles estados considerados para el diagnóstico de fallos múltiples. Recientemente [Kos-09a] un novedoso algoritmo para el aislamiento de fallos múltiples (MUFIA) fue desarrollado. En principio el algoritmo MUFIA está compuesto de dos partes: El conjunto de posibles fallos es determinado en la primera parte del algoritmo. Asumiendo esta parte como de diagnóstico primario. Si el diagnóstico es satisfactorio, entonces el algoritmo no continua. La definición de los estados del sistema con fallos es llevada a cabo en la segunda parte del algoritmo. En esta parte, exclusivamente los estados con fallos obtenidos en el diagnóstico primario se consideran.

Al inicio el diagnóstico se basa en la suposición de fallos únicos. Si un desajuste ocurre entre los valores de la señal de diagnóstico y la firma de fallos únicos, entonces el algoritmo debe llevarse a cabo. Adicionalmente, es posible reducir el número de estados considerados del sistema. En la práctica la limitación arbitraria de la máxima multiplicidad de fallos es usada. Regularmente se

establece una multiplicidad máxima doble, ya que algunas veces es aceptable tomado en cuenta la baja probabilidad que figura en los fallos múltiples con más alta multiplicidad. Este método no considera el caso de cuando tres, cuatro o cinco componentes están alimentados por la misma energía y esta se dispara, entonces se podrían averiar todos los componentes y por lo tanto no sería una situación anormal.

Gertler, Romera, Puig y Quevedo en el 2010 desarrollaron un estudio sobre la detección y diagnóstico de fugas en redes de distribución de agua usando análisis de componentes principales y residuos estructurados [Ger-10]. Esta investigación fue un trabajo alternativo al previamente desarrollado en [Per-09], con la aplicación del análisis de componentes principales (PCA) en el diagnóstico de fallos en sistemas de distribución de agua propuestos en [Kre-91] y [Wis-96].

La metodología desarrollada por Gertler está compuesta de dos fases. En la primera fase un modelo del sistema libre de fallos es construido usando análisis de componentes principales a partir de los datos del sistema fuera de línea. Entonces, el efecto de los fallos es caracterizado a través de un análisis experimental de sensibilidad. Una vez que la matriz de sensibilidad de fallos ha sido obtenida, esta es usada para generar a partir del análisis de componentes principales el conjunto de residuos estructurados [Ger-99].

La segunda fase se realiza en línea e involucra el cálculo de los residuos y su evaluación con respecto a un límite de control, como usualmente es hecho en los sistemas basados en detección y diagnóstico de fallos (FDI). Cuando una inconsistencia es detectada (algunos residuos violan el límite de control), entonces el fallo puede ser diagnosticado usando una matriz que relaciona los fallos con los residuos, previo a un diseño de residuos estructurados.

En este estudio se demuestra el uso de la técnica en una red sencilla y se analizan casos con fallos simples mostrando un leve enfoque en los fallos múltiples.

2.3. Estudios Basados en Modelos Cualitativos.

A continuación se presentan algunos de los estudios más recientes en el área de detección y diagnóstico de fallos múltiples basados en modelos cualitativos.

Samantaray Ghoshal en 2009 desarrollaron una metodología que usaba la sensibilidad de los gráficos de unión para el aislamiento de fallos múltiples a través de estimación de parámetros [Sam-09]. En esta investigación la detección y el aislamiento de fallos múltiples es basada en modelos cualitativos y se lleva a cabo mediante estimación de parámetros. La hipótesis de fallos fue generada por la evaluación de un conjunto de relaciones de redundancia analítica (ARRs) y valores de parámetros

correspondiente a la parte no estructurada del subespacio de fallos que fue estimado por la minimización como función de las ARR.

Las incertidumbres del proceso y de medición se manejaron mediante un enfoque pasivo de detección y aislamiento de fallos. La modelación mediante gráficos de unión fue usada para describir el modelo del proceso y derivar las ARR. El algoritmo desarrollado provee un rápido aislamiento a fallos porque únicamente unos pocos parámetros son estimados y estos no necesitan más que unas pocas simulaciones del modelo, haciendo esto apropiado para procesos de supervisión en tiempo real.

En los modelos de gráficos de unión con bucles estructurales y relaciones implícitas, puede ser imposible derivar explícitamente las ARR. Por esta razón, un enfoque con generación implícita de residuos es seguido en este estudio para representar los modelos de gráficos de unión en la forma de un diagnóstico de gráficos de unión. La sensibilidad y el diagnóstico de los gráficos de unión del modelo, en sí mismos pueden ser acoplados simultáneamente en un vector con representación de gráficos de unión, con lo cual se obtiene la evolución en el tiempo de los residuos y la sensibilidad de los residuos con respecto a varios parámetros al mismo tiempo.

El desarrollo de procedimientos de optimización no requiere de modelos de simulación, pero si de una simple evaluación de unas pocas relaciones algebraicas. Una aplicación reciente de los mínimos cuadrados a los residuos para el diagnóstico de fallos múltiples concernientes a la detección de fugas y su localización en sistemas de transportación por tuberías se detalla en el algoritmo presentado por Gawthrop [Gaw-00].

Mathew Daigle [Dai-07] desarrolló una metodología para el diagnóstico de fallos múltiples en sistemas físicos complejos con un enfoque cualitativo. Este estudio se basó en el marco de la evolución del fallo [Mos-99], [Man-99], el cual emplea un enfoque cualitativo para el análisis de comportamiento de fallos transitorios. El modelo de diagnóstico se utiliza para generar las firmas de fallos que representan la magnitud y el orden más alto de los efectos de los fallos en las mediciones.

El estudio anterior proporciona un método sistemático para la generación de firmas de fallos múltiples a partir de las firmas de fallos simples. Se analizan las firmas de fallos múltiples para definir el concepto de la n -diagnosticabilidad, el cual define diagnosticabilidad con respecto al conjunto mínimo de fallo más probable, donde n es el máximo permitido de multiplicidad de fallo. Luego se presenta una extensión para el algoritmo de aislamiento de fallos en línea de evolución del fallo, tal que se encuentre el conjunto mínimo de fallo más probable que es consistente con las desviaciones de las mediciones observadas. Si un sistema es n -diagnosticable para alguna n , el algoritmo aísla un único candidato de fallo múltiple, si n o menos fallos ocurren.

La metodología denominada TRANSCEND [Mos-99] es una metodología bien desarrollada para el diagnóstico de fallos abruptos de sistemas físicos complejos con dinámica continua. Se emplea un enfoque basado en modelos cualitativos para el aislamiento de fallos. Los modelos del sistema se construyen utilizando gráficos de unión. Los fallos se modelan como cambios abruptos y persistentes en los valores de los parámetros de los componentes en el modelo de gráficos de unión del sistema.

El aislamiento de fallos usando la metodología TRANSCEND se basa en un análisis cualitativo de la dinámica transitoria causada por fallos abruptos. Las desviaciones en los valores de medición después de la aparición del fallo constituyen una firma de fallo, donde se predicen las desviaciones en la magnitud y los valores de la derivada de mayor orden se asignan a los símbolos {+, 0, -}, que corresponden a una desviación por arriba de lo normal, la no desviación, y una desviación por debajo de lo normal, respectivamente.

El aislamiento de fallos usando TRANSCEND utiliza una representación en gráfico temporal de causalidad (TCG) por sus siglas en inglés, las cuales pueden ser derivadas directamente del modelo de gráficos de unión del sistema. El TCG capta las relaciones causales y temporales entre las variables del sistema. Se especifica el gráfico de flujo de señales del sistema en una forma que las aristas están etiquetadas con los valores de los componentes individuales de los parámetros o las relaciones de proporcionalidad directa o inversa.

Las firmas de fallos son generadas usando un algoritmo de propagación hacia adelante en la TCG para predecir los efectos cualitativos de los fallos en las mediciones. El efecto cualitativo de un fallo, + o -, se propaga a todos los vértices de medición en el TCG para determinar las firmas de fallo para cada medición. Se denota el conjunto de todas las fallas como: $F = \{F_1, F_2, \dots, F_k\}$ y el conjunto de todas las mediciones como $M = \{m_1, m_2, \dots, m_\lambda\}$. Para $f, f \in F$ y $m \in M$, $\sigma_{f,m}$ es la firma de fallos para la medición m dado un fallo f . Dos fallos $f_i, f_j \in F$ son distinguibles usando firma de fallos sí $(\exists m \in M) \sigma_{f_i,m} \neq \sigma_{f_j,m}$.

En el algoritmo para fallos únicos, tan pronto un fallo único es aislado, este es declarado como un fallo verdadero y las futuras mediciones pueden ser ignoradas. En el caso de fallos múltiples, el aislamiento de un fallo único no necesariamente indica un fallo verdadero. Esto únicamente indica el diagnóstico más sencillo actual, dadas las desviaciones observadas hasta el momento. Consecuentemente las futuras desviaciones de las medidas pueden resultar en un mejor entendimiento de que fallos se producen realmente en el sistema. Si hay un candidato único en cualquier punto, el algoritmo lo devolverá. Debido a que más desviaciones de las mediciones sólo se puede ampliar a este candidato, el único candidato actual es parcialmente correcto. Futuras desviaciones mayores no pueden proporcionar un diagnóstico más exacto.

Ligeza y Koscielny en el 2008 presentan un nuevo enfoque en el diagnóstico de fallos múltiples usando una combinación de matrices de diagnóstico, gráfico, modelos basados en reglas [Lig-08]. Para el caso de un modelo de dos niveles se investigan algunas de las limitaciones del modelo de diagnóstico basado en una simple matriz binaria de diagnóstico para el caso de fallos múltiples. Varias interpretaciones posibles de la matriz de diagnóstico con sistemas basados en reglas se proporcionan y se analizan. Una propuesta de diagnóstico basada en matrices de un nivel se extiende a dos niveles, basada en el análisis causal y la incorporación de matrices "OR" y "AND". Una aproximación al diagnóstico de fallos múltiples basadas en el análisis de inconsistencia se describe y un procedimiento mejorado mediante un modelo cualitativo de las dependencias entre las variables del sistema se esboza.

El enfoque presentado se basa en la teoría clásica de Reiter, e incorpora las ideas y/o gráficos causales. Su objetivo es una ampliación y mejora del enfoque de detección y diagnóstico de fallos basado en una matriz de diagnóstico tipo "OR", lo cual es insuficiente en el caso de diagnóstico de fallos múltiples.

En este estudio se presenta un análisis de la metodología de matrices binarias de diagnóstico y sus limitaciones en el caso de sistemas complejos [Kos-01], [Kos-04a], [Kos-04b], así como la ocurrencia de fallos múltiples. Se argumenta que existen algunas limitaciones para este modelo, basado en la interpretación directa hacia adelante, pero se pueden superar si los procedimientos de búsqueda hacia atrás se aplican. Además, el modelo debe extenderse para cubrir el caso de fallos múltiples. Como consecuencia, un nuevo modelo de diagnóstico, compuesto de dos tipos de matrices que forman dos niveles, se desarrolla y se presenta en detalle.

El modelo propuesto por Koscielny abarca los modelos de análisis usando el análisis de redundancia analítica y consistencia basada en razonamiento [Cor-00a], [Cor-00b]. Simultáneamente se presenta una generalización del enfoque clásico de matrices de diagnóstico [Kos-01]. Tomando en consideración el punto de vista lógico el enfoque presentando es basado en la teoría clásica de Reiter [Reiter-87], y se incorporan las ideas de los gráficos causales "AND"/"OR" [Lig-97]. Se presenta una extensión y mejora de la metodología de detección y diagnóstico de fallos basado en el diagnóstico usando la matriz "OR", la cual es insuficiente en el caso de diagnóstico de fallos múltiples.

Las limitaciones de la aplicación directa de matrices de diagnóstico en el caso del diagnóstico de fallos múltiples fueron presentadas y examinadas. El principal problema con la metodología basada en la matriz de decisión binaria es que en los sistemas realísticos es muy raro que el conjunto de manifestaciones observadas proporcione suficiente información para asegurar diagnosticabilidad determinista, es decir, todos los fallos (incluso en el caso de fallos únicos) tienen firmas diferentes.

En el 2009, He-Xuan Hu en su investigación “Qualitative Multi-faults Diagnosis Based on Automated Planning I and II: Theory, Modeling, Algorithm and Case Study”, Propone un "autómata" o modelo matemático para una máquina de estado finita y "STRIPS" (Stanford Research Instituto Problem Solver) [Fik-71], una técnica clásica de planificación automatizada [Gha-04], como el medio para construcción de modelos. STRIP se utilizó para encontrar un operador de configuración que transforme un modelo inicial dado en uno que compensa una condición con objetivo específico. En los STRIPS el lenguaje describe acciones en termino de sus precondiciones y efectos, y describe los estados como conjunciones de literales positivos.

Los estados de condición previa que deben ser ciertos en un estado antes de una acción se pueden ejecutar. El efecto se describe cómo los cambios de estado cuando se ejecuta la acción. Una acción es "aplicable" en cualquier estado que satisface la condición previa, de lo contrario, la acción no tiene ningún efecto. El modelo se va construyendo poco a poco con la información recibida, no se construye en su totalidad de una vez. Los STRIPS pueden cualitativamente definir los modelos de los fallos sin necesidad de detalle y conocimiento preciso de los componentes defectuosos.

2.4. Estudios Basados en el Análisis de Señales

A continuación se presentan algunos de los estudios más recientes en el área de detección y diagnóstico de fallos múltiples basados en análisis de las señales.

Daqi Zhu y otros en el 2010 en su investigación “Multi-Fault Diagnosis Method for Sensor Systems Base on Principal Component Analysis” presentan un modelo basado en análisis de componentes principales, una red neuronal CMAC (cerebellar model articulation controller) y una red neuronal CA-CMAC (credit assigned cerebellar model articulation controller) para el diagnóstico de múltiples fallos en sistemas de sensores. En primer lugar, los valores pronosticados de los sensores se calculan mediante el uso de datos históricos medidos en condiciones normales y un modelo de PCA. En segundo lugar, se calcula el error de predicción al cuadrado (SPE) del sistema de sensores. Un fallo puede ser detectado cuando el SPE aumenta repentinamente. Si más de un sensor en el sistema está fuera de servicio, después de combinar los diferentes sensores y la reconstrucción de las señales de los sensores combinados, el SPE se calcula para localizar los sensores defectuosos.

Como resultado de las características de rápida convergencia de la CA-CMAC, las características de tiempo real del modelo de diagnóstico de fallo basado en PCA se mejoran. Los sensores defectuosos se pueden aislar mediante la reconstrucción de las señales de los sensores combinados, y por lo tanto se realiza un diagnóstico rápido de fallos múltiples de los sistemas de sensores.

La figura 2.2 muestra el algoritmo de diagnóstico propuesto basado en PCA y CA-CMAC. En el entrenamiento de la red neuronal CA-CMAC, las entradas son los valores históricos de los

componentes principales $Y_i (i = 1, 2, \dots, s)$ en los tiempos $(k-5, k-4, k-3, k-2)$, y la salida deseada son los valores históricos al tiempo $(k-1)$, entonces tantas muestras de entrenamiento se pueden obtener.

Nieto y otros en el 2010 en su estudio “Diagnostico de múltiples fallos en sistemas eléctricos de potencia con cambios de cargas dinámicas usando PCA” [Nie-10] presentan un método de diagnóstico de fallos que es capaz de localizar fallos múltiples en un sistema eléctrico de potencia con cambios de carga. Básicamente el método está compuesto de dos fases: En la primera fase una red neuronal probabilística es entrenada con los auto-valores obtenidos de la matriz de correlación de los datos de voltaje en operación normal y con fallos simétricos y asimétricos. La segunda fase emplea una comparación entre las muestras para detectar y localizar la presencia de un fallo. Este método tiene la ventaja que únicamente necesita datos históricos de funcionamiento normal y con fallo del sistema para entrenar una red neuronal probabilística.

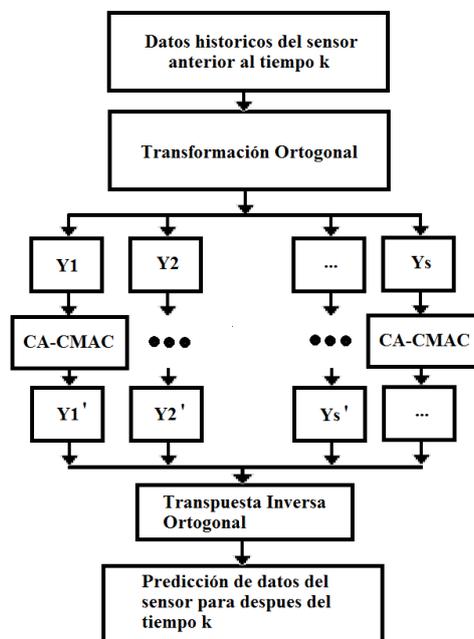


Figura 2. 2. Modelo basado en PCA Y CA_CMAC

Este método presenta como desventaja de que se necesita una gran cantidad de datos de operación normal y diferentes escenarios de fallos para un buen entrenamiento de la red neuronal. Estos algoritmos necesitan fallos desconocidos o conocidos pero con pocos datos no funciona, para sistemas complejos que cambian el punto de operación este método no funciona, ya que no hay capacidad de adaptación.

En el 2011 García Álvarez presenta un estudio comparativo de técnicas de detección de fallos basadas en el análisis de componentes principales [Gar-11] y compara diferentes métodos basados en análisis de componentes principales. Se evalúa las características de estos métodos de cara solamente a la detección de fallos. Analiza algunas de las limitaciones principales de los métodos basados en análisis

de componentes principales, dentro de las cuales se pueden nombrar: la aparición de numerosas falsas alarmas, la imposibilidad de detectar dos fallos consecutivos, o que los cambios en los puntos de operación sean detectados como falsas alarmas e incluso como fallos.

Los métodos presentados por García Álvarez son: APCA (Adaptive Principal Component Analysis) o PCA adaptativo, MSPCA (Multi-scale Principal Component Analysis) o PCA multi-escala, EWPCA (Exponentially Weighted Principal Component Analysis) o PCA pesado exponencialmente, MPCA (Moving PCA) o PCA con ventana deslizante, PCAEA (PCA using External Analysis) o PCA con análisis externo y por último NLPCA (Non Linear PCA) o análisis de componentes principales no lineales. [Zum-07], [Mis-02], [Lan-03], [Kan-04], [Kra-91].

En esta investigación no se hace énfasis en fallos múltiples, sino solamente en fallos únicos. Estaría abierta una ventana en el análisis de fallos múltiples con estas técnicas. En los métodos basados en datos históricos la detección y diagnóstico de fallos que se realizan mediante el procesamiento de un gran volumen de datos. Debido a esto, existen varias maneras en que los datos pueden ser transformados y presentados como un conocimiento a priori para el sistema de diagnóstico [Cam-07].

2.5. Diagnóstico de Fallos Intermitentes

Se denominan fallos intermitentes aquellos que ocurren de vez en cuando y no son de carácter permanente. Algunos autores como Correcher, García, Blasco y Quiles han estudiado el diagnóstico de fallos intermitentes mediante metodologías de modelos de eventos discretos y han presentado diversas estrategias para abordar el tema utilizando parámetros que caracterizan los fallos, tales como la duración, el pseudoperiodo, la persistencia y el número de fallos ocurridos [Cor-04][Cor-10].

Los parámetros que caracterizan a los fallos intermitentes y que aportan información útil al supervisor humano son:

Duración: Definido como el tiempo que permanece activo el fallo en cada una de sus apariciones.

Pseudoperiodo: Definido como el tiempo ocurrido entre dos detecciones consecutivas del fallo. Dada la naturaleza imprevisible de los fallos intermitentes, no parece razonable hablar de periodo, puesto que no se trata de señales periódicas. No obstante, la información relativa a la frecuencia con que aparecen los fallos, así como la evolución de dicha frecuencia es información que se usa para el algoritmo de diagnóstico.

Persistencia: Definida como la inversa del pseudoperiodo. Este parámetro da una estimación de la frecuencia de aparición del fallo.

La mayoría de las metodologías de detección de fallos asumen que los fallos diagnosticados son de carácter permanentes y para el caso de fallos intermitentes los diagnosticadores no son capaces de retornar al bucle de funcionamiento normal aunque el sistema quede reparado. Generalmente los casos de ocurrencia de fallos intermitentes son el prelude para la aparición de fallos permanentes y generalmente son consecuencia de una degradación o envejecimiento de los componentes o sensores.

Los fallos intermitentes podrían verse como fallos múltiples secuenciales cuando el tiempo de restitución del fallo o fallos es muy grande. El diagnóstico usando la dinámica de los fallos intermitentes presenta problemas al momento de determinarse un tamaño de ventana deslizante. En esta tesis no se estudian este tipo de fallos y se recomienda como parte de un estudio futuro.

2.6. Detección y Diagnóstico de Fallos Múltiples en Turbinas de Gas

El estudio y desarrollo de los sistemas de monitoreo de fallos en turbinas de avión es uno de los aspectos clave de la seguridad aérea. Las técnicas actuales de monitoreo de fallos consisten básicamente de inspecciones regulares y de mantenimiento, que en esencia son por naturaleza conservadoras y no son costo efectivo. Por otra parte, en circunstancias excepcionales durante la inspección rutinaria se pueden no detectar agravantes o fallos ocultos, que pueden en un futuro dar lugar a daños permanentes o a accidentes potencialmente catastróficos. Existe entonces una necesidad de desarrollar nuevas tecnologías de detección y diagnóstico para las turbinas de gas usadas en aviación. Estas tecnologías se podrían implementar ya sea durante el vuelo o mientras el avión aterriza y es estacionado, pero sin la necesidad de la instalación de sensores adicionales y/o montar la turbina en un banco de pruebas.

Una turbina de gas es un sistema dinámico complejo a gran escala cuyos subsistemas están interconectados físicamente, así como a través de lazos de control de retroalimentación. Es una tarea difícil el detectar, aislar y estimar la gravedad de los fallos. Este tema ha sido abordado por muchos investigadores en los últimos decenios.

Hoffman y Van der Merwe [Hof-02] han demostrado que los métodos tradicionales basados en el análisis de las frecuencias no pueden ser fácilmente aplicables para estimar la evolución de fallos múltiples en las turbinas de gas; patrones de fallos han sido usados para aliviar el problema, pero ha sido inviable almacenar mucha información para sistemas abordo debido al gran tamaño [Bop-96]. La literatura reciente ha reportado varios métodos, como los basados en la estimación conjunta del estado [Din-01], ecuaciones de paridad [Son-03], entre otras más.

Varias técnicas basadas en modelos y basadas en datos han sido reportados en la literatura para la detección, el diagnóstico y el diagnóstico en turbinas de gas, que incluyen linealización estadística, filtrado de Kalman [Van-01], filtro de Kalman extendido [Wan-00], filtrado convencional [Chi-06],

Cadena de Markov y de Monte Carlo [Bre-01], redes bayesianas [Wan-06], redes neuronales [Hof-02], estimación máxima probabilidad [Dav-02], herramientas basadas en Wavelet [Gha-01] y algoritmos genéticos [Yao-94].

En un sistema complejo como lo es una turbina de gas, consistente de múltiples componentes interconectados, un fallo simple en un componente puede producir una respuesta anómala que puede servir como una excitación de entrada a otros componentes en buen estado y degradar el desempeño del sistema. Por tal manera, una respuesta anómala de un componente del sistema puede producirse debido a dos posibles causas: a) fallo en el componente, y/o b) fallo de excitación debido a algún otro componente. En un escenario extremo y adverso, la gradual evolución de pequeños fallos en componentes individuales puede conducir a fallos en cascada porque las fuertes interconexiones de entrada-salida y retroalimentación entre los componentes de la turbina de gas y producir fallos catastróficas y forzar el paro total de la misma.

Sistemas de detección de fallos basados en algoritmos de reconocimiento de patrones han sido experimentalmente validados para ejecuciones en tiempo real en diferentes aplicaciones, tales como en circuitos electrónicos [Chi-05] y monitoreo de fallos debidos a fatiga en el material [Gup-07a][Gup-07b]. Estudios recientes implementan el análisis de componentes principales (PCA) y redes neuronales artificiales [Chi-05][Gup-07b].

En general, los fallos en una turbina de gas se clasifican en tres categorías principales:

- 1- Pérdida de eficiencia en subsistemas (compresor y turbina).
- 2- Degradación de los sensores (por ejemplo: interrupciones y fluctuaciones).
- 3- Fallos del actuador (por ejemplo: lenta respuesta, excesiva banda muerta e histéresis).

Los anteriores escenarios son los implementados en esta tesis para demostrar el desempeño de los algoritmos de detección y diagnóstico propuestos.

2.6. Sumario

En general hay diferentes técnicas y combinaciones de ellas utilizadas en la detección y diagnóstico de fallos múltiples, pero dependerá de la naturaleza del problema el seleccionar la técnica que tiene el mejor rendimiento. Todos estos métodos no son excluyentes entre sí, sino que son complementarios. La complejidad de la tarea para la que se diseñan hace que no tenga que ser abordada desde una única perspectiva.

Considerando que la mayoría de los sistemas complejos son complicados de modelar matemáticamente o mediante modelos de gráficos de unión, se ha enfocado esta tesis al uso de

análisis de las señales, específicamente análisis de componentes principales lineal y no lineal como métodos de detección de fallos múltiples y los residuos estructurados como método de diagnóstico.

En un sistema complejo como lo es una turbina de gas, consistente de múltiples componentes interconectados, un fallo simple en un componente puede producir una respuesta anómala que puede servir como una excitación de entrada a otros componentes en buen estado y degradar el desempeño del sistema completo. En general, los fallos en una turbina de gas se clasifican en tres categorías principales: pérdida de eficiencia en subsistemas; degradación de los sensores; fallos en actuadores.

2.7. Bibliografía

- [Bre-01] Bremer, C. y Kaplan, D. (2001). Markov chain and Monte Carlo estimation of nonlinear dynamics from time series. *Phys D*.
- [Bro-74] Broersen P. (1974). Estimation of parameters of nonlinear dynamical systems. *Int. J. Non-linear Mech.*
- [Bop-96] Boppana, V., Hartanto, I. y Fuchs, W. (1996). Fault diagnosis using state information. *Proceeding of the 26th annual international symposium on fault tolerant computing.*
- [Chi-05] Chin, S. y Ray, A. (2005). Symbolic time series analysis for anomaly detection: a comparative evaluation. *Signal process.*
- [Chi-06] Ching, J., Beck J. y Porter, K. (2006) Bayesian state and parameter estimation of uncertain dynamical systems. *Probab. Eng. Mech.*
- [Cor-04] Correcher, A. y García, E. (2004). Diagnóstico de Fallos Intermitentes basado en modelos de eventos discretos. *WODES04.*
- [Cor-10] Correcher, A. y García, E. (2010). *Factory Automation. Chapter 29. Diagnosis of Intermittent Fault Dynamics. InTech.*
- [Dav-02] David, B. y Bastin, G. (2002). Parameter estimation in nonlinear system with auto and cross correlated noise. *Automatic.*
- [Din-01] Ding, Y., Wu, Z. y Zhang, Y. (2001). Multi-fault diagnosis method based on a joint estimation of states and estimation parameters. *J. Tsinghua Univ.*
- [Fao-05] Faouzi, M., Mourot, G. y Ragot, J. (2005). Sensor Fault and Isolation of an Air Quality Monitoring Network using Nonlinear Principal Component Analysis. *Proceeding of the 16th IFAC World Congress.*

- [Fao-03] Faouzi, M., Mourot, G. y Ragot, J. (2003). Nonlinear PCA Combining Principal curves and RBF-Networks for Process Monitoring. Proceeding of the 42nd IEEE Conference on Decision and Control.
- [Gar-11] García, D. y Fuente, M. (2011). Estudio Comparativo de detección de fallos basadas en Análisis de Componentes Principales. Elsevier.
- [Ger-98] Gertler, J. (1998). Fault Detection and Diagnosis in Engineering Systems, Marcel Decker,(Inc. New York- Basel- Hong Kong.
- [Ger-99] Gertler, J., Li, W., Huang, Y. y McAvoy. (1999). Isolation Enhanced Principal Component Analysis. AIChE Journal, 45, pp. 323-334.
- [Ger-10] Gertler, J., Romera, J., Puig, V. y Quevedo, J. (2010). Leak Detection and Isolation in water distribution networks using Principal Component Analysis and Structured Residuals. Conference on Control and Fault Tolerant Systems (Systol).
- [Gha-01] Ghanem, R. y Romero, E. (2001). A wavelet-based approach for model and parameter identification of nonlinear systems. Int. J. Non-linear Mech.
- [Gup-07a] Gupta, S. y Ray, A. (2007). Fatigue damage monitoring by ultrasonic measurements: a symbolic dynamic approach. Int. J. Fatigue.
- [Gup-07b] Gupta, S. y Ray, A. (2007). Ray. Real time fatigue life estimation in mechanical systems. Meas. Sci. Technol.
- [Gaw-00] Gawthrop, P. (2000). Sensitivity bond graphs. J. Franklin Inst., pp. 907-922.
- [Hof-02] Hoffman, A. y Van der Merwe, N. (2002). The application of neural network to vibrational diagnostics for multiple fault conditions. Comput. Stand. Interfaces.
- [Kor-04] Korbicz, J., Koscielny, J., Kowalczyk y Cholewa, W. (2004). Fault Diagnosis. Models, Artificial Intelligence, Applications, Springer-Verlag, Berlin.
- [Kem-06] Kempowsky, T., Subias, A. y Aguilar, J. (2006). Process situation assessment: from a fuzzy partition to a finite state machine” Engineering Applications of Artificial Intelligence. Vol.19, N°5, pp.461-477.
- [Kos-03] Koscielny, J. y Bartys, M. (2003). Multiple Faults Isolation in Diagnostic of Industrial Processes. European Control Conference.

- [Kos-09a] Koscielny, J. y Bartys, M. (2009). Application of the Method of Dynamic Decomposition for Recognition of Multiple Faults in the Large Scale systems. 7th Workshop on Advanced Control and Diagnostic ACD 2009 (a).
- [Kos-09b] Koscielny, J., Bartys, M. y Syfert M. (2009). Method of Multiple Isolation in Diagnostics of Large Systems. 7th Workshop on Advanced Control and Diagnostic ACD 2009 (b).
- [Kre-91] Kresta, J., , MacGregor, J. y Marlin, T.E. (1991). Multivariate Statistical Monitoring of processes, "Can. J. Chem. Eng., pp. 69- 35.
- [Lig-08] Ligeza, J. y Koscielny, J. (2008) A New Approach to Multiple Diagnosis: A Combination of Diagnostic matrices, Graphs, Algebraic and Rule-based Models. The case of two layer models. Int. Journal Appl. Math. Comput. Sci., Vol. 18, No.4, 465-476.
- [Moh-07] Mohamed, H., Salah, D., Noureddine, D. y Mohamed, B. (2007). Sensor Fault Detection, Isolation and Reconstruction Using Nonlinear Principal Component Analysis. International Journal of Automation and Computing.
- [Mor-07] Mora-Flórez, J., Barrera-Núñez, V. y Carrillo-Caicedo, G. (2007). Fault Location in Power Distribution Systems Using a Learning Algorithm for Multivariable Data Analysis. IEEE Transactions on Power Delivery, VOL. 22, No. 3.
- [Moh-12] Mohamed, H. y Messaoud, B. (2012). Nonlinear Principal Component Analysis combining TT Net and RBF neural networks for process monitoring. Proceeding of the 20th Mediterranean conference on Control and Automation.
- [Mos-99] Montermán, P. y Biswas, G. (1999). Diagnosis of continuous valued systems in transients operating regions. IEEE Transaction on systems, Man and Cybernetics, Part A.
- [Oga-02] Ogaji, S., Singh, R. y Probert, S. (2002). Multiple-Sensor fault diagnoses for a 2 shaft stationary gas turbine.
- [Per-09] Pérez, R., Puig, V., Pascual J., Peralta A., Landeros, E. y Jordanas, Ll. (2009). Pressure sensor distribution for leak detection in Barcelona water distribution network". Water Science & Technology, Vol. 9, No 6, pp. 715–721.

- [Pul-05] Pulido, B., Puig, V., Escobet, T. y Quevedo, J. (2005). A new fault algorithm that improves the integration between fault detection and localization in dynamic systems. Decision and control 2005. European Control Conference.
- [Rei-87] Reiter, R. (1987). A theory of diagnosis from first principles, *Artificial Intelligence* 32(1): pp. 57–95.
- [Sch-12] Scholz, M. (2012). Validation of nonlinear PCA. *Neural Proceeding Letters*, Volume 36, Number 1. pp. 21-30.
- [Sch-02] Scholz, M. y Vigarrio, R. (2002). Nonlinear PCA: a new hierarchical approach. In M. AVerleysen, editor, *Proceedings ESANN*.
- [Sch-07] Scholz, M., Fraunholz, M. y Selbig, J. (2007). Nonlinear Principal Component analysis: Neural Network Model and Applications. In *Principal Manifolds for Data Visualization and Dimension Reduction*. Volume 58 of LNCSE, pp. 44-67. Springer Berlin Heidelberg.
- [Son-03] Song, H. y Zhang, H. (2003). Approach for multiple faults diagnosis based on parity equations and parameter estimation. *Control Decis.*
- [Sim-03] Simani, S., Fantuzzi, C. y Patton, R. (2003). Model based fault diagnosis in dynamical systems using identification techniques.
- [Tol-06] Tolani, D., Yasar, M. y Ray, A. (2006). Anomaly detection in aircraft gas turbine engines. *J. Aerosp. Comput. Info. Commun.*
- [Van-01] Van, L. y Witteveen, F. (2001). Multiple nonlinear parameter estimation using PI feedback control. *Control engineering pract.*
- [Ven-03] Venkatasubramanian, V. (2003). A review of process fault detection and diagnosis. *El Sevier, Computer and Chemical Engineering* 27 pp. 327-346.
- [Wis-96] Wise, B. y Gallagher, N. (1996). "The Process Chemometrics Approach to Chemical Process Fault Detection and Supervision," *Journal of Process Control*, 6(6), pp. 329-348.
- [Wan-00] Wan, E. y and Van der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. In *proceedings of the IEEE symposium*.

- [Wan-06] Wang, Y. y Geng, I. (2006). Bayesian network based fault section estimation in power systems. In proceeding of the IEEE region 10th annual international conference TENCON.
- [Yao-94] Yao, I. y Sethares, W. (1994). Nonlinear parameter estimation via the genetic algorithm. IEEE Trans. Signal Process.

CAPITULO 3:

DESCRIPCION DEL SISTEMA

TURBINA DE GAS SR-30

Contenido

- 3.1. Descripción del sistema MiniLab Gas Turbine
- 3.2. Principio de funcionamiento
- 3.3. Modelo Matemático de la Turbina de Gas SR-30
- 3.4. Modelo en Simulink de la Turbina de Gas SR-30
- 3.5. Fallos estudiados en la Turbina de Gas SR-30
- 3.6. Sumario
- 3.7. Bibliografía

3.1. Descripción del Sistema "MiniLab Gas Turbine"

El sistema "MiniLab Gas Turbine" es un equipo de laboratorio diseñado por la compañía Turbine Technologies Ltd., compuesto por una turbina jet modelo SR-30, especialmente diseñado para propósitos educativos y de investigación en ingeniería. El sistema MiniLab Gas Turbine permite explorar todos los aspectos de la teoría de turbinas de gas de una manera fácil y práctica. La turbina de gas SR-30 es una representación de todas las turbinas tipo jet rectos, en los cuales la combustión da lugar a la expansión de un gas que es suficientemente capaz de producir trabajo útil y empuje propulsivo. Estos tipos de turbinas se utilizan en los aviones, sistemas de defensa, propulsión de marina y para la producción de energía usada a nivel industrial.

El sistema "MiniLab" básicamente está compuesto de la turbina de gas SR-30, el sistema de control multivariable y el sistema de adquisición de datos. La turbina de gas SR-30 consiste esencialmente de un compresor de flujo centrífugo, una cámara de combustión anular y una turbina de flujo axial. Todos los sistemas de control de combustible y aceite de lubricación están integrados en el gabinete del MiniLab. Las predicciones teóricas se pueden medir en el sistema de adquisición de datos incluido. Los sensores situados a lo largo de la trayectoria del flujo del gas permiten obtener medidas exactas, en tiempo real de las condiciones de funcionamiento en estos puntos. La figura 3.1 muestra el sistema "Minilab Gas Turbine" instalado en el laboratorio de la Universidad Interamericana de Puerto Rico, Recinto de Bayamón.



Figura 3. 1. Sistema "Minilab" de la compañía Turbine Technologies Inc.

3.1.1. Turbina de Gas SR-30

La turbina de gas SR-30 es una versión reducida a escala, que suena y huele como una verdadera usada para propulsión en los aviones comerciales. Tiene un largo total de 61 centímetros y de un

diámetro medio de 16.5 centímetros y otras especificaciones que se detallan más adelante. La figura 3.2 muestra la vista seccionada de la turbina de gas SR-30.

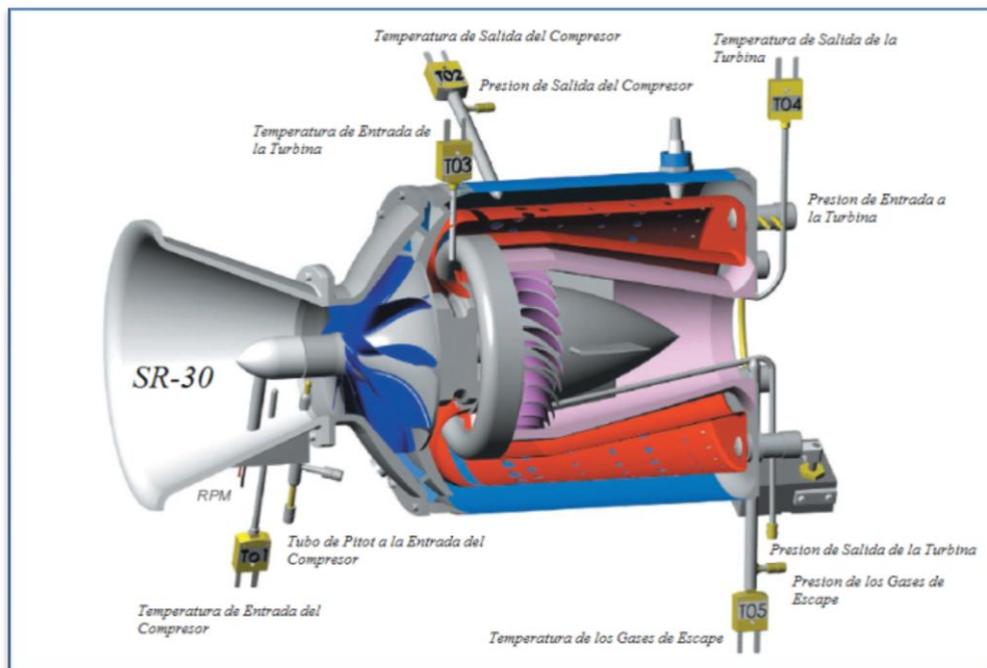


Figura 3. 2. Vista seccionada de la turbina SR-30

El fabricante proporciona la siguiente descripción de los sensores y de su localización, según se presenta en la tabla 3.1.

Tabla 3. 1 Especificaciones de los sensores instalados en la turbina de gas SR-30

Sensor	Variable	Tipo	Especificaciones
(P01)	Presión estática de entrada del compresor.	Transductor de presión tipo diafragma piezo-resistivo.	Entrada: 9-30 VDC. Salida: 0.5- 5.5 VDC. Rango: 0 - 1378.951KPa Precisión: 0.25% valor de la escala completa. Temperatura: 20-80°C
(P02)	Presión de estancamiento de salida del compresor.		
(P03)	Presión de la cámara de combustión		
(P04)	Presión de estancamiento de salida de la turbina.		
(P05)	Presión de estancamiento de la salida de la tobera.		
(T01)	Temperatura estática de entrada del compresor.	Termopar	Tipo K, Rango: 20 - 1335°C. Buena estabilidad a alta temperatura.
(T02)	Temperatura de estancamiento de la salida del compresor.		
(T03)	Temperatura de estancamiento de la entrada de la turbina.		
(T04)	Temperatura de estancamiento de la salida de la turbina.		
(T05)	Temperatura de estancamiento de la salida de la tobera.		
(Fw)	Flujo de combustible	Sensor volumétrico	Precisión del 0.5% del valor de la escala completa.

(Tn)	Torque	Celda de carga	Celda de carga Sensy 2712 capacidad máxima 200N. Sensitividad 1.85 mV/V.
(N)	Revoluciones por minuto (RPM)	Sensor efecto Hall	Máximo 87000 rpm

El sistema MiniLab Gas Turbine provee en el panel de control indicadores digitales para la temperatura de entrada de la turbina (T03), temperatura de salida de los gases (T05) y para la velocidad rotacional de la turbina (RPM). Adicional provee indicadores análogos para la presión de aceite, presión de la turbina, presión de aire para el arranque de la turbina. La turbina SR-30 de fábrica no provee control automático de las RPM, simplemente se abre o cierra el flujo de combustible que pasa por una válvula con actuador manual que es accionado por un mecanismo de palanca-cable como se muestra en la figura 3.3.

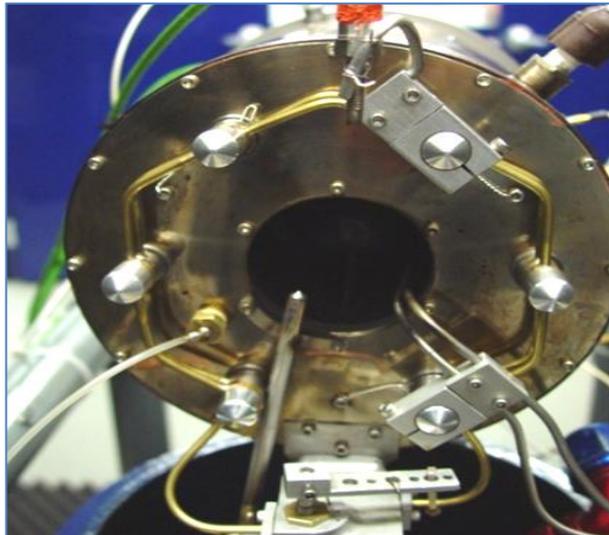


Figura 3. 3. Válvula manual de control de combustible

La turbo jet SR-30 es capaz de correr usando varios grados de combustible jet, incluyendo militares y grados comerciales de kerosene y combustible Diesel. La tabla 3.2 muestra los varios tipos de combustible que pueden ser usados. Esta también muestra las densidades y el calor de combustión para cada combustible. Para el estudio de la presente tesis se utilizó Jet-A. En la tabla 3.3 se presentan las condiciones de operación de la turbina de gas SR-30.

Tabla 3. 2. Propiedades para combustible Jet

Propiedades Típicas para Combustibles Jet						
	Militar / Kerosene comercial			Diesel de baja temperatura		Diesel Militar
Propiedad	JP-4 (Jet B)	JP-5	JP-8 (Jet A/Jet A-1)	DL-1	DL-2	F-54
Densidad (Kg/L)	0.755	0.817	0.797	0.812	0.852	0.830
Calor de Combustión (KJ/Kg)	43571	42929	43008	43219	42917	42851

Tabla 3. 3. Condiciones de operación de la turbina SR 30

Condiciones de Operación	Limitaciones
Máximo Torque de Diseño	178 N
Combustibles Aprobados	JET A, A-1, B, JP-4, 5, 8. Kerosene, Diésel, Aceite Combustible #1 o #2.
Temperatura de los gases de escape	720°C
Flujo másico de aire	0.5 Kg/s
Sistema de encendido	Espacio de Aire, descarga de alto voltaje de un capacitor tipo bobina de encendido herméticamente sellada y enchufe de encendido.
Tipo de Compresor	Centrifugo de una sola fase (Flujo de salida radial).
Tipo de Turbina	Flujo axial de una sola fase.
Máxima RPM de diseño	87,000
Soporte del motor	Soporte de pivote con cojinete que permite torque directo para colocar una celda de carga.
Razón de compresión del motor	3.4
Razón de presión del motor	30.0
Consumo específico de combustible	1.2

3.1.2. Sistema de Adquisición de Datos

El sistema MiniLab Gas Turbine viene de fábrica con un módulo de adquisición de datos marca IOtech's Personal Daq™ 100. Para propósito de esta investigación este módulo fue remplazado por un sistema Compact Rio de la compañía National Instrument y programado con Labview 2013 debido a que el sistema original tenía limitaciones con el tiempo de muestreo y no permitía adicionar más sensores y control de actuadores. La tabla 3.4 muestra los módulos instalados para conectar los diferentes sensores y el servomotor DC. La figura 3.4 muestra el sistema completo instalado en el laboratorio de la Universidad Interamericana de Puerto Rico, Bayamón.

Tabla 3. 4. Módulos Compact Rio instalados para la conexión de sensores

Modulo	Función
NI-CRIO 9004	Controlador en Tiempo Real con 64 MB DRAM, 512 MB Compact Flash. Rango de temperatura de -40 a 70 °C; doble entrada de suministro de 9 a 35 VDC.
NI-9211	Módulo para 4 termopares o entradas analógicas de ± 80 mv. Resolución de 24 bits, rechazo de ruido de 50/60 Hz. Velocidad de muestreo de 14 muestras/s.
NI-9274	Módulo de entradas y salidas digitales. Velocidad de muestreo de 14 muestras/s.
NI-9201	Módulo de Voltaje de 8 Canales de 500 kS/s para USB, Ethernet o Wi-Fi. Rango de entrada ± 10 V. Entradas de una sola terminal, resolución de 12 bits. Velocidad de muestreo de 14 muestras/s.
NI-9233	Módulo 4 entradas analógicas muestreadas simultáneamente hasta 50 kS/s. ADCs de resolución de 24 bits con rango dinámico de 102 dB. Rango de entrada de ± 5 V. Velocidad de muestreo de 14 muestras/s.
NI-9219	Módulo Universal de 4 Canales, Medidas de termopares, RTD, resistencia, voltaje y corriente. Velocidad de muestreo de 14 muestras/s.
NI-9505	Módulo de Drive Servo de DC de Escobillas con Puente H Completo.



Figura 3. 4. Sistema de adquisición Compact Rio instalado

Para mayor flexibilidad en cambios en el tiempo de muestreo, manipulación y control de las diferentes variables en el sistema se elaboró un programa en Labview 2013, con el cual se automatizó el almacenamiento de datos correspondientes a los sensores instalados (temperatura, presión, revoluciones, flujo y torque). La válvula de control de combustible se automatizó con un servomotor DC interconectado con un mecanismo tipo trapezoide, según se observa en la figura 3.5 y de esta manera lograr controlar las revoluciones de la turbina.



Figura 3. 5. Control automático de la válvula de combustible mediante un servomotor DC

3.2. Principio de funcionamiento

La turbina de gas SR-30 es modelada por el ciclo de Brayton, tal como se presenta en [Ail-01], [Leo-08], [Wat-04], [Rah-11], [Fre-03], [Per-03], [Wit-03] y [Sar-01] que emplea el aire como el fluido operante. El proceso básico es comprimir el aire, agregar el combustible, quemar la mezcla aire-combustible, y utilizar la energía generada para desarrollar empuje. Puesto que el motor SR-30 no se ata a un aeroplano móvil, una sección acampanada a la entrada permite crear un perfil uniforme de la velocidad del aire al ingresar en el compresor y así reducir al mínimo las pérdidas. El ciclo de Brayton para la turbina SR-30 se compone de los siguientes procesos:

- 1) Aceleración isoentrópica a través de una tobera de entrada.
- 2) Compresión isoentrópica a través del compresor.
- 3) Adición de calor a presión constante a través de la cámara de combustión.
- 4) Expansión isoentrópica a través de la turbina.
- 5) Expansión isoentrópica a través de una tobera de salida de gases.

La figura 3.6 muestra un diagrama esquemático del sistema y de su ciclo ideal asociado.

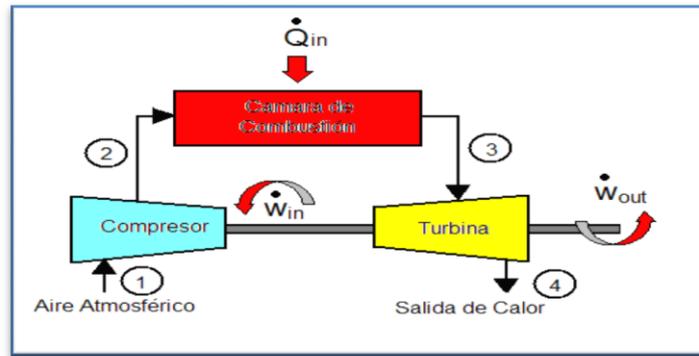


Figura 3. 6. Diagrama esquemático del ciclo de Brayton

Sobre el ciclo básico existen variantes que lo complican para mejorar alguna de sus prestaciones, como la potencia o el rendimiento. Ejemplos de estas variantes son el ciclo Brayton con regenerador, el ciclo Brayton de múltiples etapas (de compresión u expansión), o la combinación de un ciclo Brayton con un ciclo Rankin para dar lugar al denominado ciclo combinado. La figura 3.7 muestra el diagrama de la turbina de gas SR-30 con la nomenclatura para cada punto de medición.

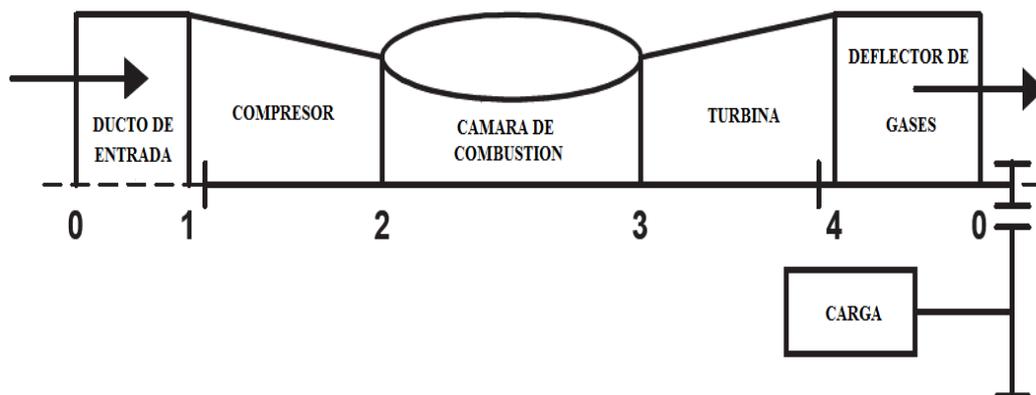


Figura 3. 7. Diagrama de la turbina de gas

3.3. Modelo Matemático de una Turbina de Gas SR-30

Mientras que los diseños basados en modelos lineales son adecuados en ciertas circunstancias, nuevas técnicas para el análisis y diseño de turbinas de gas se han empleado con el objeto de acomodar los efectos no lineales. En consecuencia el modelado de la turbina de gas ha recibido especial atención en los últimos años con los trabajos dedicados a la identificación y análisis de sistemas no lineales [Ail-01], [Leo-08], [Wat-04], [Rah-11], [Fre-03], [Per-03], [Wit-03]. Para este estudio con el fin de obtener un modelo dinámico de bajo orden para propósitos de simplificación del modelo se realizan las siguientes suposiciones tal como se presenta en [Ail-01].

Suposiciones Generales:

- a) Las propiedades de las constantes físico químicas son asumidas en las partes principales de la turbina de gas, tales como el calor específico a presión y volumen constante, constante específica de los gases y el exponente adiabático.
- b) Las pérdidas de calor (conducción, convección y radiación) son insignificantes.

Otras Suposiciones

c) El coeficiente de pérdida de presión en el ducto de entrada (σ_1) se asume constante. Lo cual significa que el total de pérdida de presión en el ducto de entrada es un porcentaje fijo de la presión total (P_0^*).

d) En el compresor:

- El flujo másico es constante ($\dot{m}_{Cin} = \dot{m}_{Cout} = \dot{m}_C$).
- No hay efecto de energía almacenada ($U^*_2 = constante$).

e) En la cámara de combustión:

- El coeficiente de pérdida de presión se asume constante (σ_{comb}).
- La eficiencia de la combustión se asume constante (η_{comb}).
- La entalpia del combustible es insignificante.

f) En la turbina:

- El flujo másico es constante: $\dot{m}_{Tin} = \dot{m}_{Tout} = \dot{m}_T$.
- No hay efecto de energía almacenada ($U^*_4 = constante$).

g) En el deflector de gases el coeficiente de presión se asume constante (σ_N).

3.3.1. Desarrollo de las ecuaciones del modelo.

El desarrollo de las ecuaciones del modelo de la turbina de gas está simplificado en los siguientes pasos y basado en el modelo presentado en [Ail-01][Rin-12][Rin-13]:

Realización de un balance de conservación de masa.

$$\frac{dm}{dt} = \frac{dm_{in}}{dt} - \frac{dm_{out}}{dt} \quad (3.1)$$

Realización de un balance de conservación de energía interna, donde el flujo de calor y el trabajo son tomados en cuenta:

$$\frac{dU}{dt} = \frac{dm_{in}}{dt} h_{in} - \frac{dm_{out}}{dt} h_{out} + Q + W \quad (3.2)$$

Transformando la ecuación de conservación de energía y considerando la dependencia de la energía interna de la medida de la temperatura se obtiene:

$$\frac{dU}{dt} = C_v \frac{d}{dt} (m * T) \quad (3.3)$$

$$\frac{dU}{dt} = C_v m \frac{dT}{dt} + C_v T \frac{dm}{dt} \quad (3.4)$$

De la ecuación 3.4 se obtiene la ecuación para la temperatura como variable de estado.

$$\frac{dT}{dt} = \frac{\frac{dU}{dt} - C_v T \frac{dm}{dt}}{C_v m} \quad (3.5)$$

Reemplazando de (3.1) y (3.2) se obtiene:

$$\frac{dT}{dt} = \frac{\frac{dm_{in}}{dt} h_{in} - \frac{dm_{out}}{dt} h_{out} + Q + W - C_v T (\frac{dm_{in}}{dt} - \frac{dm_{out}}{dt})}{C_v m} \quad (3.6)$$

La ecuación de los gases ideales $pV = mRT$ es usada para desarrollar una ecuación alternativa para la presión:

$$\frac{dp}{dt} = \frac{p}{m} \frac{dm}{dt} + \frac{p}{T} \frac{dT}{dt} \quad (3.7)$$

$$\frac{dp}{dt} = \frac{p}{m} \left(\frac{dm_{in}}{dt} - \frac{dm_{out}}{dt} \right) + \frac{p}{T} \left(\frac{dm_{in}}{dt} h_{in} - \frac{dm_{out}}{dt} h_{out} + Q + W - C_v T \left(\frac{dm_{in}}{dt} - \frac{dm_{out}}{dt} \right) \right) \quad (3.8)$$

Al realizar el balance de conservación de energía mecánica del eje turbina-compresor se tiene:

$$\frac{dE_{eje}}{dt} = \frac{dm_t}{dt} C_{pg} (T_{03} - T_{04}) \eta_{mec} - \frac{dm_c}{dt} C_{pa} (T_{02} - T_{01}) - 2\pi \frac{3}{50} n M_{load} \quad (3.9)$$

El primer y más importante requerimiento para la simulación de la turbina de gas es identificar los parámetros que definen la planta completamente. Estos parámetros son conocidos como variables de estado, los cuales en presencia de condiciones externas de flujo de combustible, carga y condiciones atmosféricas pueden establecer los puntos de operación de cada componente de la turbina.

El sistema puede ser representado completamente por las tres ecuaciones diferenciales (3.6), (3.8) y (3.9).

El vector de estados (\vec{X}) para el modelo es:

$$\vec{X} = (T_3, P_3, N)^T$$

La descripción de cada variable de estado se presenta en la tabla 3.5 a continuación:

Tabla 3. 5. Variables de estado del modelo de la turbina

Estado	Variable	Descripción
1	T_3	Temperatura Total de Combustión
2	P_3	Presión Total de Combustión
3	N	RPM del Eje de la Turbina

Entonces, las tres ecuaciones que definen las tres variables de estado son:

$$\frac{dP_3}{dt} = \frac{P_3}{T_3} \left(\frac{dT_3}{dt} \right) + \frac{RT_3}{V_{comb}} (m_c + m_f + m_t) \quad (3.10)$$

$$\frac{dT_3}{dt} = \frac{1}{C_v m} [(C_{pa} T_2 m_c + LHV \eta_{cc} m_f - C_{pg} T_4 m_t) - C_v T_3 (m_c + m_f - m_t)] \quad (3.11)$$

$$\frac{dN}{dt} = \left(\frac{30}{\pi} \right)^2 \frac{1}{\Theta * N} (\text{Trabajo}_{Turbina} - \text{Trabajo}_{Compresor} - 2\pi \frac{180}{50} N * M_{load}) \quad (3.12)$$

Dónde:

$$\text{Trabajo}_{\text{ Turbina}} = m_t C_{pg} (T_{03} - T_{04}) \eta_{mech} \quad (3.13)$$

$$\text{Trabajo}_{\text{ Compresor}} = m_c C_{pa} (T_{02} - T_{01}) \quad (3.14)$$

3.3.2. Relaciones Fundamentales.

Algunas relaciones fundamentales necesitadas para completar el modelo no lineal de la turbina son:

$$p_2 = \frac{p_3}{\sigma_{comb}} \quad (3.15)$$

$$p_4 = \frac{p_1}{\sigma_I \sigma_N} \quad (3.16)$$

La ecuación de los gases ideales es usada para la cámara de combustión como sigue:

$$T_3 = \frac{p_3 V_{comb}}{m_{comb} R_{med}} \quad (3.17)$$

La temperatura después del compresor es calculada usando la eficiencia isoentrópica η_c de la siguiente manera:

$$T_2 = T_1 \left(1 + \frac{1}{\eta_c} \left(\left(\frac{p_2}{p_1} \right)^{\frac{k_{aire}-1}{k_{aire}}} - 1 \right) \right) \quad (3.18)$$

La temperatura después de la turbina es calculada usando la eficiencia isoentrópica η_T de la siguiente manera:

$$T_4 = T_3 \left(1 - \eta_T \left(1 - \left(\frac{p_4}{p_3} \right)^{\frac{k_{aire}-1}{k_{aire}}} \right) \right) \quad (3.19)$$

Las siguientes ecuaciones constitutivas definen los flujos másicos y las eficiencias en el compresor y la turbina.

$$\dot{m}_c = const(1) q(\lambda_1) \frac{p_1}{\sqrt{T_1}} \quad (3.20)$$

$$m_T = \text{const}(2)q(\lambda_3) \frac{P_3}{\sqrt{T_3}} \quad (3.21)$$

Dónde:

$$q(\lambda_1) = f_1\left(\frac{n}{\sqrt{\frac{T_1}{288.15}}}, \frac{P_2}{P_1}\right) \quad (3.22)$$

$$q(\lambda_3) = f_2\left(\text{const}(3), \frac{n}{\sqrt{T_3}}, \frac{P_3}{P_4}\right) \quad (3.23)$$

$q(\lambda_1)$ es función de las revoluciones corregidas y de la relación de presión en el compresor, y $q(\lambda_3)$ es función de la velocidad a dimensional y de la relación de presión en la turbina.

Las ecuaciones de las eficiencias isoentrópicas del compresor y de la turbina pueden ser calculadas como siguen:

$$\eta_c = g_1\left(\frac{n}{\sqrt{\frac{T_1}{288.15}}}, q(\lambda_1)\right) \quad (3.24)$$

$$\eta_T = g_2\left(\text{const}(3) \frac{n}{\sqrt{T_3}}, \frac{P_3}{P_4}\right) \quad (3.25)$$

Los parámetros y las constantes de estas funciones pueden ser determinados con la ayuda de los resultados de mediciones realizadas de las características del compresor y la turbina.

El flujo másico adimensional del compresor se expresa como una función del número de revoluciones corregida y como función lineal de la relación de presión en el compresor.

$$q(\lambda_1) = a_1 \frac{n}{\sqrt{\frac{T_1}{288.15}}} \frac{P_2}{P_1} + a_2 \frac{n}{\sqrt{\frac{T_1}{288.15}}} + a_3 \frac{P_2}{P_1} + a_4 \quad (3.26)$$

La eficiencia isoentrópica del compresor es una función lineal del número de revoluciones corregida y una función lineal del flujo másico adimensional del compresor.

$$\eta_c = b_1 \frac{n}{\sqrt{\frac{T_1}{288.15}}} q(\lambda_1) + b_2 \frac{n}{\sqrt{\frac{T_1}{288.15}}} + b_3 q(\lambda_1) + b_4 \quad (3.27)$$

El flujo másico adimensional de la turbina es una función lineal de la velocidad adimensional y función lineal de la relación de presión en la turbina.

$$q(\lambda_3) = c_1 \left(\frac{n}{\sqrt{T_3}} \right) \frac{p_3}{p_4} + c_2 \frac{n}{\sqrt{T_3}} + c_3 \frac{p_3}{p_4} + c_4 \quad (3.28)$$

La eficiencia isoentrópica de la turbina es función lineal de la velocidad adimensional y función lineal de la relación de presión en la turbina.

$$\eta_T = d_1 \left(\frac{n}{\sqrt{T_3}} \right) \frac{p_3}{p_4} + d_2 \frac{n}{\sqrt{T_3}} + d_3 \frac{p_3}{p_4} + d_4 \quad (3.29)$$

Las constantes de estas funciones (a_i, b_i, c_i, d_i) se calculan con la ayuda del método de los mínimos cuadrados usando la herramienta de Matlab para regresiones múltiples (Ver Apéndice). Los datos usados para los cálculos de las constantes fueron obtenidos de manera experimental al poner en funcionamiento la turbina y tomando datos a 45.000, 50.000, 55.000, 60.000, 70.000 y 77.000 revoluciones por minuto y las respectivas presiones, temperaturas y flujo de combustible cada 0.5 segundos.

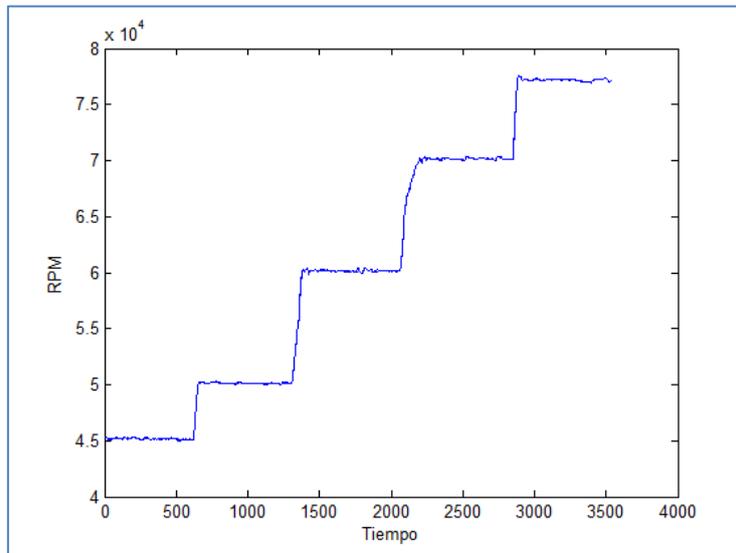


Figura 3. 8. RPM versus tiempo

La figura 3.8 muestra la velocidad medida del eje de la turbina para diferentes flujos de combustible. La figura 3.9 muestra las relaciones de presión calculadas para las diferentes revoluciones, las figuras 3.10 y 3.11 muestran la eficiencia del compresor y relaciones de presión de la turbina calculadas respectivamente para las diferentes revoluciones.

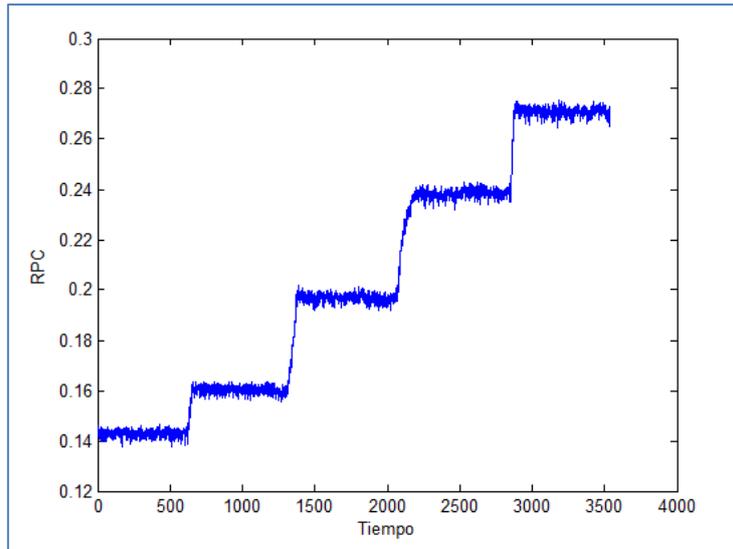


Figura 3. 9. Relación de Presiones en el compresor a diferentes RPM

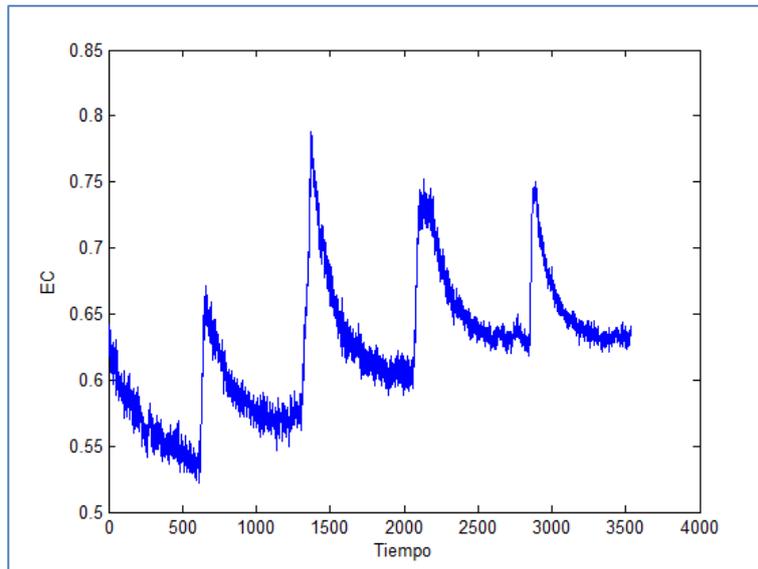


Figura 3. 10. Eficiencia en el compresor a diferentes RPM

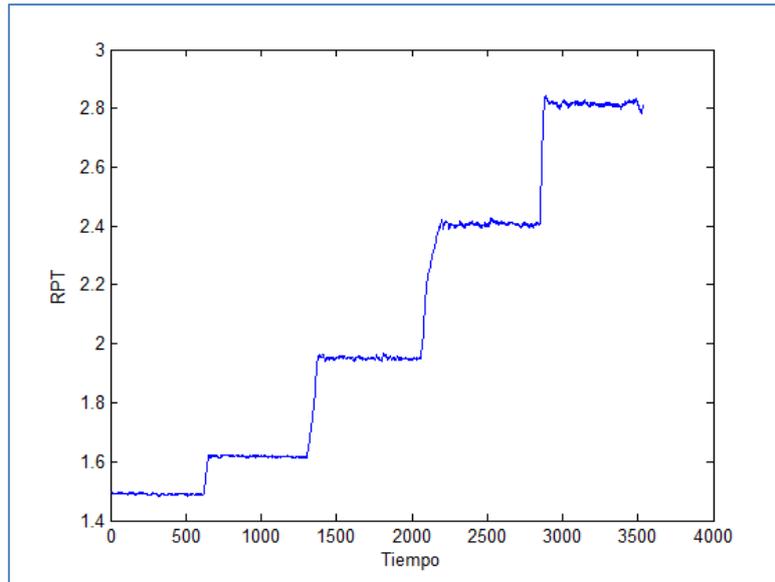


Figura 3. 11. Relación de Presiones en la Turbina a diferentes RPM

La tabla 3.6 muestra las constantes calculadas usando las ecuaciones (3.26-3.29).

Las figuras 3.12 a la 3.15 muestran las regresiones múltiples realizadas para los diferentes parámetros de eficiencia y flujo másico en el compresor y la turbina.

Tabla 3. 6. Constantes calculadas para los parámetros de la turbina

i	a_i	b_i	c_i	d_i
1	-0.0134 ± 0.0025	0.0120 ± 0.0018	0.0020 ± 0.08	1.9877 ± 0.03
2	0	0	0	0
3	0.0285 ± 0.025	3.4835 ± 0.23	0.0137 ± 0.0021	-44.5889 ± 1.2
4	0	0	0	0.0002 ± 0.0001

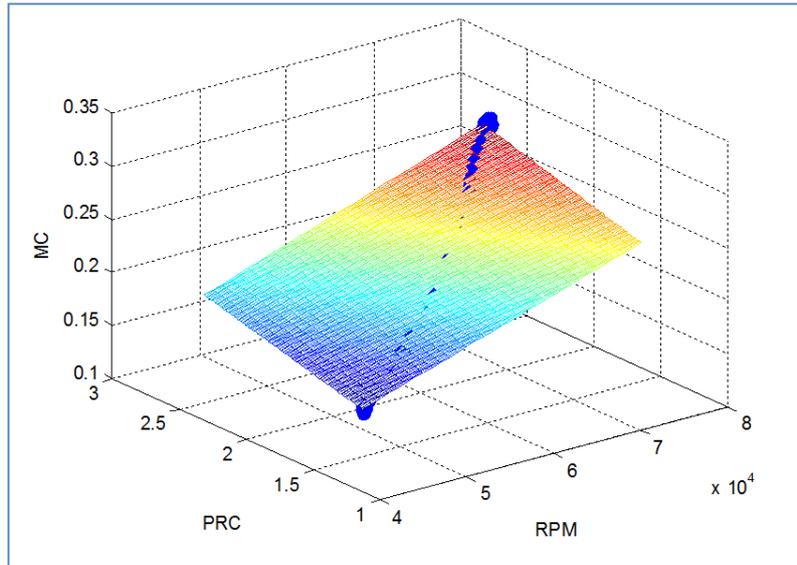


Figura 3. 12. Regresión múltiple entre la masa de aire, la relación de presión en el compresor y rpm.

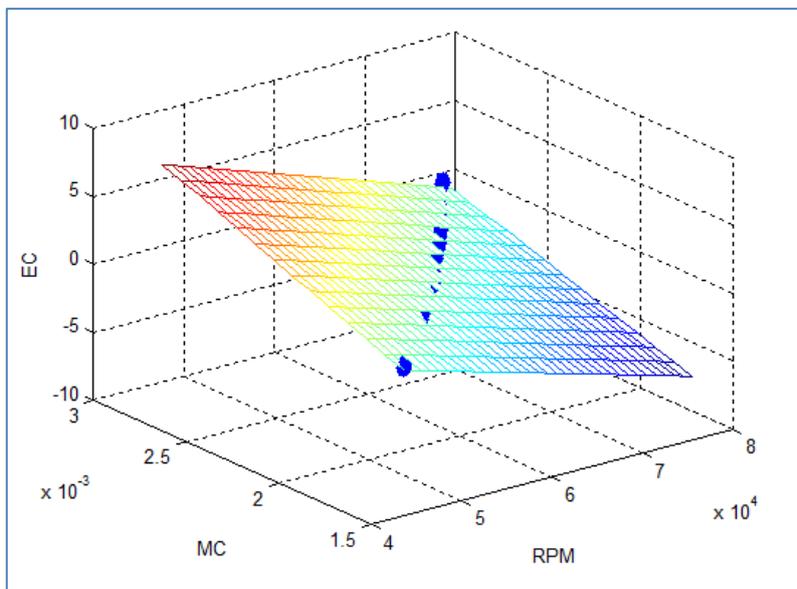


Figura 3. 13. Regresión múltiple entre la masa de aire, la relación de presión del compresor y rpm.

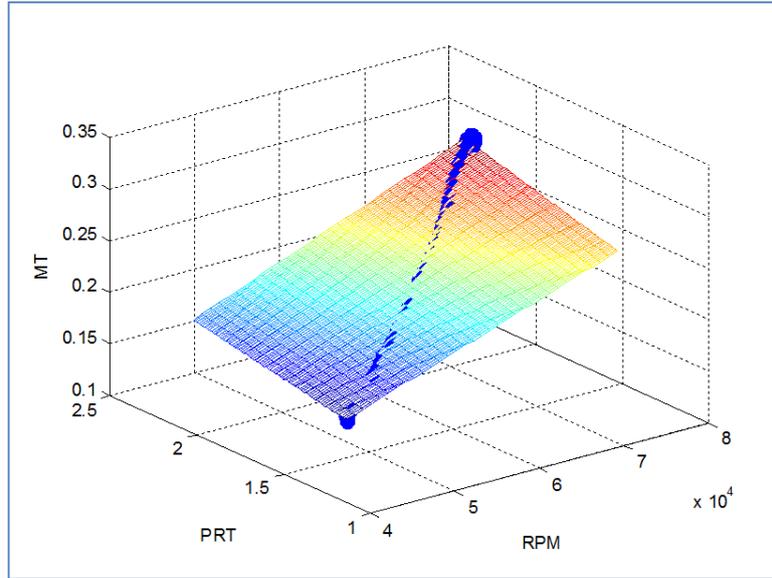


Figura 3. 14. Regresión múltiple entre la masa de combustión, relación de presión en la turbina y rpm.

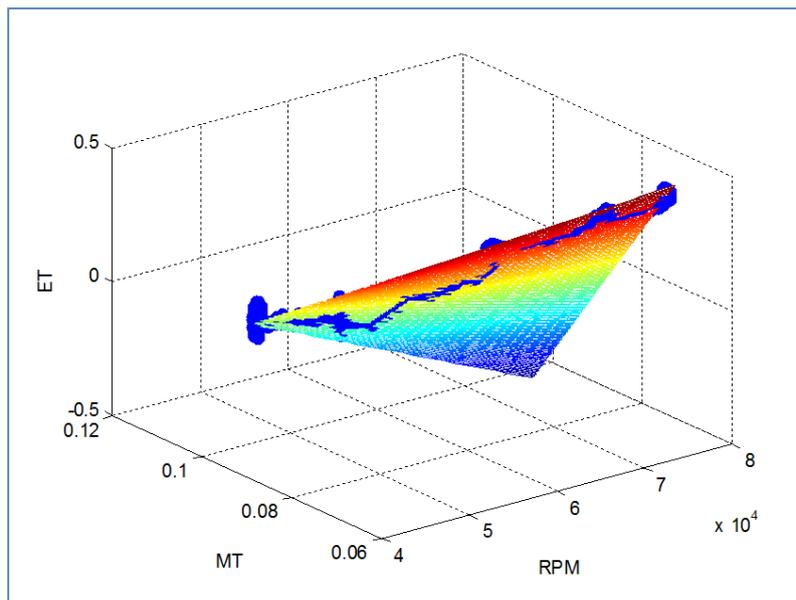


Figura 3. 15. Regresión múltiple entre la masa de combustible, la velocidad del rotor y la eficiencia de la turbina

3.4. Modelo en Simulink de la Turbina de Gas SR-30

Un modelo en Simulink fue desarrollado para la turbina de gas SR-30 según se muestra en la figura 3.16. Para la programación de cada bloque se usaron las ecuaciones presentadas en la sección 3.3.

El modelo desarrollado tiene como entrada la variable de flujo de combustible y las variables de salida son las diferentes temperaturas y presiones en cada una de las etapas de la turbina, adicional la

velocidad del eje y el torque. Este modelo en Simulink usó como referencia el modelo desarrollado en [Kim-00] y [Ail-01].

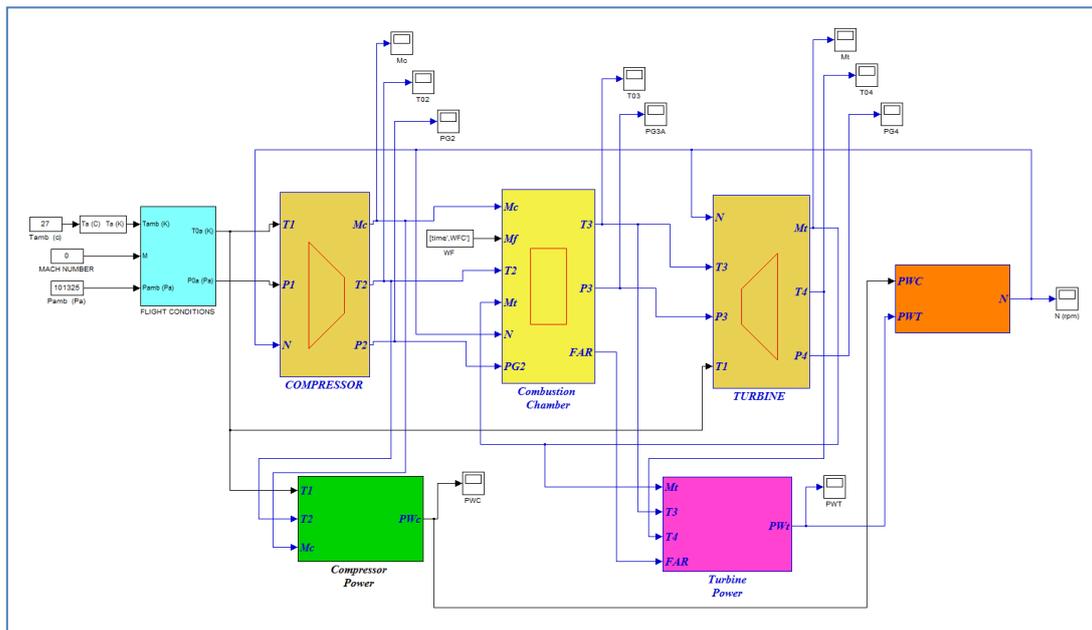


Figura 3. 16. Modelo en Simulink para la turbina de gas SR30

La figura 3.17 muestra el diagrama de bloque realizado para la simulación del compresor donde las variables de entrada son la temperatura de entrada del compresor, la velocidad de rotación del eje y las variables de salida con la temperatura de salida del compresor, la eficiencia del compresor, la presión de salida del compresor y el flujo másico de aire a través del compresor.

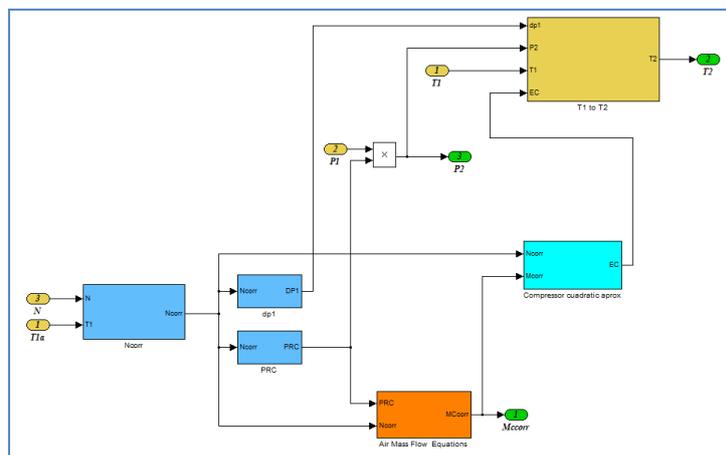


Figura 3. 17. Diagrama en Simulink del compresor

La figura 3.18 muestra el diagrama de bloques realizado para la simulación de la turbina donde las variables de entrada son la temperatura de entrada a la turbina, la presión de entrada a la turbina, la velocidad de rotación del eje, la temperatura ambiente y las variables de salida son la temperatura de salida de la turbina, la eficiencia de la turbina y el flujo másico de aire y combustible a través de la turbina.

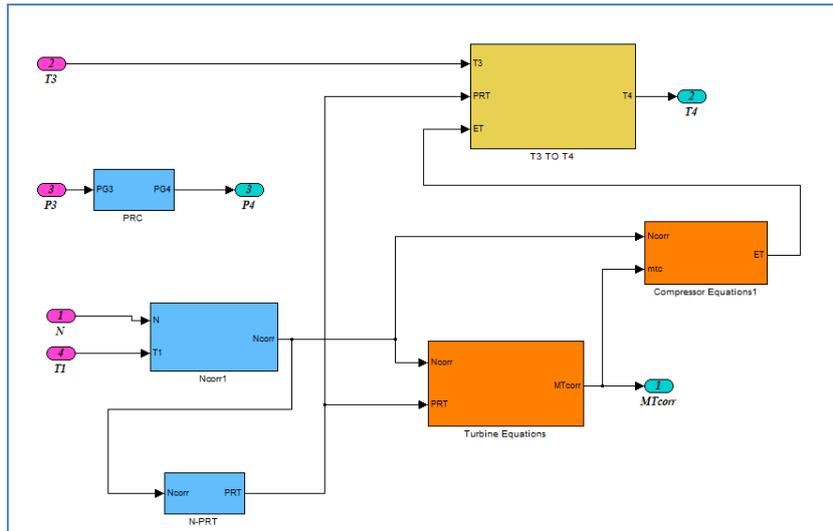


Figura 3. 18. Diagrama en Simulink de la turbina

La figura 3.19 muestra el diagrama de bloques realizado para la simulación de la cámara de combustión donde las variables de entrada son el flujo másico de aire, flujo másico de combustible, el flujo másico de aire y combustible para la combustión, la velocidad del eje y las variables de salida son la eficiencia de la combustión, la temperatura de combustión y la presión.

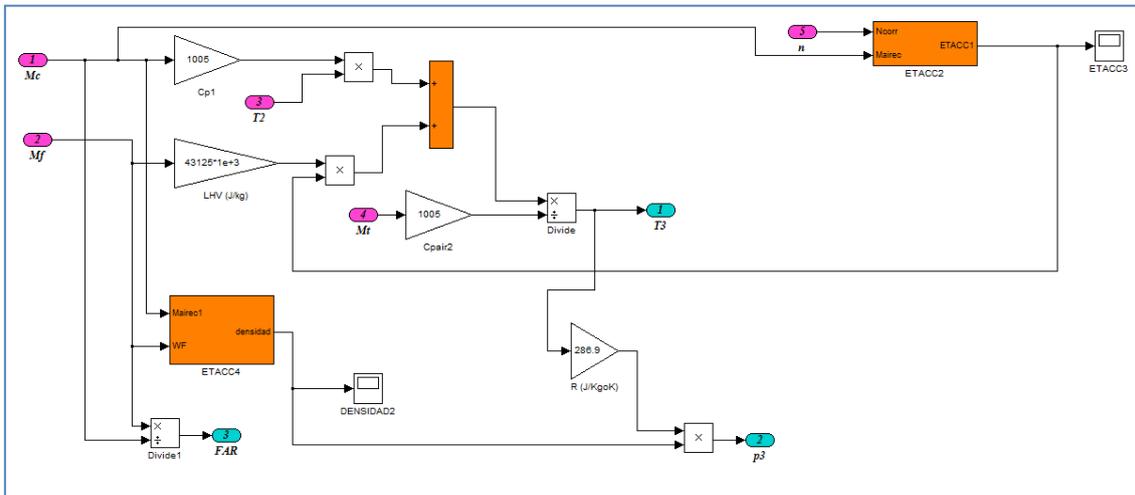


Figura 3. 19. Diagrama en Simulink de la cámara de combustión.

El diagrama presentado en la figura 3.20 representa el balance de energía realizado en la turbina, donde se relaciona la potencia del compresor (figura 3.21), la potencia de la turbina (figura 3.22) y las pérdidas mecánicas para poder determinar el cambio en la velocidad del eje.

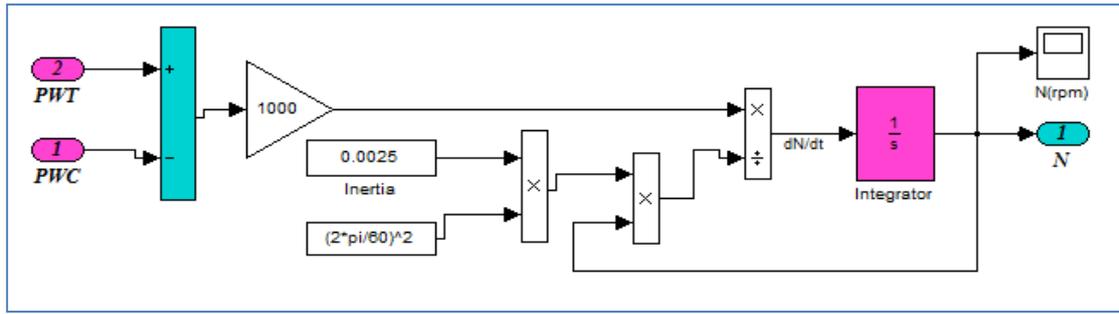


Figura 3. 20. Diagrama en Simulink para obtener la revoluciones en el eje

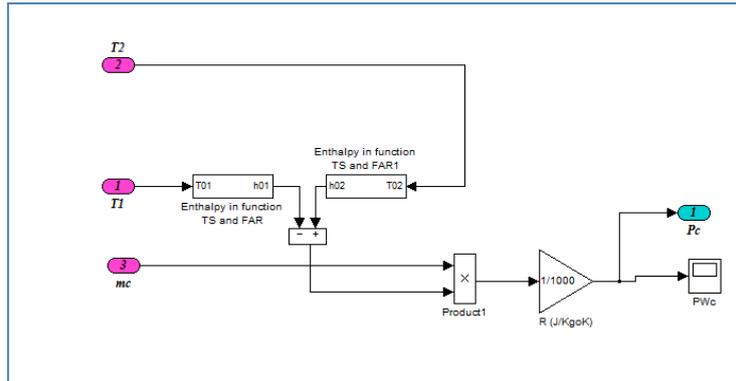


Figura 3. 21. Diagrama en Simulink para obtener la potencia del compresor.

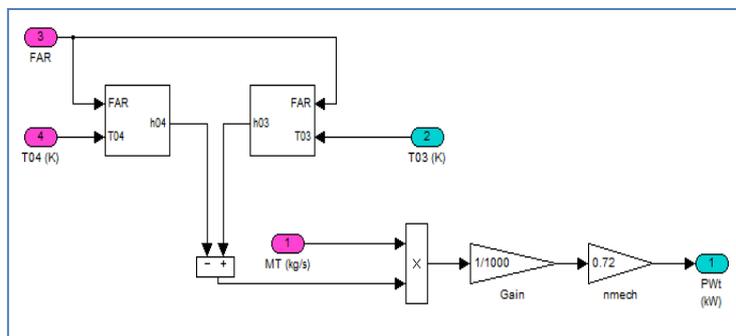


Figura 3. 22. Diagrama en Simulink para obtener la potencia en la turbina

3.4.1. Validación del Modelo

Para validar el modelo matemático se realizaron varias experimentaciones en la turbina real para determinar los parámetros experimentales y de ajuste en las ecuaciones planteadas.

La figura 3.23, 3.24, 3.25 y 3.26 muestran el flujo de combustible, las revoluciones del eje de la turbinas, las temperaturas y las presiones del comportamiento real versus el simulado para los experimentos realizados en la turbina SR-30.

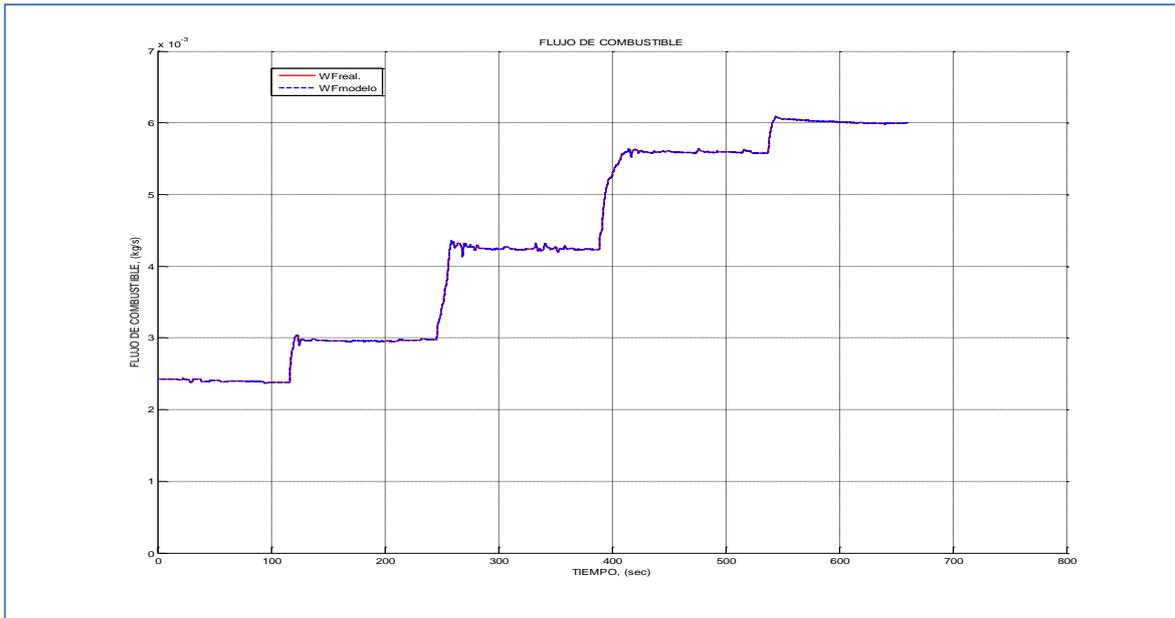


Figura 3. 23. Flujo de combustible en su comportamiento real versus simulado

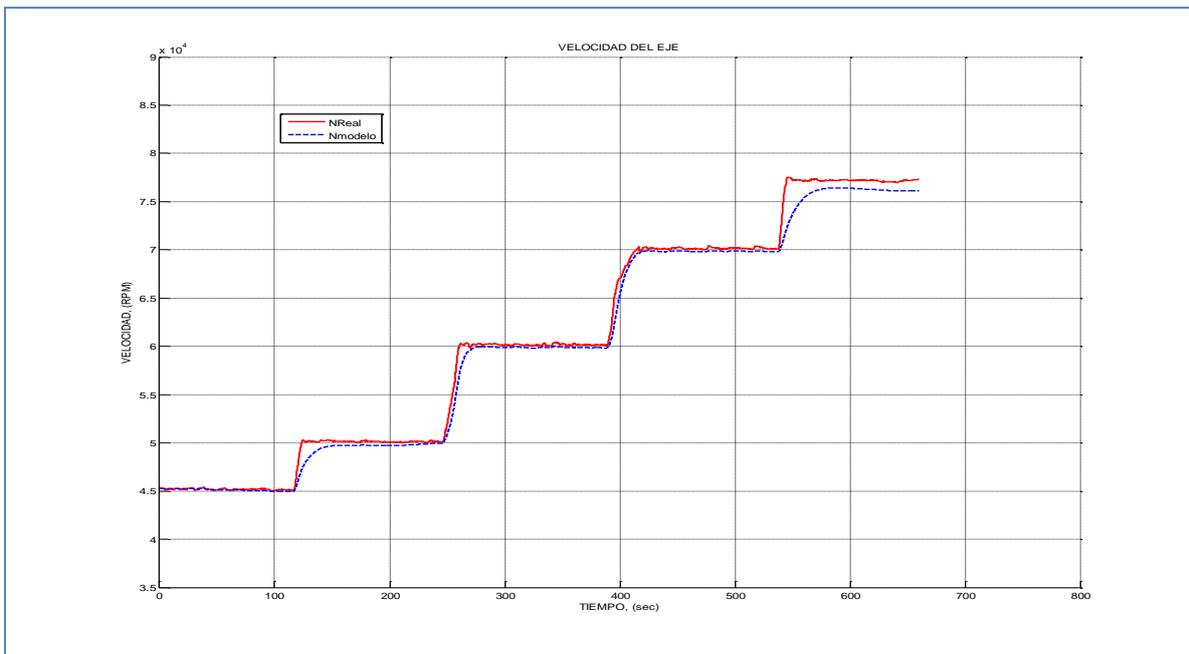


Figura 3. 24. Velocidad del eje de la turbina en su comportamiento real versus simulado

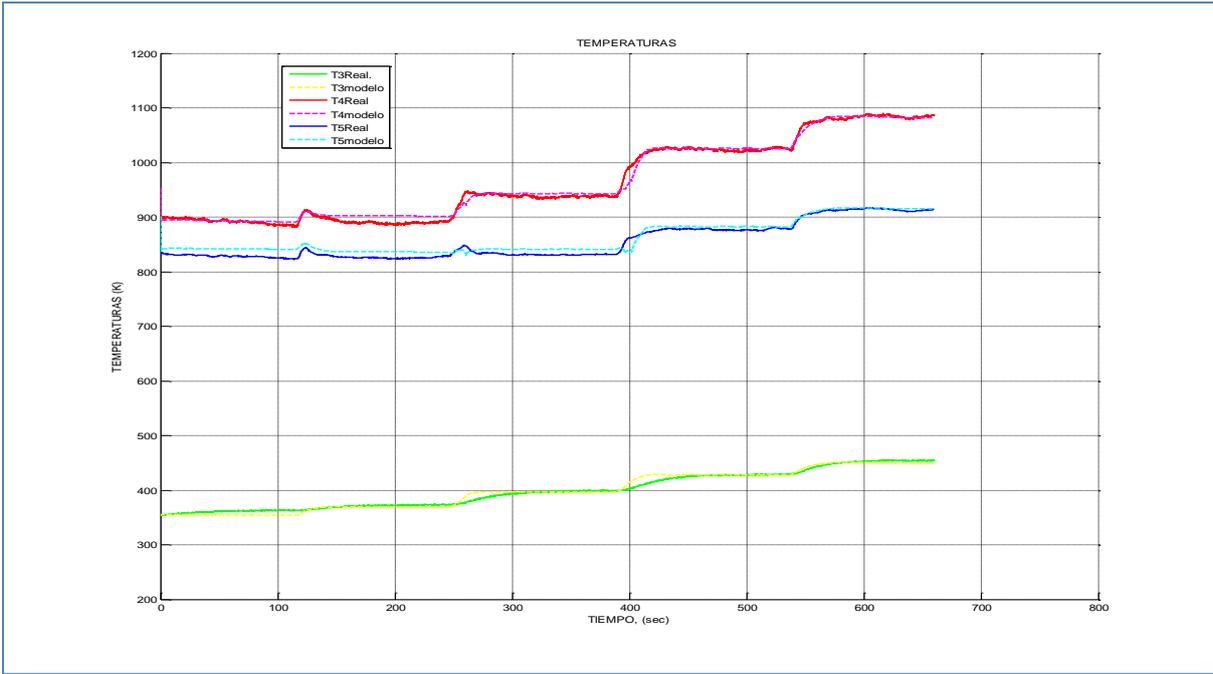


Figura 3. 25. Temperaturas en su comportamiento real versus simulado

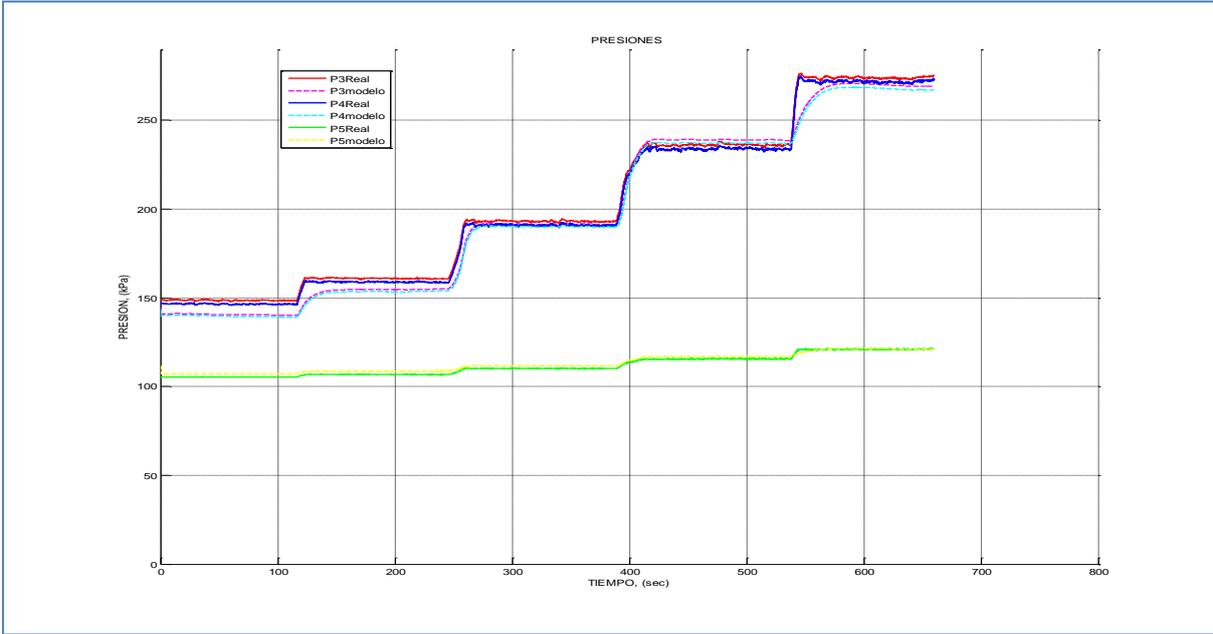


Figura 3. 26. Presiones en su comportamiento real versus simulado

Los datos reales versus simulados muestran un ajuste de menos de 5% de error, lo cual implica que para efectos de este estudio el modelo podría representar una realidad muy cercana cuando se deseen simular fallos en componentes y sensores.

3.5. Fallos Estudiados en la Turbina de Gas

El estudio de los posibles fallos simples se llevó a cabo en los sensores de acuerdo con la tabla 3.7 y en los componentes físicos de acuerdo con la tabla 3.8. Los sensores que se incluyen en la tabla 3.7 corresponden a los sensores que trae de fábrica la turbina Minilab SR-30. Luego de realizar el análisis termodinámico de la turbina en la cual se realizó la fase experimental se determinó que los sensores utilizados son los mínimos necesarios para poder obtener toda caracterización y desempeño de la turbina.

Tabla 3. 7. Fallos estudiados en los sensores de la turbina de gas

Fallo	Tipo de Fallo
F1	Sensor de Temperatura Estática de entrada del Compresor (T01)
F2	Sensor de Temperatura de Estancamiento de salida de la etapa de compresor (T02)
F3	Sensor de Temperatura del Estancamiento de entrada de la etapa de la turbina (T03)
F4	Sensor de Temperatura de Estancamiento de salida de la etapa de la turbina (T04)
F5	Sensor de Temperatura de Estancamiento de salida de la tobera (T05)
F6	Sensor de Presión estática de entrada del compresor (P01)
F7	Sensor de Presión de estancamiento de salida de la etapa de compresor (P02)
F8	Sensor de Presión de la cámara de combustión (P03)
F9	Sensor de Presión de estancamiento de salida de la turbina (P04)
F10	Sensor de Presión de estancamiento de la salida de la tobera (P05)
F11	Sensor de Flujo de combustible (Q1)
F12	Sensor de Revoluciones del eje de la turbina. (RPM)

La turbina de gas SR-30 tiene tres componentes principales: El compresor, La turbina, y la válvula de entrada de combustible. De acuerdo con los estudios realizados estos componentes son lo que habitualmente presentan fallos [Leo-08][Kim-00][Wat-04]. Fallos menos significativos pueden ocurrir pero se reflejan siempre en los componentes principales [Per-03][Laz-01]. La tabla 3.8 presenta los fallos habituales en la turbina SR-30.

Tabla 3.8 Fallos Estudiados en los Componentes de la Turbina de Gas

Fallo	Tipo de Fallo
F13	Obstrucción de la masa de aire a través del compresor (MC)
F14	Disminución en la Eficiencia en el compresor (EC)
F15	Obstrucción de la masa de aire y combustible a través de la turbina (MT)
F16	Disminución en la Eficiencia en la turbina (ET)
F17	Obstrucción en la válvula de Combustible (WF)

Los fallos en los sensores fueron generados mediante la adición computacional de una cantidad a la señal real. Los fallos en los actuadores fueron simulados en fallos limitando el rango de salida de la tarjeta de adquisición de datos. Los posibles fallos múltiples resultan de la combinación de los fallos presentados en la tabla 3.8.

En este estudio no se analizan fallos a nivel de elementos mecánicos (compresor, turbina, cámara de combustión, etc.) debido a las implicaciones de seguridad que este tipo de experimentos amerita. Por la razón anterior, un modelo fue usado para poder simular fallos en componentes y luego aplicar los diferentes algoritmos de detección y diagnóstico de fallos simples y múltiples.

En los capítulos 4, 5 y 6 se detallan los fallos múltiples estudiados. Básicamente es la combinación de los fallos presentados en la tabla 3.7 y 3.8.

Las figuras 3.27 y 3.28 muestran la implementación realizada en Simulink para representar fallos abruptos y fallos graduales en los sensores y en los componentes de la turbina.

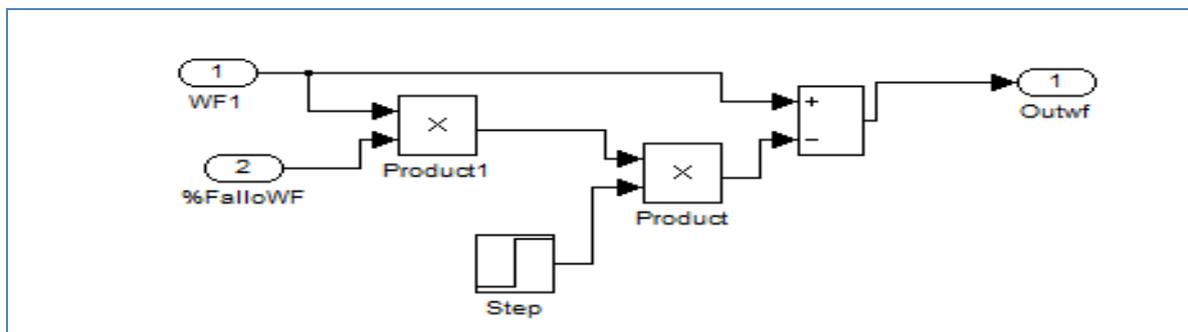


Figura 3. 27. Representación de fallos abruptos

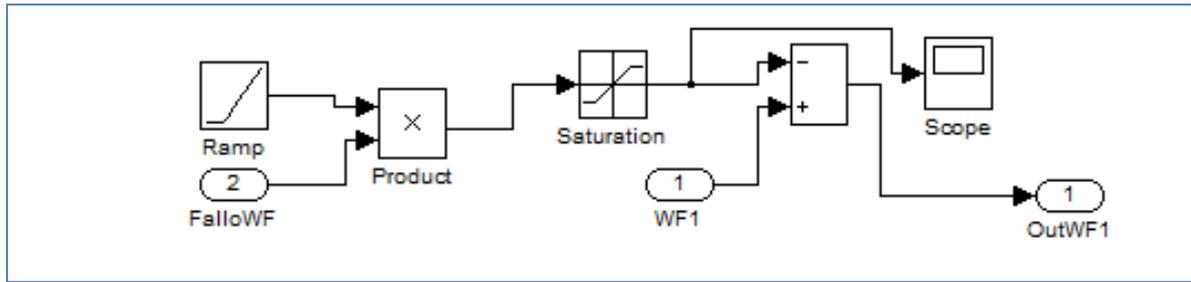


Figura 3. 28. Representación de fallos graduales

3.6. Trabajo experimental

Las turbinas de gas por su complejidad mecánica son altamente sensibles a desbalance mecánico y altamente peligrosas si las condiciones de operación son las no adecuadas. En esta tesis se usaron los datos obtenidos de los sensores ubicados en una turbina real SR30 para poder calibrar y ajustar el modelo teórico. Las ecuaciones de balance de energía, masa y fuerzas fueron usadas para obtener un modelo bastante simple.

La figura 3.29 muestra los alabes de una turbina SR30 a los cuales se le modificó levemente uno de ellos para poder observar la variación en la eficiencia mecánica. Para poder realizar los experimentos se usaron las normas de seguridad y un cuarto especialmente preparado para ello en la compañía Turbine Technologies, USA.



Figura 3. 29. Muestra extraída de un alabe de la turbina para observar el comportamiento de la eficiencia mecánica.

La figura 3.30 muestra el diseño experimental realizado a la turbina para poder desbalancear el rotor de la turbina hasta un rango de operación seguro.



Figura 3. 30. Desbalanceo del rotor de la turbina SR30

3.6. Sumario

En este capítulo se presentó la descripción del sistema MiniLab Gas Turbine, la teoría de operación y las ecuaciones matemáticas que caracterizan a una turbina de Gas SR-30.

El desarrollo de las pruebas experimentales ayudó al refinamiento del modelo teórico y permitió que se usara el modelo para simular fallas en componentes que de otra manera no se podía obtener.

En este estudio no se analizan fallos a nivel de elementos mecánicos debido a las implicaciones de seguridad que este tipo de experimentos amerita. Por la razón anterior, un modelo en Simulink se desarrolló usando Simulink y poder simular fallos en componentes y luego aplicar los diferentes algoritmos de detección y diagnóstico de fallos simples y múltiples.

3.7. Bibliografía

- [Ail-01] Ailer, P. y Santa, I. (2001). Nonlinear Model Building of a Low Power Gas. Periódica Polytechnica SER. Transp. eng. vol. 29, No. 1-2, pp. 117–135.
- [Cam-07] Camacho, O. (2007). Fault Diagnostic base on multivariate statistical techniques. Revista técnica de la Universidad de Zulia. Vol. 30, No 3, pp. 253-262.
- [Fre-03] French, K. (2003). Recycled Fuel Performance in the SR-30 Gas Turbine. Proceedings of the 2003 American Society for Engineering Education Annual Conference and Exposition.
- [Kim-00] Kim, S., Pilidis, P. y Yin, J. (2000). Gas Turbine Dynamic simulation using simulink”. Department of propulsion Engineering and power. Cranfield University, United Kingdom. 2000-01-3647. Society of Automotive Engineers.
- [Laz-01] Lazzaretto, A. y Toffolo, A. (2001). Analytical and Neural Network Model for Gas Turbine Design and Off-Design simulation. Int. J. Applied Thermodynamics. Vol. 4 (No. 4) pp173-182, December.
- [Leo-08] Leonard, O., Thomas, J. y Borguet, S. (2008). Ten Years of Experience with a Small Jet Engine as a Support for Education. Proceedings of ASME Turbo Expo 2008: Power for Land, Sea and Air GT2008, June 9-13, Berlin, Germany.

- [Per-03] Perez-Blanco, H. (2003). Activities around the SR-30 Minilab at PSU. Proceeding of the 2003 American Society for Engineering Education Annual Conference and Exposition copyright. American Society for engineering education.
- [Rah-11] Rahman, M., Ibrahim, T. y Abdalla, A. (2011). Thermodynamic performance analysis of gas-turbine power-plant. International Journal of the Physical Sciences Vol. 6(14), pp. 3539-3550, 18 July.
- [Rin-12] Rincón, A. (2012). Multiple fault detection and diagnosis in a Gas Turbine using principal component analysis and structured residuals. 20th Mediterranean Conference on Control & Automation (MED 2012).
- [Rin-13] Rincón, A. Quevedo, J. (2013). Multiple Fault Detection and Diagnosis in Gas Turbine Using Nonlinear Principal Component Analysis and Structured Residuals. ASME 2013 International Mechanical Engineering Congress & Exposition.
- [Sar-01] Saravanamuttoo, H. (2001). Gas Turbine Theory. Fifth Ed. Pearson Prentice Hall.
- [Wat-04] Watanabe, A., Ölçmen, S., Leland, R. y Whitaker K. (2004). Soft Computing Applications on SR-30 Turbojet Engine. 1st AIAA Intelligent Systems AIAA 2004-6444, Technical Conference Chicago, IN, September 20-22.
- [Wit-03] Witkowski, T., White, S. y Ortiz, C. (2003). Characterizing the performance of the SR30 Turbojet Engine. Proceeding of the 2003 American Society For Engineering education Annual Conference and exposition.

CAPITULO 4:

DETECCION DE FALLOS MULTIPLES EN UNA TURBINA DE GAS USANDO ANALISIS DE COMPONENTES PRINCIPALES (PCA)

Contenido

- 4.1. Introducción
- 4.2. Fundamentos del análisis de componentes principales
- 4.3. Modelado Implícito
- 4.4. Detección de Fallos múltiples basado en PCA
- 4.5. Algoritmo para la detección de fallos múltiples mediante PCA
- 4.6. Resultados
- 4.7. Conclusiones
- 4.8. Bibliografía

4.1. Introducción

Los procesos industriales son sistemas complejos y altamente instrumentados, en los cuales se maneja un número considerable de variables asociadas ya sea al control o para fines de monitoreo. Un enfoque convencional de monitoreo es el univariado, donde para cada variable se define un umbral de condición normal. Así, la detección consiste en evaluar cada observación con respecto al umbral, activándose una alarma cuando una observación se sale de los límites establecidos. Sin embargo, el enfoque univariado no toma en cuenta la correlación entre las variables (correlación cruzada o paralela) y la correlación entre mediciones a diferentes tiempos de una misma variable (autocorrelación o correlación en serie), lo cual genera insensibilidad y falsas alarmas a ciertas condiciones de fallo [Mas-02].

La limitante de las gráficas de monitoreo univariado para detección de fallos ha llevado al desarrollo de enfoques basados en análisis estadístico multivariado. En este sentido, dada la cantidad de variables disponibles, el número de correlaciones puede ser muy elevado, por lo tanto, se sugiere transformar las variables a un número tal que se conserve la mayor variabilidad de los datos originales [Peñ-02].

El análisis de componentes principales (PCA) permite de manera óptima, transformar un conjunto de datos multivariados a un conjunto ortogonal de menor dimensión, esto en términos de la captura de la varianza de los datos y tomando en cuenta la correlación entre variables, lo cual constituye una forma de modelado implícito. Cabe destacar que a pesar de las ventajas del modelado basado en PCA para monitorear procesos complejos, este está limitado a la tarea de detección de fallas debido a la ausencia de procedimientos para diagnosticar fallos [Min-06].

En este capítulo se revisan los fundamentos del análisis de componentes principales (PCA); se describe el procedimiento de modelado basado en PCA, se presenta la extensión de PCA al caso de fallos múltiples y finalmente se presenta la implementación de la metodología al caso de una turbina de gas SR-30. En el capítulo 5 se presenta una alternativa para el diagnóstico usando análisis de componentes principales no lineales.

4.2. Fundamentos de PCA

El análisis de componentes principales tiene como objetivo: dado n observaciones de p variables, analizar si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales. La técnica de componentes principales es debida a Hotelling en 1933, aunque sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados introducidos por K. Pearson en 1901.

En las herramientas de análisis estadístico multivariado como lo es el análisis de componentes principales (PCA), se consideran las siguientes suposiciones.

Suposición 1: Las mediciones obtenidas en cada instante de muestreo son observaciones serialmente independientes.

Suposición 2: Las observaciones asociadas a las variables tienen una distribución normal.

Supongamos que se dispone de los valores de p -variables en n elementos de una población dispuestos en una matriz \mathbf{X} de dimensión $n \times p$, donde las columnas contienen las variables y las filas los elementos. El problema que se desea resolver es encontrar un espacio de dimensión más reducida que represente adecuadamente los datos. Puede abordarse desde tres perspectivas equivalentes.

a) Enfoque descriptivo

Bajo este enfoque la idea es encontrar un subespacio de dimensión menor que p tal que al proyectar sobre él los puntos conserven su estructura con la menor distorsión posible. Considerando el caso de dos dimensiones $p = 2$, la figura 4.1 indica el diagrama de dispersión y una recta que, intuitivamente, proporciona un buen resumen de los datos, ya que la recta pasa cerca de todos los puntos y las distancias entre ellos se mantienen aproximadamente en su proyección sobre la recta.

La condición de que la recta pase cerca de la mayoría de los puntos puede concretarse exigiendo que las distancias entre los puntos originales y sus proyecciones sobre la recta sean lo más pequeñas posibles. En consecuencia, si se considera un punto x_i y una dirección $a_1 = (a_{11}, \dots, a_{1p})'$, definida por un vector a_1 de norma unidad, la proyección del punto x_i sobre esta dirección es el escalar:

$$z_i = a_{11}x_{i1} + \dots + a_{1p}x_{ip} = a_1'x_i \quad (4.1)$$

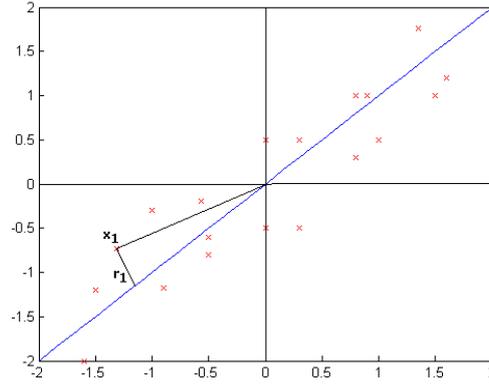


Figura 4. 1. Recta que minimiza las distancias ortogonales de los puntos

Y el vector que representa esta proyección será $z_i a_i$, este criterio implica:

$$\text{Minimizar } \sum_{i=1}^n r_i^2 = \sum_{i=1}^n |x_i - z_i a_i|^2, \quad (4.2)$$

Donde $|u|$ es la norma Euclídea o módulo del vector u .

La figura 4.1 muestra que al proyectar cada punto sobre la recta se forma un triángulo rectángulo donde la hipotenusa es la distancia del punto al origen, $(x_i' x_i)^{1/2}$, y los catetos la proyección del punto sobre la recta (z_i) y la distancia entre el punto y su proyección (r_i). Por el teorema de Pitágoras se puede escribir:

$$x_i' x_i = z_i^2 + r_i^2 \quad (4.3)$$

Y sumando esta expresión para todos los puntos, se obtiene:

$$\sum_{i=1}^n x_i' x_i = \sum_{i=1}^n z_i^2 + \sum_{i=1}^n r_i^2 \quad (4.4)$$

Como el primer miembro es constante, minimizar la suma de las distancias a la recta de todos los puntos, es equivalente a maximizar la suma al cuadrado de los valores de las proyecciones y por lo tanto su varianza.

b) Enfoque estadístico

En este enfoque se busca representar puntos p dimensionales con la mínima pérdida de información en un espacio de dimensión uno, lo cual es equivalente a sustituir las p variables originales por una nueva variable, z_i , que resuma óptimamente la información. Esto supone que las nuevas variables

deben tener globalmente máxima correlación con las originales o, en otros términos debe permitir proveer las variables originales con la máxima precisión.

c) Enfoque geométrico

En este enfoque el problema puede abordarse desde el punto de vista geométrico con el mismo resultado final. Si se considera que la nube de puntos que se muestran en la figura 4.2 se sitúa siguiendo una elipse y podemos describirlos por su proyección en la dirección del eje mayor de la elipse. Puede demostrarse que el eje es la recta que minimiza las distancias ortogonales. En varias dimensiones tendremos elipsoides, y la mejor aproximación a los datos es la proporcionada por su proyección sobre el eje mayor del elipsoide. Intuitivamente la mejor aproximación en dos dimensiones es la proyección sobre el plano de los dos ejes mayores del elipsoide y así sucesivamente. Considerar los ejes del elipsoide como nuevas variables originales supone pasar de variables correlacionadas a variables ortogonales o no correlacionadas.

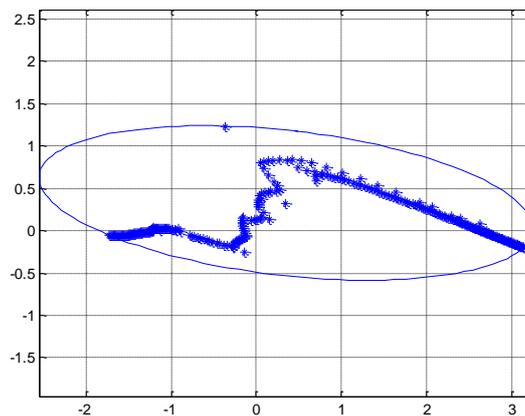


Figura 4. 2. Elipse que envuelve las distancias ortogonales de los puntos

4.2.1. Cálculo de los componentes

Sea un conjunto de n observaciones de p variables x_i de un proceso operando en condiciones nominales, las cuales se ordenan en la siguiente matriz de datos:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (4.5)$$

Asumiendo que \mathbf{X} satisface las Suposiciones 1-2, el análisis de componentes principales puede transformar este conjunto n-dimensional, en un espacio n-dimensional de variables no correlacionadas con varianzas maximizadas.

La primera componente en el espacio ortogonal se define como:

$$p_1 = Xq_1 \quad (4.6)$$

Con varianza

$$\gamma_{p_1}^2 = q_1^T \Gamma q_1 \quad (4.7)$$

Dónde: Γ es la matriz de covarianza de \mathbf{X} , y $q_1^T = [q_{11} \ q_{21} \ \dots \ q_{n1}]$. El problema consiste en encontrar el vector q_1 que maximice $\gamma_{p_1}^2$. Además se impone la restricción $q_1^T q_1 = 1$, es decir, q_1 debe ser ortonormal.

Por lo tanto, el problema de determinar q_1 se puede plantear en términos de maximizar la siguiente función:

$$\max v_1 = q_1^T \Gamma q_1 - \lambda_1 (q_1^T q_1 - 1) \quad (4.8)$$

Donde λ_1 es un multiplicador de LaGrange. Por lo tanto:

$$\frac{\partial v_1}{\partial q_1} = 2\Gamma q_1 - 2\lambda_1 q_1 = 0 \quad (4.9)$$

Es una condición necesaria de la solución, que puede escribirse como:

$$(\Gamma - \lambda_1 I_n) q_1 = 0 \quad (4.10)$$

O equivalentemente

$$\Gamma q_1 = \lambda_1 q_1 \quad (4.11)$$

Considerando que (4.10) es un sistema homogéneo de n ecuaciones con n incógnitas, este tiene una solución diferente de la trivial sí y sólo sí $|\Gamma - \lambda_1 I_n| = 0$. Por lo tanto, el multiplicador de LaGrange λ_1 es un valor propio de Γ , y la solución q_1 es su correspondiente vector propio normalizado. Dado que $\Gamma \in R^{n \times n}$ tiene n vectores propios, es necesario determinar el que maximice la varianza de p_1 .

Pre multiplicando (4.11) por q_1^T se obtiene:

$$\begin{aligned} q_1^T \Gamma q_1 &= \lambda_1 q_1^T q_1 \\ q_1^T \Gamma q_1 &= \lambda_1 = \gamma_{p1}^2 \end{aligned} \quad (4.12)$$

Es decir, λ_1 es la varianza de p_1 . Por lo tanto, dado que se busca la máxima varianza entonces q_1 es el vector propio asociado al valor propio más grande de Γ .

De manera similar a la primera componente, la segunda componente en el espacio ortogonal está dada por la siguiente combinación lineal de los datos:

$$p_2 = Xq_2 \quad (4.13)$$

Con varianza

$$\gamma_{p2}^2 = q_2^T \Gamma q_2 \quad (4.14)$$

Dónde: $q_2^T = [q_{12} \ q_{22} \ \dots \ q_{n2}]$. El problema consiste en encontrar el vector q_2 que maximice γ_{p2}^2 , con $q_2^T q_2 = 1$ y la restricción adicional $q_2^T q_1 = q_1^T q_2 = 0$, es decir, q_2 debe ser ortogonal a q_1 .

En este caso el problema de optimización, dada las dos restricciones, requiere de dos multiplicadores de LaGrange, y se busca maximizar.

$$\max v_2 = q_2^T \Gamma q_2 - \lambda_2 (q_2^T q_2 - 1) - \Psi (q_2^T q_1) \quad (4.15)$$

Dónde: λ_2 y Ψ son los multiplicadores de LaGrange. Derivando v_2 con respecto a q_2 e igualando a cero resulta:

$$\frac{\partial v_2}{\partial q_2} = 2\Gamma q_2 - 2\lambda_2 q_2 - \Psi q_1 = 0 \quad (4.16)$$

La cual puede reescribirse como:

$$2(\Gamma - \lambda_2 I_n)q_2 - \Psi q_1 = 0 \quad (4.17)$$

O equivalentemente

$$(\Gamma - \lambda_2 I_n)q_2 = \frac{1}{2} \Psi q_1 \quad (4.18)$$

Pre multiplicando (4.18) por q_1^T resulta en:

$$q_1^T \Gamma q_2 = \frac{1}{2} \Psi = q_2^T \Gamma q_1 \quad (4.19)$$

La igualdad en (4.19) es válida ya que la transpuesta de un escalar es el mismo escalar y Γ es una matriz simétrica. Pre multiplicando (4.11) por q_2^T se obtiene:

$$q_2^T \Gamma q_1 = \lambda_1 q_2^T q_1 = 0 \quad (4.20)$$

Lo que significa que $\Psi = 0$, y por lo tanto, (4.18) resulta en:

$$(\Gamma - \lambda_2 I_n) q_2 = 0 \quad (4.21)$$

Así, de manera similar al análisis de (4.10) para la primera componente, de (4.21) se deduce que:

$$\lambda_2 = q_2^T \Gamma q_2 = \gamma_{p_2}^2 \quad (4.22)$$

Por lo tanto, q_2 es el vector propio normalizado asociado al segundo valor propio más grande de Γ .

El análisis anterior se puede llevar a cabo para cada una de las n componentes del espacio ortogonal. Así, para la j -ésima componente definida por la combinación lineal $p_j = Zq_j$, se busca el vector q_j que maximice la varianza de la componente p_j , que satisfaga $q_j^T q_j = 1$ y sea ortogonal a todos los previos vectores q_j ($q_j^T q_i = 0, \forall i < j$). La maximización de la varianza ante estas restricciones implica la maximización de una expresión v_j que involucra j multiplicadores de LaGrange.

En resumen, el problema de determinar los vectores q_j que generan un espacio ortogonal, se reduce al problema de obtener la estructura propia de la matriz Γ de covarianza de los datos, esto es:

$$\begin{aligned} \Gamma Q &= Q\Lambda \\ \Gamma &= Q\Lambda Q^T \end{aligned} \quad (4.23)$$

Donde $Q \in \mathfrak{R}^{n \times n}$ es la matriz de vectores propios ortonormales y $\Lambda \in \mathfrak{R}^{n \times n}$ es la matriz diagonal de los correspondientes valores propios, ordenados en forma decreciente, esto es $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

4.2.2. Propiedades de la matriz de covarianza Γ

Propiedad 1. Dado que Γ es una matriz simétrica, entonces sus valores propios son reales.

Propiedad 2. Si Γ es positiva definida, entonces sus valores propios satisfacen $\lambda_j > 0$, para $j = 1, 2, \dots, n$.

Propiedad 3. Si Γ es positiva semidefinida de rango p , entonces exactamente p de sus n valores propios son positivos mientras que los $n - p$ valores propios restantes son cero.

4.2.3. Propiedades de las componentes del espacio ortogonal

Propiedad 4. Las varianzas de las componentes ortogonales p_j están dadas por los valores propios de Γ .

Dada la matriz de componentes ortogonales

$$P = [Xq_1 \ Xq_2 \ \dots \ Xq_n] = XQ \quad (4.24)$$

Su covarianza está definida como.

$$\text{cov}(P) = Q^T \Gamma Q \quad (4.25)$$

De donde al sustituir (4.23) en (4.25) resulta:

$$\text{cov}(P) = Q^T (Q \Lambda Q^T) Q = \Lambda \quad (4.26)$$

Dado que $Q^T Q = I_n$.

Propiedad 5. La varianza total de la matriz de observaciones \mathbf{X} está dada por la suma de las varianzas de cada variable, y es igual a la suma de los valores propios λ_j .

Definiendo la varianza total de los datos como:

$$V_T = \text{tr}(\Gamma) = \sum_{j=1}^n \gamma_{jj}^2 \quad (4.27)$$

Por lo tanto, al sustituir (4.23) en (4.27) se obtiene:

$$V_T = \text{tr}(Q \Lambda Q^T) = \text{tr}(Q^T Q \Lambda) = \text{tr}(\Lambda) = \sum_{j=1}^n \lambda_j \quad (4.28)$$

Es decir, la varianza total de \mathbf{X} es igual a la suma de las varianzas de las n componentes principales.

4.2.4. Dimensión del espacio de componentes principales.

Uno de los usos más generalizados del análisis de componentes principales es su habilidad de transformar un conjunto de datos n -variado, en un conjunto reducido de p variables no correlacionadas denominadas componentes principales, las cuales retienen la mayor variabilidad de los datos. A pesar de la ventaja señalada anteriormente, desafortunadamente no existe un método aceptado de manera general para determinar el valor de p o el número de componentes principales.

Un método es el criterio del porcentaje de varianza. Recordando de (4.28) que la varianza total de los datos está dada por la suma de los valores propios, entonces un cierto porcentaje de la varianza se retiene con un número p de componentes, esto es:

$$Rv = \frac{v_p}{v_T} 100 = \frac{\sum_{j=1}^p \lambda_j}{\sum_{j=1}^n \lambda_j} \quad (4.29)$$

El porcentaje de varianza Rv se selecciona para cada proceso particular y está determinado por el grado de correlación entre las variables.

4.3. Modelado implícito

Considerando el conjunto de datos nominales \mathbf{X} de la ecuación (4.5), el modelado implícito basado en análisis de componentes principales se resume en dos pasos, la estandarización de los datos y la generación de los subespacios de componentes principales y residuales.

4.3.1. Estandarización de datos

Debido a la naturaleza de las variables, las cuales pueden tener valores en rangos y/o unidades diferentes, es conveniente estandarizar los datos en \mathbf{X} con respecto a sus medias y desviación estándar. Esta estandarización permite una ponderación equitativa de la variabilidad de los datos en el proceso de obtención de las componentes ortogonales.

Sea el conjunto de datos nominales \mathbf{X} dado en la ecuación (4.5), sus correspondientes medias están dadas por:

$$\hat{\boldsymbol{\mu}} = \left[\frac{1}{N} \mathbf{1X} \right] = [\mu_{x1} \dots \mu_{xn}]_{(1 \times n)} \quad (4.30)$$

Donde $\mathbf{1} = [1, 1, \dots, 1] \in \mathfrak{R}^{1 \times N}$. Así, los datos con extracción de medias se obtiene a partir de:

$$\bar{X} = X - 1^T \hat{\mu}_X \quad (4.31)$$

Por otro lado, la matriz de covarianzas de \mathbf{X} está dada por:

$$\hat{\Gamma} = \frac{1}{N-1} X^T \bar{X} \quad (4.32)$$

Y el vector de desviaciones estándar de \mathbf{X} es:

$$\hat{\sigma}_X = \sqrt{\text{diag}(\Gamma)} = [\hat{\sigma}_{X1} \dots \hat{\sigma}_{Xn}]_{(1 \times n)} \quad (4.33)$$

Por lo tanto, la estandarización de los datos se obtiene a partir de la siguiente expresión:

$$\tilde{X} = \bar{X} D_X^{-1} \quad (4.34)$$

Donde; $D_X = \text{diag}(\hat{\sigma}_X)$

4.3.2. Subespacios de componentes principales y subespacios residual

A partir de la matriz de correlación entre variables, $\hat{\Sigma} = \frac{1}{N-1} \tilde{X}^T \tilde{X} \in \mathfrak{R}^{n \times n}$, es posible obtener su correspondiente estructura propia, esto es:

$$\hat{\Sigma} Q = Q \Lambda \text{ o } \hat{\Sigma} = Q \Lambda Q^T \quad (4.35)$$

En general, Q puede ser descompuesta en dos submatrices como:

$$Q = [Q_p \ Q_r] \quad (4.36)$$

Dónde: $Q_p \in \mathfrak{R}^{n \times p}$ y $Q_r \in \mathfrak{R}^{n \times r}$, $r = n - p$.

Hecho 1 Dado que los vectores propios en Q son ortonormales entonces se satisface:

$$Q_p^T Q_p = I_p$$

$$Q_r^T Q_r = I_r$$

$$Q_p Q_p^T + Q_r Q_r^T = I_n$$

Demostración. Dado que Q es una matriz cuadrada ortonormal, se satisface:

$$Q^T Q = I_n = Q Q^T \quad (4.37)$$

Sustituyendo (4.37) en (4.38)

$$Q^T Q = \begin{bmatrix} Q_p^T \\ Q_r^T \end{bmatrix} [Q_p \ Q_r] = \begin{bmatrix} Q_p^T Q_p & Q_p^T Q_r \\ Q_r^T Q_p & Q_r^T Q_r \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_r \end{bmatrix} = I_n \quad (4.38)$$

$$Q Q^T = [Q_p \ Q_r] \begin{bmatrix} Q_p^T \\ Q_r^T \end{bmatrix} = Q_p Q_p^T + Q_r Q_r^T = I_n$$

Así, \mathbf{Q}_p el subconjunto de vectores propios asociados a los valores propios más significativos, constituye una base para el subespacios denominado de las componentes principales; por su parte, \mathbf{Q}_r el subconjunto de vectores propios complementarios, asociados a los valores propios menos significativos, es una base para el subespacios denominado como residual. Las dimensiones de los subespacios se pueden determinar de acuerdo con los procedimientos descritos en la sección donde se discutió sobre la dimensión del espacio de componentes principales.

En el análisis convencional del modelado basado en PCA, el conjunto de datos \mathbf{X} se maneja sin hacer diferencia entre variables de entrada y salida. Sin embargo, considérese que \mathbf{X} está conformado por mediciones de l-variables de entrada y s-variables de salida de un proceso, las cuales se representan por los siguientes vectores:

$$u(t) = u^0(t) + \eta(t)$$

$$y(t) = y^0(t) + v(t)$$

Donde $\eta(t)$ y $v(t)$ son vectores de ruido blanco adicionado a la entrada y a la salida del sistema, con media cero y varianzas γ_η^2 y γ_v^2 respectivamente. Asumiendo la existencia de relaciones lineales estáticas entre $u(t)$ e $y(t)$, expresadas en la forma:

$$y(t) = \mathbf{B}u(t) + v(t) \quad (4.39)$$

Donde: $\mathbf{B} \in \mathfrak{R}^{s \times l}$ y $\text{rank}(\mathbf{B}) = l$, entonces se satisface el siguiente hecho.

Hecho 2 sea un conjunto de datos $X = [U^T \ Y^T]$ del sistema (4.39). En ausencia de ruido en las mediciones, dadas las s-relaciones lineales de X , su correspondiente matriz de covarianza $\Gamma \in \mathfrak{R}^{n \times n}$ satisface $\text{Rank}(\Gamma) = l$, por lo tanto, de acuerdo **con la propiedad 3**, los l-primeros valores propios de Γ son positivos, mientras que los últimos $s = n - l$ valores propios son ceros.

Demostración. Los parámetros estadísticos de media y covarianza de las variables de entrada son:

$$\mu_u = E[u]$$

$$\Gamma_u = E[u^T u] - \mu_u \mu_u^T \quad (4.40)$$

Y los de las variables de salida

$$\mu_y = E[y] = B\mu_x$$

$$\Gamma_y = E[yy^T] - \mu_y\mu_y^T = B\Gamma_u B^T \quad (4.41)$$

La covarianza entre las señales de entrada y salida está dada por:

$$\Gamma = E[xx^T] - \mu_x\mu_x^T = \begin{bmatrix} \Gamma_u & \Gamma_u B^T \\ B\Gamma_u & \Gamma_y \end{bmatrix} = \begin{bmatrix} \Gamma_u & \Gamma_u B^T \\ B\Gamma_u & B\Gamma_u B^T \end{bmatrix} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \quad (4.42)$$

Dónde:

$$\mu_x = \begin{bmatrix} \mu_u \\ \mu_y \end{bmatrix} = \begin{bmatrix} \mu_u \\ B\mu_u \end{bmatrix} \quad (4.43)$$

Dado que Γ_{12} , Γ_{21} y Γ_{22} son combinaciones lineales de Γ_{11} , entonces $\text{Rank}(\Gamma) = \text{Rank}(\Gamma_{11}) = \text{Rank}(\Gamma_u) = 1$. por lo tanto, los l primeros valores propios de Γ serán positivos, mientras que los últimos $s = n - 1$ valores propios tendrán valor cero. Ante la presencia de ruido en las mediciones los s -últimos valores propios serán aproximadamente cero.

Es decir, si se sabe que existe una relación lineal (4.39) entre las variables de entrada y salida, entonces, de acuerdo con el **Hecho 2**, se puede seleccionar *a priori* las dimensiones de los subespacios de componentes principales y residual con $p = l$ y $r = s$, respectivamente.

En conclusión, a partir de un conjunto de datos nominales \mathbf{X} de un proceso multivariado, a través de PCA, este puede ser modelado implícitamente mediante el conjunto de parámetros siguientes:

$$\{\hat{\mu}_x, \hat{\sigma}_x, Q_p, Q_r\} \quad (4.44)$$

Con base en este modelo implícito, el monitoreo del proceso se puede llevar a cabo mediante la evaluación en línea de la proyección de nuevas observaciones, ya sea en el subespacios de componentes principales o en el residual.

A continuación se describe el procedimiento de detección de fallos múltiples basado en PCA.

4.4. Detección de fallos múltiples basados en PCA

4.4.1. Umbrales de los residuos

El procedimiento de detección de fallos múltiples consiste en la evaluación de residuos que se pueden generar en el subespacios de componentes principales o en el subespacios residual. Específicamente, en el subespacios de componentes principales se usa como residuo el parámetro de Hotelling T^2 , y en el subespacios residual se usa como residuo el error cuadrático de predicción, SPE.

Cada observación $\tilde{x}(i) \in \mathfrak{R}^{1 \times n}$ del conjunto de datos \mathbf{X} dado en (4.5), se puede expresar en el subespacios de las componentes principales como se define a continuación.

$$p(i) = \tilde{x}(i)Q_p \quad (4.45)$$

Y adicionalmente en el subespacio residual como:

$$r(i) = \tilde{x}(i)Q_r \quad (4.46)$$

Tomando en cuenta los datos en el subespacios de las componentes principales (4.45) y en el subespacio residual (4.46), a continuación se describe cómo generar los residuos de cada subespacio y determinar los respectivos umbrales.

4.4.1.1. Umbral de los residuos Hotelling (T^2)

El residuo Hotelling T^2 para cada una de las observaciones $\mathbf{p}(i)$ está dado por:

$$T_p^2(i) = p(i)\Lambda_p^{-1}p(i)^T \quad (4.47)$$

Dónde: Λ_p es la matriz de \mathbf{p} , $\Lambda_p = \text{diag}(\lambda_1 \lambda_2 \dots \lambda_p)$. Sustituyendo (4.45) en (4.47), se puede obtener una expresión de T_p^2 en términos de $\tilde{z}(i)$, esto es:

$$T_p^2(i) = \tilde{x}(i)Q_p\Lambda_p^{-1}Q_p^T\tilde{x}(i)^T = \tilde{x}(i)F\tilde{x}(i)^T \quad (4.48)$$

Donde

$$F = Q_p\Lambda_p^{-1}Q_p^T = F^T \quad (4.49)$$

$$F = F_1 + \dots + F_p = q_1\lambda_1^{-1}q_1^T + \dots + q_p\lambda_p^{-1}q_p^T \quad (4.50)$$

Donde \mathbf{F} y \mathbf{F}_j son matrices simétricas.

Este estadístico representa una medida de la variación del proceso, ya que se construye a partir del subespacio de la variación dominante, formados por los a primeros auto vectores.

Si este estadístico supera un umbral se puede inferir que el proceso estará fuera de control, es decir, que se ha producido un fallo. Este umbral se puede calcular fuera de línea a partir de los datos de muestreo de la siguiente forma:

A partir de la distribución de probabilidad del conjunto de parámetros $T_p^2(i)$, para $i = 1, \dots, N$, se calcula el umbral de condición normal UCL_p . La distribución de probabilidad de T_p^2 puede ser χ^2 F o β , la diferencia es menos notable conforme $N \rightarrow \infty$. Asumiendo una distribución tipo β , UCL_p se determina como sigue:

$$UCL_p = \frac{(N-1)^2 \left(\frac{p}{N-p-1} \right) F\left(\frac{\alpha}{2}; p, N-p-1\right)}{N \left(1 + \left(\frac{p}{N-p-1} \right) F\left(\frac{\alpha}{2}; p, N-p-1\right) \right)} \quad (4.51)$$

Donde N y p son las dimensiones de \mathbf{P} y α es el nivel de confianza. La confiabilidad de este umbral depende de la validez de la **Suposición 2**.

El estadístico T^2 se basa en los a primeros componentes principales, por eso, proporciona una medida de las desviaciones de las variables latentes, que son aquellas que mejor explican la variabilidad del proceso. De este modo, este sólo detecta un evento si la variación en las variables latentes es mayor que la variación explicada por causas comunes, pero pueden existir otro tipo de eventos que producen grandes variaciones en el espacio de ruido desechado para el cálculo de este estadístico. Estos eventos se detectan con el estadístico Q que se explica más adelante.

Existen distintos métodos para la selección del número de componentes principales, [Jac-91]; [Chi-00]; [Wei-01]; [Peñ-02]. Uno de esos métodos consiste en realizar un gráfico de barras que representa los valores propios ordenados de mayor a menor y en buscar un codo, es decir un punto a partir del cual los valores propios son aproximadamente iguales. Otro método consiste en seleccionar el número de componentes hasta cubrir una cota de varianza (CPV). Se suelen seleccionar valores de CPV de entre 80% y 90% [Peñ-02]. Otra opción consiste en despreciar los valores los componentes con valores propios menores que la unidad, si se trabaja con la matriz de correlación el valor medio de los componentes es 1. Todos estos métodos tienen un planteamiento un tanto heurístico y el proceso de selección del número de a componentes principales óptimo se continua trabajando y desarrollando para investigaciones futuras.

4.4.1.2. Umbral de los residuos SPE

Sea una observación $\tilde{x}(i)$ del conjunto de datos X , esta se puede escribir como:

$$\tilde{x}(i) = \tilde{x}(i)QQ^T \quad (4.52)$$

Dónde: $QQ^T = I_n$. Dado que $Q = [Q_p \quad Q_r]$, y de acuerdo con el **Hecho 1**, $\tilde{x}(i)$ se reescribe como:

$$\tilde{x}(i) = \tilde{x}(i)[Q_p Q_p^T \quad Q_r Q_r^T] = \tilde{x}(i)Q_p Q_p^T + \tilde{x}(i)Q_r Q_r^T \quad (4.53)$$

Sustituyendo (4.45) y (4.46) en (4.53) se obtiene:

$$\tilde{x}(i) = p(i)Q_p^T + r(i)Q_r^T = \tilde{x}^p(i) + \tilde{x}^r(i) \quad (4.54)$$

Es decir, (4.54) está compuesto por dos términos, donde $\tilde{x}^p(i)$ corresponde al estimado de $\tilde{x}(i)$ a partir del subespacio de componentes principales, y $\tilde{x}^r(i)$ corresponde al estimado a partir del subespacio residual y constituye el error de estimación.

Así, el residuo SPE para cada una de las observaciones $r(i)$, para $i = 1, \dots, N$, está dado por:

$$SPE_r(i) = \tilde{x}^r(i)\tilde{x}^r(i)^T = r(i)Q_r^T Q_r r(i)^T = r(i)r(i)^T \quad (4.55)$$

Sustituyendo (4.46) en (4.55), se obtiene una expresión de SPE_r en términos de $\tilde{x}(i)$, esto es:

$$SPE_r(i) = \tilde{x}(i)Q_r Q_r^T \tilde{x}(i)^T \quad (4.56)$$

El umbral fijo de condición normal UCLr para el conjunto de parámetros univariado, $SPE_r(i)$, asociados al subespacio residual, se calcula con los datos tomados fuera de línea mediante la expresión:

$$UCL_r = \theta_1 \left[\frac{h_o c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_o (h_o - 1)}{\theta_1^2} \right]^{\frac{1}{h_o}} \quad (4.57)$$

Dónde: $\theta_k = \sum_{j=p+1}^n \lambda_j^k$; $h_o = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2}$, c_α es la desviación normal correspondiente al percentil $(1 - \alpha)$,

α es un nivel de confianza del estimado UCLr [Chi-00].

Cuando el proceso está bajo control, el estadístico Q representa fluctuaciones del ruido. Si ocurriese un evento inusual que afectase la covarianza de X, se detectaría porque la estadística Q superaría el umbral.

4.5. Algoritmo para la detección de fallos múltiples mediante PCA.

Para implementar un sistema de monitorización con PCA hay que llevar a cabo dos fases: por un lado el análisis previo donde se realiza el análisis de componentes principales, y por otro lado la monitorización propiamente dicha.

4.5.1. Calculo del PCA (fuera de línea)

Esta fase es previa a la detección de fallos y se realiza fuera de línea. Una vez que se tienen los datos del proceso en condiciones normales de operaciones se genera la matriz X y se lleva a cabo el análisis de componentes principales. Previo al análisis es conveniente normalizar los valores de las variables medidas a valores de media cero y varianza uno.

En esta fase también se calculan los umbrales o límites de operación de los estadísticos.

4.5.2. Calculo del PCA (En línea)

Una vez que se tienen los datos relativos al PCA se puede realizar la monitorización y detección de fallos del proceso. Esta fase se realiza en línea con el proceso. Para cada una de las nuevas lecturas que se obtengan del proceso monitorizándose tienen que calcularse los estadísticos T^2 y Q. Previo al cálculo de los estadísticos hay que normalizar el vector de medidas del proceso con la media y la varianza de los datos con los que se realizó el PCA fuera de línea. Si alguna de las estadísticas supera el umbral se notifica una alarma, si el número de alarmas consecutivas supera un número preestablecido se notificará que ha sucedido un fallo.

Sea una nueva observación

$$x^a = [x_1 \quad x_2 \quad \dots \quad x_n] \tag{4.58}$$

El monitoreo del proceso consiste en evaluar en línea la consistencia de $x^a \in \mathfrak{R}^{1 \times n}$ con respecto al modelo implícito (4.46) según se muestra en la figura 4.3. Esto se lleva a cabo mediante la evaluación de los residuos de x^a , generados en los subespacios de componentes principales y residuales, con respecto a los correspondientes umbrales (4.51) y (4.57).

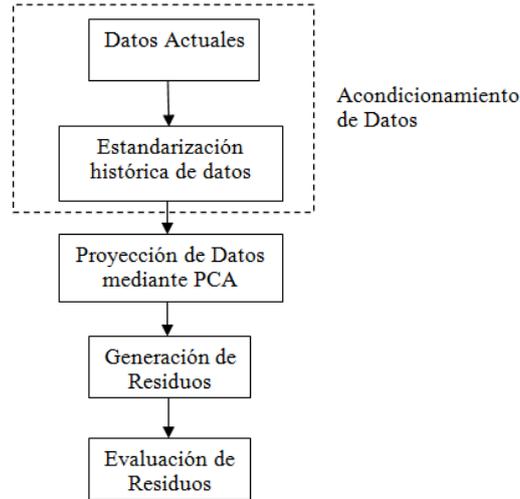


Figura 4. 3. Esquema de detección de fallos múltiples mediante PCA

El procedimiento de monitoreo se resume en los siguientes pasos:

Paso 1. Estandarizar (4.58) con respecto a las medias (4.30) y las desviaciones estándar (4.33), esto es:

$$\bar{x}^a = x^a - \hat{\mu}_x$$

$$\tilde{x}^a = \bar{x}^a D_x^{-1} \quad (4.59)$$

Paso 2. Transformar el vector estandarizado \tilde{x}^a en los subespacios de componentes principales y residuales, mediante Q_p y Q_r .

$$p^a = \tilde{x}^a Q_p$$

$$r^a = \tilde{x}^a Q_r \quad (4.60)$$

Paso 3. Generar los correspondientes residuos para p^a y r^a

$$T_{p^a}^2 = p^a \Lambda_p^{-1} p^{aT}$$

$$SPE_{r^a} = r^a r^{aT} \quad (4.61)$$

Paso 4. Si alguno de los residuos es mayor a su respectivo umbral, UCL_p o UCL_r, entonces la presencia de un fallo en el proceso es detectada.

4.6. Resultados:

A continuación se muestra la metodología del análisis de componentes principales para la detección de fallos múltiples en una turbina de gas SR-30 ubicada en la Universidad Interamericana de Puerto Rico, Recinto de Bayamón [Rin-12].

4.6.1. Cálculo del análisis de componentes principales fuera de línea.

En esta fase se usaron los datos obtenidos del modelo desarrollado en Simulink para la turbina de gas SR-30 cuando no existen fallos en el sistema. Las velocidades de rotación de la turbina usadas para verificar la metodología fueron: 55.000, 65.000 y 70.000 revoluciones por minuto, según se muestra en las figuras 4.4 a la 4.6.

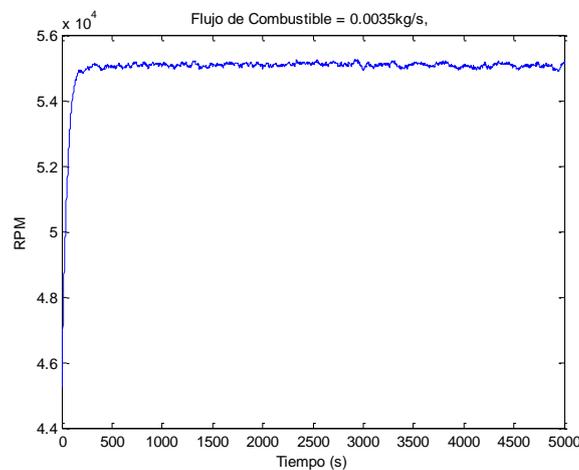


Figura 4. 4. Turbina de gas SR30 a 55000 rpm.

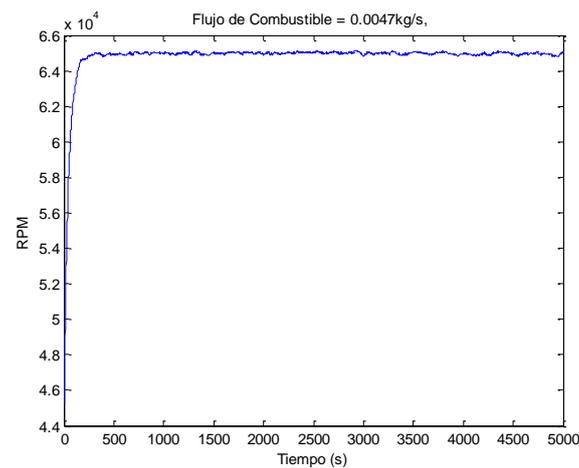


Figura 4. 5. Turbina de gas SR30 a 65000 rpm.

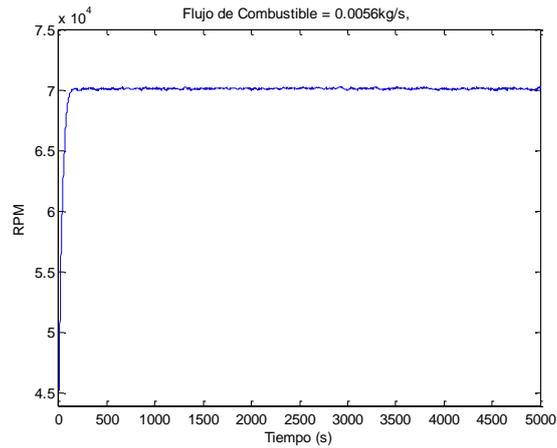


Figura 4. 6. Turbina de gas SR30 a 70000 rpm.

Para cada una de las velocidades de rotación de la turbina se almacenaron los siguientes datos de los sensores acorde con la tabla 4.1.:

Tabla 4. 1. Variables medidas en la turbina de gas SR30

SENSOR	VARIABLE MEDIDA
s1	Temperatura Estática de entrada del Compresor (T01)
s2	Temperatura de Estancamiento de salida de la etapa de compresor (T02)
s3	Temperatura del Estancamiento de entrada de la etapa de la turbina (T03)
s4	Temperatura de Estancamiento de salida de la etapa de la turbina (T04)
s5	Temperatura de Estancamiento de salida de la tobera (T05)
s6	Presión estática de entrada del compresor (P01)
s7	Presión de estancamiento de salida de la etapa de compresor (P02)
s8	Presión de la cámara de combustión (P03)
s9	Presión de estancamiento de salida de la turbina (P04)
s10	Presión de estancamiento de la salida de la tobera (P05)
s11	Revoluciones del eje de la turbina. (RPM)
S12	Par generado por la turbina (Tr)

Las figuras 4.7 a la 4.9 muestran el comportamiento de las temperaturas, presiones y torque en la turbina de gas SR30 a condiciones normales de operación para 65000 rpm.

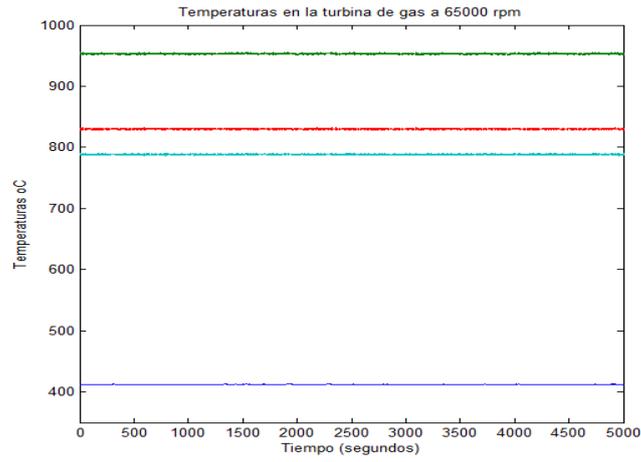


Figura 4. 7. Temperaturas en la turbina

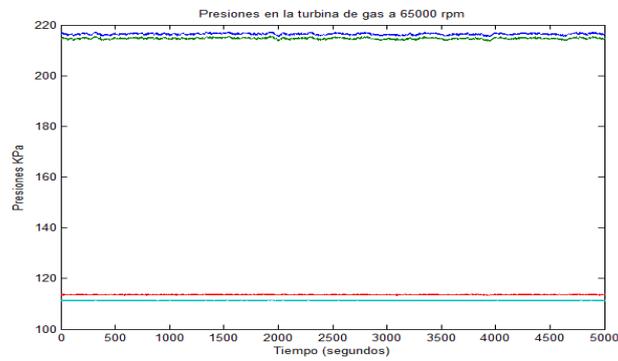


Figura 4. 8. Presiones en la turbina

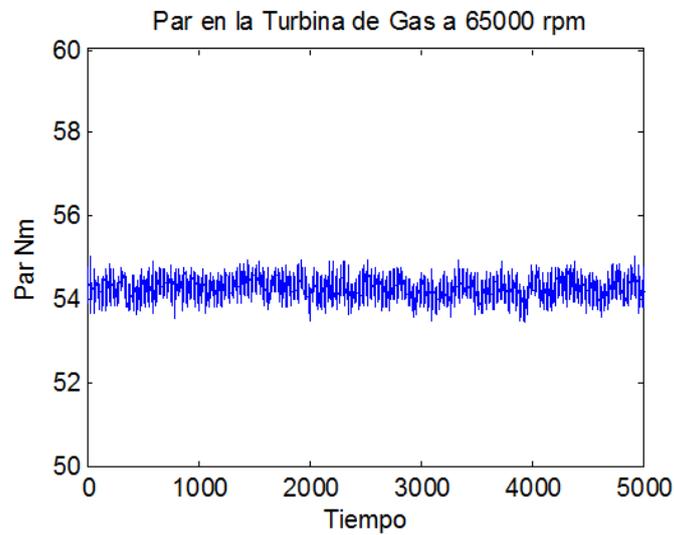


Figura 4. 9. Par en la turbina

Los datos obtenidos se estandarizaron para poder obtener media cero y varianza uno. Para la construcción del modelo usando análisis de componentes principales se siguieron los siguientes pasos:

- a- Se obtuvo la matriz de covarianza usando las lecturas estandarizadas de los diez sensores definidos en la tabla 4.1.

$$C = \begin{bmatrix} 1.0000 & 0.5754 & 0.2537 & 0.2537 & 0.9866 & 0.9862 & 0.4856 & 0.4325 & 0.9977 & 0.6434 \\ 0.5754 & 1.0000 & 0.9368 & 0.9368 & 0.6794 & 0.6820 & 0.9929 & 0.9851 & 0.5745 & 0.9939 \\ 0.2537 & 0.9368 & 1.0000 & 1.0000 & 0.3799 & 0.3831 & 0.9651 & 0.9786 & 0.2518 & 0.8986 \\ 0.2537 & 0.9368 & 1.0000 & 1.0000 & 0.3799 & 0.3831 & 0.9651 & 0.9786 & 0.2518 & 0.8986 \\ 0.9866 & 0.6794 & 0.3799 & 0.3799 & 1.0000 & 1.0000 & 0.6032 & 0.5543 & 0.9905 & 0.7449 \\ 0.9862 & 0.6820 & 0.3831 & 0.3831 & 1.0000 & 1.0000 & 0.6060 & 0.5572 & 0.9900 & 0.7473 \\ 0.4856 & 0.9929 & 0.9651 & 0.9651 & 0.6032 & 0.6060 & 1.0000 & 0.9982 & 0.4888 & 0.9814 \\ 0.4325 & 0.9851 & 0.9786 & 0.9786 & 0.5543 & 0.5572 & 0.5572 & 1.0000 & 0.4356 & 0.9682 \\ 0.9977 & 0.5745 & 0.2518 & 0.2518 & 0.9905 & 0.9900 & 0.4888 & 0.48356 & 1.0000 & 0.6470 \\ 0.6434 & 0.9939 & 0.8986 & 0.8986 & 0.7449 & 0.7473 & 0.9814 & 0.9682 & 0.6470 & 1.0000 \end{bmatrix}$$

- b- Utilizando la metodología propuesta por Peña 2002 se calcula los tres primeros auto valores y auto vectores:

$$\text{autovalores} = \begin{bmatrix} 7.5482 & 0 & 0 \\ 0 & 2.4397 & 0 \\ 0 & 0 & 0.0121 \end{bmatrix}$$

$$\text{autovectores} = \begin{bmatrix} 0.2699 & 0.4274 & -0.6095 \\ 0.3546 & -0.1430 & -0.2673 \\ 0.3037 & -0.3522 & -0.3099 \\ 0.3037 & -0.3522 & -0.3099 \\ 0.3010 & 0.3598 & 0.2191 \\ 0.3017 & 0.3579 & 0.2194 \\ 0.3445 & -0.2054 & 0.2996 \\ 0.3369 & -0.2414 & 0.2967 \\ 0.2705 & 0.4284 & 0.0129 \\ 0.3605 & -0.0853 & 0.3015 \end{bmatrix}$$

Graficas como la mostrada en la figura 4.10 se desarrollaron para las diferentes velocidades de rotación de la turbina. Los valores de velocidades se seleccionaron cada 5000 revoluciones por minuto, usando como valor mínimo 45000 rpm y como valor máximo 75000 rpm.

La elipse mostrada en la figura 4.10 representa el área dentro del cual deben ubicarse los componentes principales para representar un sistema sin ningún fallo. Al presentarse fallos únicos o fallos múltiples los valores de las componentes principales quedan fuera de esta elipse y podría determinarse la presencia de los fallos.

Al usar el método gráfico se tiene la ventaja de que visualmente se puede detectar más fácil el fallo. Cuando se usan dos componentes principales el método básicamente busca el área mínima que encierra los componentes principales sin fallo en un plano x-y y la complejidad computacional es aceptable en el consumo de memoria y velocidad de ejecución de los algoritmos.

Cuando se usan tres componentes principales el método básicamente busca el volumen mínimo que encierra los componentes principales sin fallo en un plano 3D y la complejidad computacional empieza a ponerse más rigurosa en el consumo de memoria y velocidad computacional. Para el caso en el que se seleccionan más de tres componentes principales el método gráfico se vuelve ineficiente y ya la visualización no es posible.

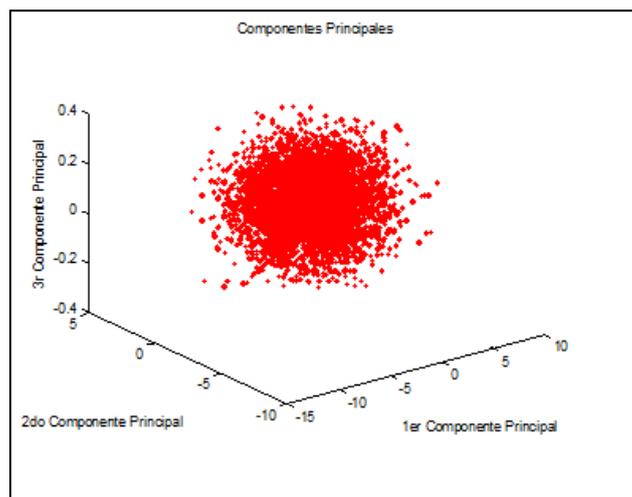


Figura 4. 10. Componentes principales para 65000 rpm

Las figuras 4.11 y 4.12 muestran los umbrales de detección de fallos mediante el uso de los criterios de Hotelling T2 y error de predicción al cuadrado SPE cuando en el sistema no hay fallos presentes.

Al usar los criterios de Hotelling (T2) y Error de predicción al cuadrado (SPE) se puede trabajar de manera sencilla la manera de visualizar los datos con o sin la presencia de error en un plano 2D.

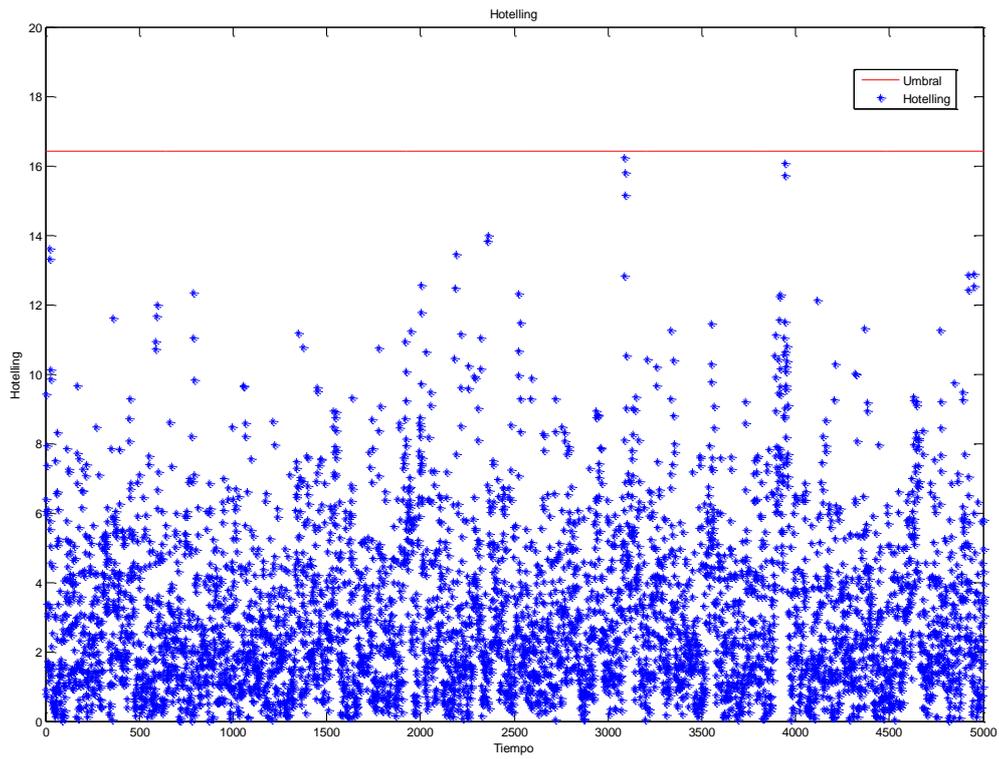


Figura 4. 11. Criterio Hotelling para condiciones sin fallo a 65000 rpm

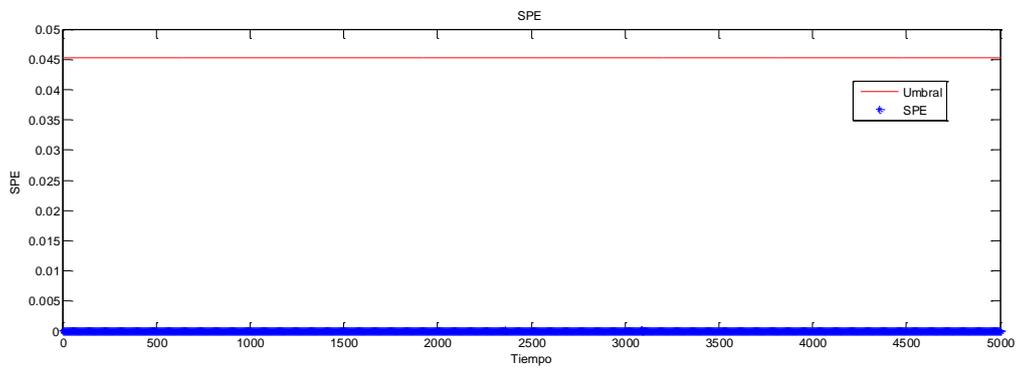


Figura 4. 12. Criterio de Error de Predicción al cuadrado (SPE) para condiciones sin fallo a 65000 rpm

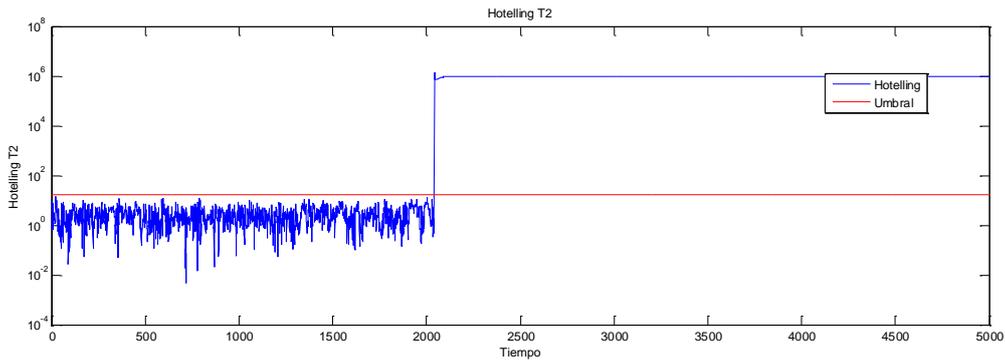


Figura 4.13. Criterio de Hotelling para fallo de 10% en la eficiencia del compresor.

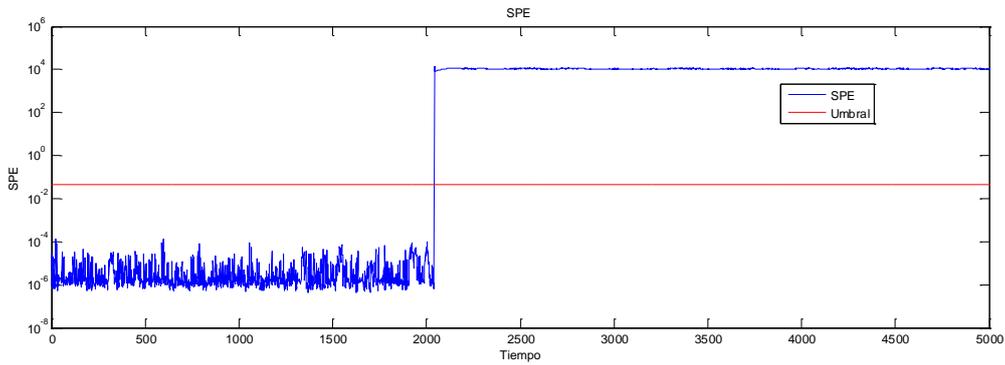


Figura 4.14. Criterio de SPE para fallo de 10% en la eficiencia del compresor.

Las figuras 4.13 y 4.14 muestran los umbrales de detección de fallos mediante el uso de los criterios de Hotelling y error de predicción al cuadrado SPE cuando en el sistema experimenta un fallo de 10% en la eficiencia del compresor.

La figura 4.15 muestra el criterio de Hotelling implementado cuando el sistema experimenta a los 2000 segundos un fallo de 10% en la eficiencia del compresor y a los 3500 segundos un fallo de 10% en la eficiencia de la turbina.

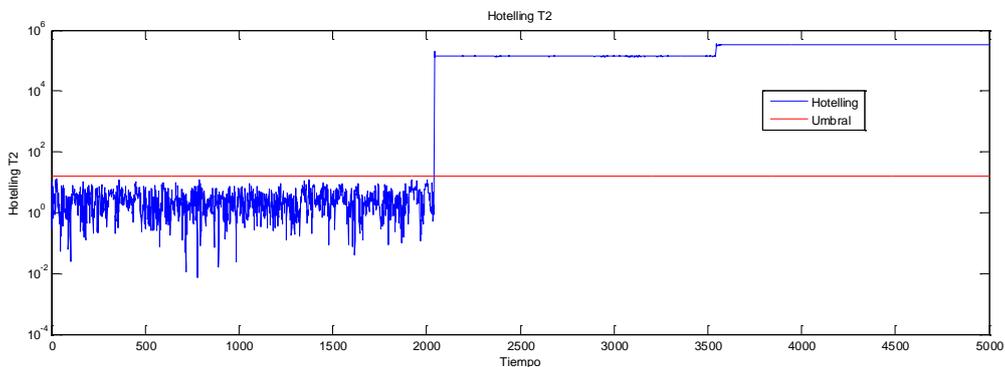


Figura 4.15. Criterio de Hotelling T2 para fallos secuenciales

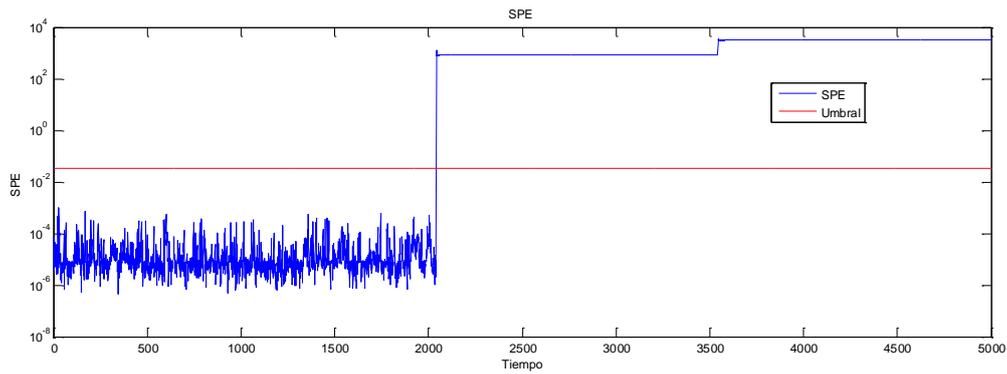


Figura 4. 16. Error de predicción al cuadrado para dos fallos secuenciales 5%WC y 5%ET a 55000 rpm

La figura 4.16 muestra el criterio de error de predicción al cuadrado (SPE) implementado cuando el sistema experimenta a los 2000 segundos un fallo de 5% en el flujo de aire en el compresor y a los 3500 segundos un fallo de 5% en la eficiencia de la turbina.

Las figuras 4.17 y 4.18 muestran el criterio de Hotelling T2 y SPE aplicado cuando el sistema presenta a los 2000 segundos dos fallos simultáneos de 5% en la eficiencia del compresor y 10% en la eficiencia de la turbina.

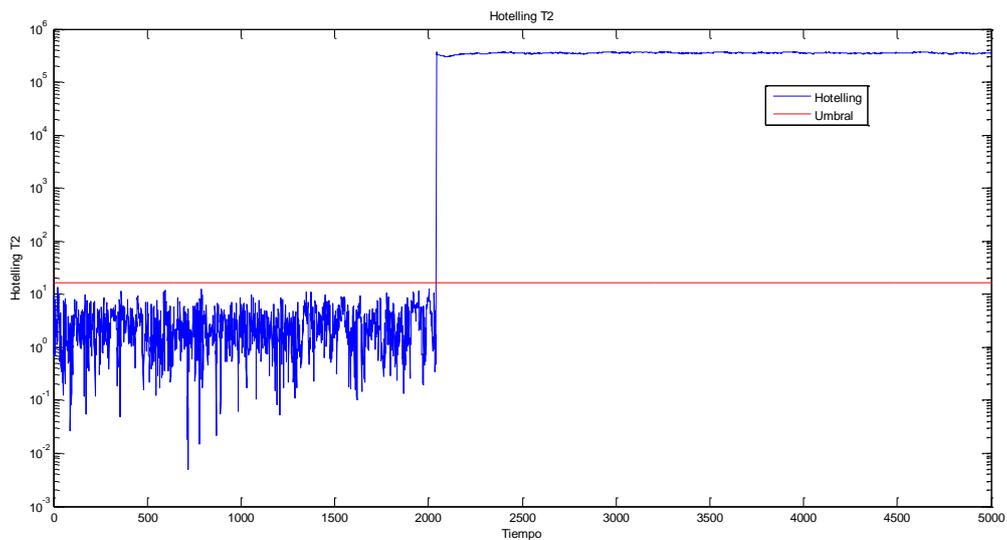


Figura 4. 17. Hotelling T2 aplicado al sistema con fallos simultáneos 5EC 10ET a 65000 rpm.

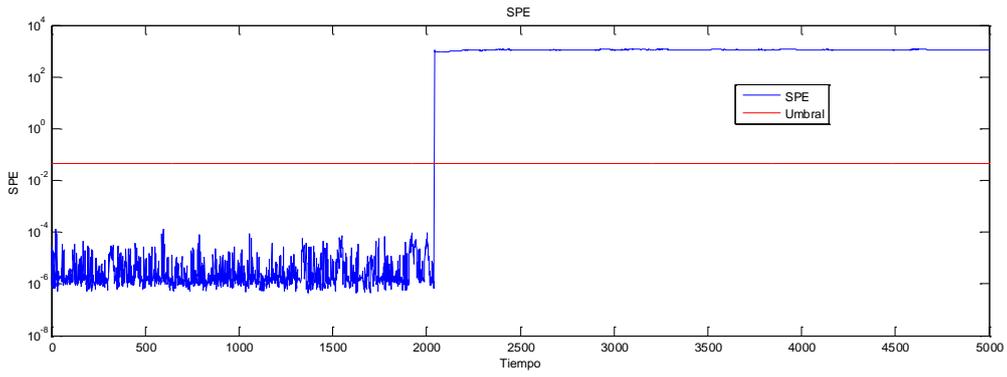


Figura 4. 18. SPE aplicado al sistema con fallos simultáneos 5EC 10ET a 65000 rpm.

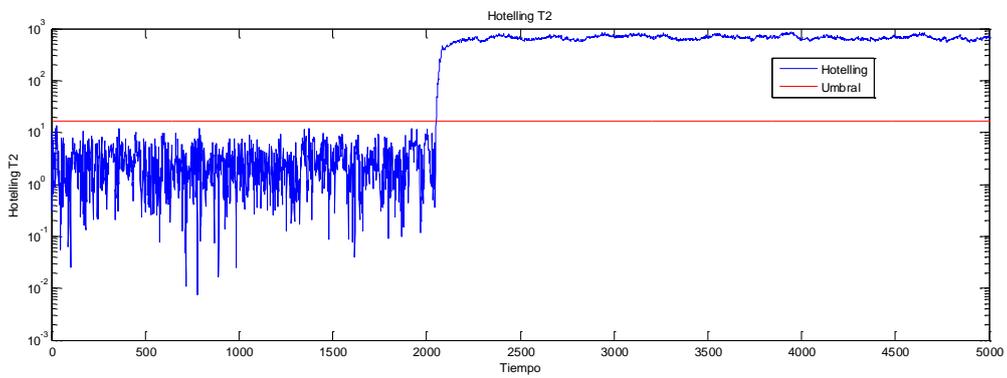


Figura 4. 19. Hotelling para fallos simultáneos de 2% EC y 3%ET

Las figuras 4.19 y 4.20 muestran los criterios de Hotelling (T2) y SPE para condiciones de fallo de un 2% en la eficiencia en el compresor y 3% en la eficiencia en la turbina. Los fallos se simulan simultáneamente a 2000 segundos.

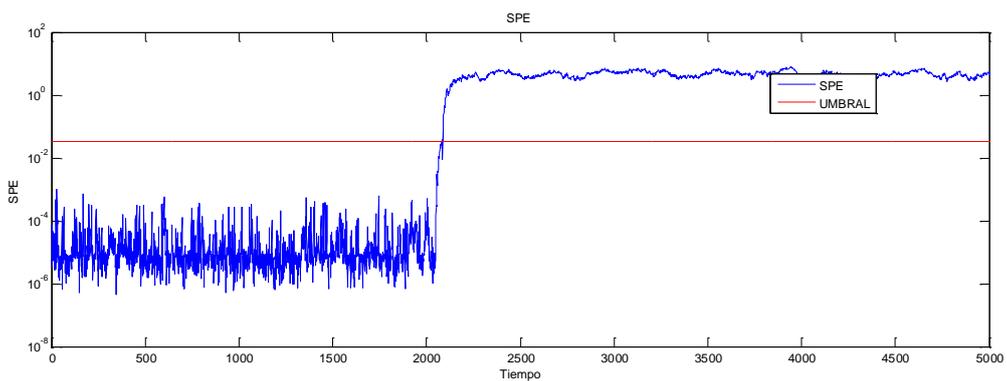


Figura 4. 20. SPE fallos simultáneos 2%EC y 3%ET

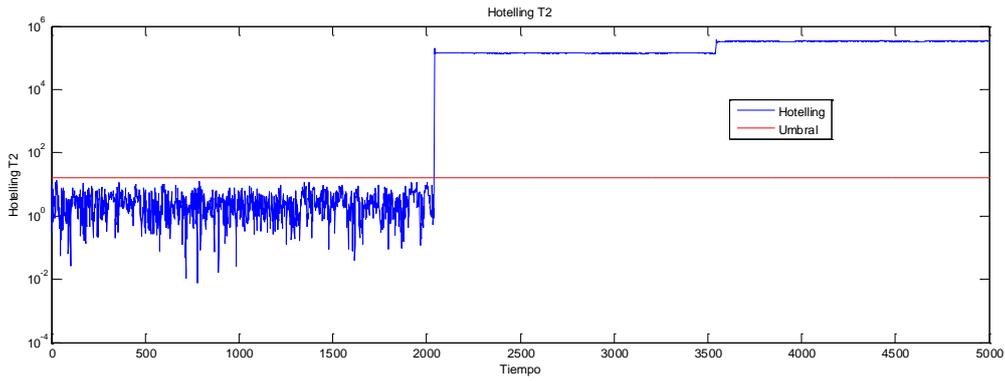


Figura 4. 21. Hotelling fallos secuenciales 5%WC y 5%ET a 55000 rpm.

Las figuras 4.21 y 4.22 muestran los criterios de Hotelling (T2) y SPE para condiciones de fallo de un 5% en el flujo de aire en el compresor y 5% en la relación de presiones en la turbina. Los fallos se simulan consecutivamente a 2000 segundos y 3500 segundos.

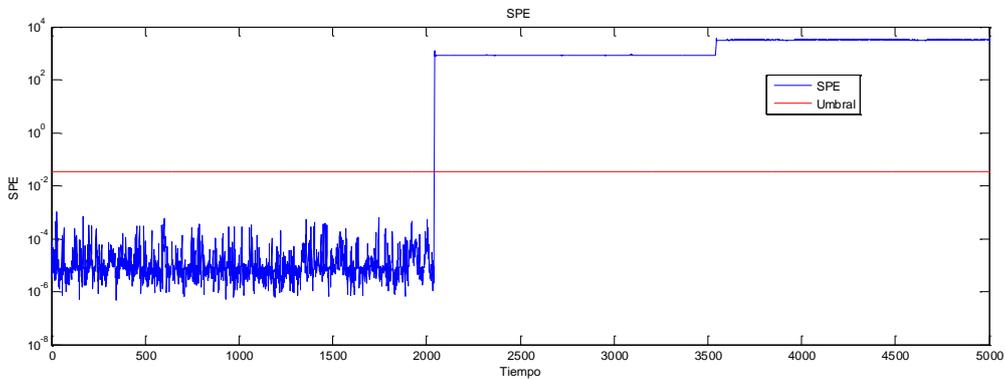


Figura 4. 22. SPE fallos secuenciales 5%WC y 5%ET

Las figuras 4.23 y 4.24 muestran los criterios de Hotelling (T2) y SPE para condiciones de fallo de un 5% en la eficiencia del compresor y 5% en la eficiencia de la turbina. Se simulan fallos incipientes y simultáneos a 2000 segundos.

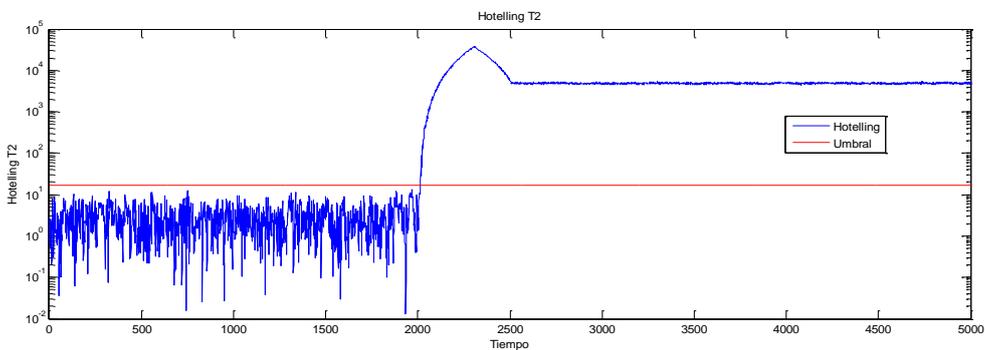


Figura 4. 23. Hotelling T2 fallos incipientes simultáneos 5EC 5ET 2000s

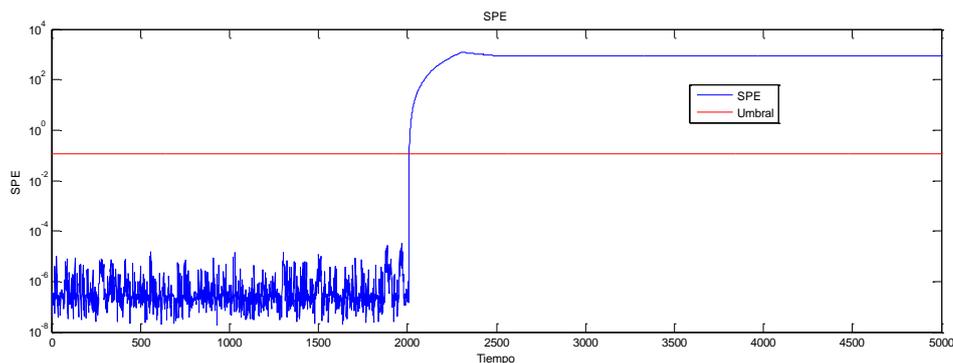


Figura 4. 24. SPE fallos incipientes simultáneos 5EC 5ET 2000s

Es importante destacar que los fallos estudiados en esta tesis son fallos en los componentes de la turbina y no en los sensores. La detección y el diagnóstico de fallos en sensores ha sido estudiada y analizada en muchas investigaciones y no ocurre así con los fallos en componentes. Los fallos que se simularon fueron de naturaleza abrupta e incipientes.

La tabla 4.2 muestra los diferentes tipos de fallos analizados y probados en la turbina SR30.

Tabla 4. 2. Fallos analizados

Fallo	Tipo de Fallo
F1	Perdida de presión en los alabes del compresor (PRC)
F2	Disminución en la Eficiencia en el compresor (EC)
F3	Obstrucción de la masa de aire a través del compresor (WC)
F4	Perdida de presión en los alabes de la Turbina (PRT)
F5	Disminución en la Eficiencia en la turbina (ET)
F6	Obstrucción de la masa de aire y combustible a través de la turbina (MT)
F7	Obstrucción o fallo en el actuador de la válvula de combustible (WF)

La tabla 4.3 muestra las combinaciones de los fallos en componentes estudiados y verificados mediante simulación. Es importante destacar que todos estos fallos múltiples dobles y triples tienen un 5% de porcentaje de éxito en la detección y el sistema de detección necesita 0.4 segundos para conocer un fallo.

Tabla 4. 3. Fallos múltiples analizados

Fallos Múltiples dobles
F1F2; F1F3; F1F4; F1F5; F1F6; F1F7; F2F3; F2F4; F2F5; F2F6; F2F7; F3F4; F3F5; F3F6 F3F7; F4F5; F4F6; F4F7; F5F6; F5F7; F6F7.
Fallos Múltiples triples
F1F2F3; F1F2F4; F1F2F5; F1F2F6; F1F2F7; F1F3F5; F1F3F6; F1F3F7; F1F4F5; F1F4F6; F1F4F7 F1F5F6; F1F5F7; F1F6F7.

4.7. Conclusiones

Como resultado del análisis desarrollado en este capítulo, resulta aconsejable en los casos en que las variables estén expresadas en diferentes unidades de medida trabajar con la matriz de correlaciones, para aislar así los efectos nocivos de las distintas unidades de medida.

Por otro lado, en los casos en que las variables se encuentren en la misma unidad de medida, debe asegurarse que la información que proveen las diferencias en las varianzas al análisis estadístico sea un elemento útil al propósito del análisis, para usar el enfoque de la matriz de varianzas y covarianzas. Esto es así, ya que al usar la matriz de correlación no solo se anulan los efectos de unidad de medida si no toda otra información referente a la diferencia de varianza que, en algunos casos, podría aportar riqueza al estudio bajo análisis y su no inclusión podría distorsionar el significado y en análisis de las componentes.

Adicionalmente merece la pena destacar varias conclusiones importantes: En primer lugar, si las varianzas de las variables son mayores que uno, utilizar la matriz de correlaciones, que indirectamente estandariza las varianzas de todas las variables a uno, tiende a incrementar la proporción de la varianza total explicada por la primer componente principal.

En segundo lugar, alcanzar una representación en un espacio de dimensión reducida con la técnica de componentes principales requiere que exista correlación en algunas de sus variables, aquellas variables que están altamente correlacionadas tienen la misma representación y esta técnica de análisis no es adecuada para el proceso que se evalúe. En procesos industriales donde se tiene suficiente información de las variables es posible aplicar técnicas estadísticas, sin embargo, lograr una representación con menor número de variables depende de las propiedades estadísticas en el modelo.

El método de detección de fallos múltiples mediante PCA está limitado solamente a la detección y no al diagnóstico. Otra limitante del método es que solamente permite analizar sistemas lineales y en el caso de sistemas complejos no lineales presenta muchas falsas alarmas de detección.

Si en el proceso de monitorización aparecen dos fallos diferentes en distintos instantes de tiempo la monitorización con PCA no es capaz de distinguir el segundo fallo detectado ya que la detección del primer fallo provoca el disparo de los estadísticos Hotelling y SPE por encima de su umbral

Si se quiere monitorizar la misma planta pero con un funcionamiento que recorra diferentes puntos de operación hay que recoger datos de los diferentes modos de operación para que el modelo PCA tenga en cuenta las distintas correlaciones de variables que pueden aparecer. Si se monitoriza este tipo de funcionamiento se puede comprobar que los diferentes transitorios entre los modos disparan los estadísticos el suficiente número de muestras para que sean detectados como fallos.

La detección de fallos se muestra eficiente para sistemas de control donde no hay grandes cambios en el modo de operación. Aparecen diferentes alarmas que por su corta duración no pueden ser confundidas con fallos positivos.

Cuando el proceso está bajo control, el SPE representa fluctuaciones de ruido. Si ocurriese un evento inusual que afectase a la covarianza de X , se detectaría porque la estadística Q superaría el umbral.

4.8. Bibliografía

- [Cam-07] Camacho, O. (2007). Fault Diagnostic base on multivariate statistical techniques. *Revista técnica de la Universidad de Zulia*. Vol. 30, No 3, pp. 253-262.
- [Chi-00] Chiang, L. (2000). *Fault detection and diagnosis in Industrial Systems*, Springer, Nueva York.
- [Doa-04] Doan, T., Khiang, L., y Liu, J. (2004). Comparative study of pca approaches in process monitoring and fault detection. *The 30th annual conference of the IEEE industrial electronics society*, pages 2594-2599.
- [Gar-11] García, D. y Fuentes, M. (2011). Estudio comparativo de técnicas de detección de fallos basadas en el análisis de componentes principales. *Revista Iberoamericana de Automática e Informática Industrial*. Elsevier.
- [Kre-91] Kresta, J. (1991). *Multivariate Statistical Monitoring of Process Operating Performance*. *The Canadian Journal of Chemical Engineering*.
- [Jac-91] Jackson, J. (1991). *A user's guide to principal components* . Wiley.
- [Kui-95] Ku, W., Storer, H. y Geogakis, Ch. (1995). *Disturbance Detection and Isolation by Dynamic Principal Component Analysis*. *Chemometrics and Intelligent Laboratory Systems*.
- [Mas-02] Mason, R., Young, J. (2002). *Multivariate Statistical Process Control with Industrial Applications*. ASA SIAM Series.
- [Min-06] Mina, J. y Verde, C. (2006). *Proceso de Estandarización Adaptable para Detección de Fallas en Sistemas MIMO*. *Congreso Nacional de Control Automático*. UNAM México DF.
- [Pen-02] Peña, D. (2002). *Análisis de Datos Multivariantes*. Mc Graw Hill.
- [Rin-12] Rincón, A. (2012). Multiple fault detection and diagnosis in a Gas Turbine using principal component analysis and structured residuals. *20th Mediterranean Conference on Control & Automation (MED 2012)*.
- [Shl-05] Shlens, J. (2005). *A Tutorial on principal component analysis*. La Jolla, CA 92037: Salk Institute for Biological Studies.
- [Wei-01] Weighell, M. (2001). The statistical monitoring of a complex manufacturing process. *Journal of applied statistics*.

CAPITULO 5:

DETECCION DE FALLOS MULTIPLES EN TURBINAS DE GAS USANDO ANALISIS DE COMPONENTES PRINCIPALES NO LINEAL (NLPCA)

Contenido

- 5.1. Introducción
- 5.2. Estado del arte del NLPCA
- 5.3. Análisis de Componentes Principales no Lineales
- 5.4. Algoritmo de la curva principal
- 5.5. Mapeo de la red
- 5.6. Demapeo de la red
- 5.7. Detección de fallos múltiples usando NLPCA.
- 5.8. Resultados
- 5.9. Bibliografía

5.1. Introducción.

Una alternativa conocida como análisis de componentes principales no lineal (NLPCA), se ha desarrollado para evitar las limitaciones presentadas por el PCA. Una primera versión de este método fue descrito por Guttman en 1941, y otras importantes contribuciones a la literatura sobre este tema son de [Kru-65], [She-66], [You-78], y [Win-83]. Para una revisión histórica, ver [Gif-90].

El método de NLPCA tiene los mismos objetivos que el análisis tradicional de componentes principales, pero es adecuado para las variables de distintos niveles de medición (nominal, ordinal y numérico), que pueden estar o no linealmente relacionadas entre sí. En el tipo de PCA no lineal que se describe en el presente capítulo, todas las variables son vistas como categóricas, y cada valor distinto de una variable se le conoce como una categoría. En consecuencia, el método también se conoce como PCA categórico.

Este capítulo proporciona una introducción del método NLPCA. En la primera sección, se discute el estado del arte del análisis de componentes principal no lineal, el NLPCA basado en redes neuronales y el método de detección y aislamiento de fallos múltiples. La solución del NLPCA se compara con la solución del PCA con los mismos datos. En la sección final se resumen los aspectos más importantes del NLPCA, centrándose en sus fortalezas y limitaciones como método de análisis exploratorio de datos.

5.2. Estado del arte del NLPCA

En los últimos años el análisis de componentes principales (PCA) se ha utilizado en el área de control de procesos estadísticos tales como la supervisión de procesos [Kre-91], la detección de errores graves [Ton-95], la identificación de fallos en sensores [Dun-96]. Sin embargo, el análisis de componentes principales es un método lineal, y la mayoría de los problemas de ingeniería son no lineales.

Hastie y Stuetzel proponen la metodología de la curva principal para proporcionar un abreviado conjunto m -dimensional de datos no lineales [Has-89]. Sin embargo, este enfoque es no paramétrico y no se puede utilizar para el análisis continuo de nuevos datos. Para superar el problema de parametrización se utiliza una red neuronal auto-asociativa [Kra-91], [Don-94], [Tan-95]. Kramer propuso un análisis no lineal de componentes principales (NLPCA) sobre la base de una red neuronal de cinco capas auto-asociativa para el descubrimiento de las relaciones lineales y no lineales entre las variables [Kra-91]. Esta red neuronal tiene la primera capa (entrada) y la quinta capa (salida) lineales, y la segunda y cuarta capa con no linealidades sigmoidales. Sin embargo, la presencia de múltiples mínimos locales en la función de peso y de optimización pueden llevar a problemas de estabilidad en la NLPCA.

Dong y MacAvoy [Don-94] desarrollaron un NLPCA basados en el algoritmo de las curvas principales de Hastie y dos redes neuronales de tres capas. Un entrenamiento supervisado se utiliza para la determinación de los pesos de cada red, donde la salida deseada de la primera red y la entrada de la segunda representan los componentes principales dados por el algoritmo de Hastie, se observará que este algoritmo es predispuesto.

Tan y Mavrouniotis [Tan-95] propusieron una técnica de reducción de dimensionalidad no lineal basada en la optimización de las entradas de la red neuronal. La idea básica es la de reducir las cinco capas de la red auto-asociativa a una red de tres capas simultáneamente, para la optimización de los parámetros de la red y las entradas de la red que representan los componentes principales, reduciendo al mínimo el error de salida.

Este enfoque sólo considera la sección de reducción de componentes del NLPCA e implementa una búsqueda no lineal para el cálculo de los componentes principales no lineales, pero requiere más cálculo en la aplicación debido a la necesidad de un optimizador no lineal en línea. Para este tipo de redes, sin embargo, el problema de entrenamiento se convierte en una tarea complicada de optimización no lineal, que se define en un espacio multidimensional.

Webb [Web-96] propuso un método para el análisis no lineal de componentes principales utilizando una red neuronal con función de base radial (RBF). Dos redes RBF se utilizan en este enfoque, la primera un proyecto conjunto de datos de dimensión reducida usando una transformación no lineal cuyos parámetros están determinados por la solución de una ecuación simétrica generalizada de los vectores propios, maximizando la varianza de sus salidas.

5.3. Análisis de Componentes Principales No Lineal (NLPCA).

Mientras PCA identifica las relaciones lineales entre las variables del proceso, el objetivo del NLPCA es extraer las relaciones lineales y no lineales. Esta generalización se consigue mediante la proyección de las variables del proceso con relación a curvas o superficies en lugar de líneas o planos. La figura 5.1 ilustra el concepto de PCA lineal. El primer componente principal minimiza la suma de los cuadrados de las desviaciones ortogonales entre la línea recta y todas las variables. El concepto de NLPCA se ilustra en la figura 5.2. El enfoque no lineal es similar al PCA lineal, excepto que proyecta los datos contra una curva suave dimensional que se determina por la relación no lineal entre las variables. La curva se define para minimizar las desviaciones ortogonales entre los datos y la curva. Los modelos de NLPCA pueden ser representados por dos submodelos (modelos de mapeo y de demapeo). En la figura 5.3 se muestra el resultado producido por el PCA versus el NLPCA.

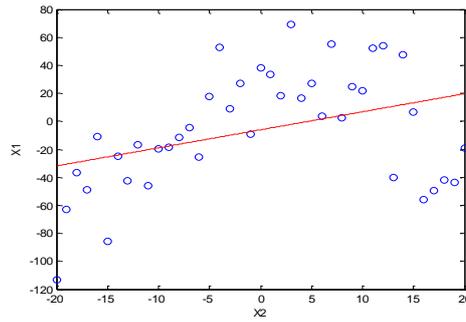


Figura 5. 1. El PCA minimiza la suma de los cuadrados de la desviación ortogonal usando una línea recta

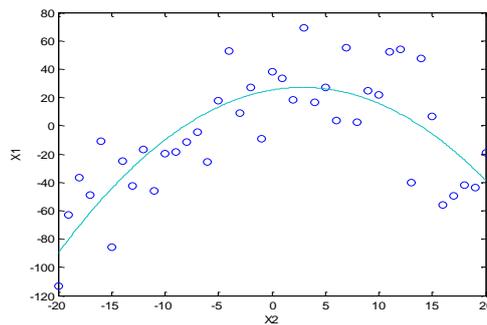


Figura 5. 2. El NLPCA minimiza la suma de los cuadrados de la desviación ortogonal utilizando una curva

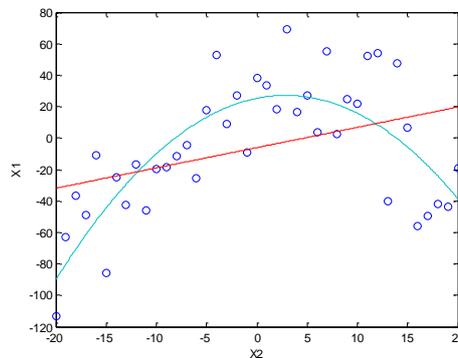


Figura 5. 3. PCA versus NLPCA para los mismos datos

La función de mapeo proporciona a partir de la matriz de datos X los componentes principales no lineales U y estimación de la función de demapeo. En este caso, la función de mapeo no lineal tiene la siguiente forma:

$$z = G(x) \tag{5.1}$$

Donde x y z son las filas de X y Z respectivamente, G es la función de mapeo. La función de demapeo da la estimación de \hat{x} a partir de x usando los componentes principales no lineales z y que tiene la forma:

$$\hat{x} = F(z) \tag{5.2}$$

Donde F representa la función de demapeo.

Por lo tanto, un conjunto de datos X, incluyendo m variables puede ser expresado en términos de componentes principales no lineales, es decir,

$$X = \hat{X} + E = F(Z) + E \tag{5.3}$$

Donde $Z = [Z_1, \dots, Z_l]$ es la matriz de componentes principales no lineal $Z = G(X)$, y E es la matriz de los residuos. El problema consiste en identificar la proyección no lineal de las funciones G y F. Para hacer esto, la función objetivo a minimizar es la suma de cuadrados de las desviaciones ortogonales:

$$\min \sum_{i=1}^N \|x_i - \hat{x}_i\|^2 = \min \sum_{i=1}^N \|x_i - F(G(x_i))\|^2 \tag{5.4}$$

Donde x_i es la fila i-esima de X y \hat{x}_i es la estimación por las cinco capas de la red neuronal del modelo NLPCA. Donde G es la función de mapeo y F representa la función de demapeo.

Dada una matriz $N \times m$, donde N representa las mediciones de m variables, la reducción de la dimensionalidad de los datos transforma una matriz de datos originales en una matriz de dimensión mucho más pequeña con una distorsión mínima y luego se realiza una proyección de demapeo. La matriz reducida describe las variables de componentes principales extraídos de la matriz inicial (ver Figura 5.4).

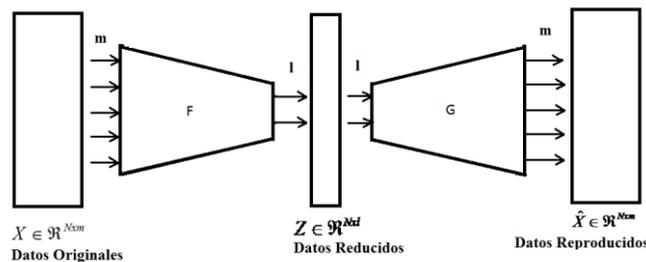


Figura 5.4. Reducción de dimensionalidad de los datos

Para realizar NLPCA, la red neuronal en la figura 5.5 contiene 3 capas ocultas de neuronas entre la entrada y las capas de salida de variables.

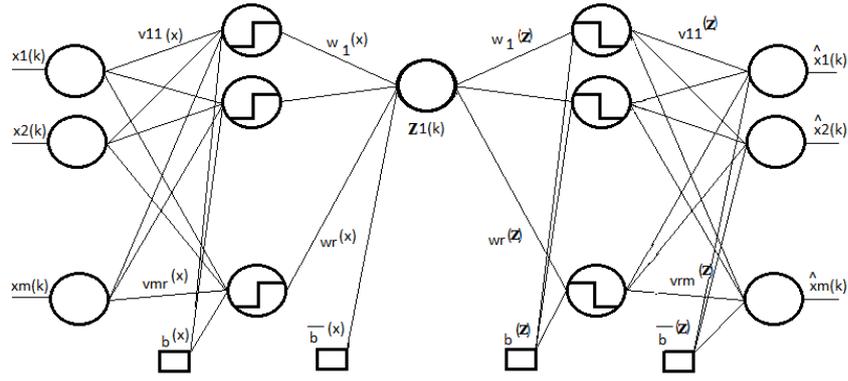


Figura 5. 5. Uso de una red neuronal para la extracción de los componentes principales no lineales

Una función de transferencia $G1$ mapea a partir de \mathbf{x} , el vector de entrada tipo columna de longitud m , a la primera capa oculta, representada por $\mathbf{h}(\mathbf{x})$ y un vector columna de longitud r , con elementos:

$$h_j^{(x)} = G_1((V^{(x)}(\mathbf{x}) + \mathbf{b}^{(x)})) = G_1\left(\sum_{i=1}^m v_{ij}^x x_i + b_j^{(x)}\right) \quad (5.5)$$

Donde con letra mayúscula y negrita se reserva para las matrices y con letra minúscula en negrita se reserva para los vectores. $V^{(x)}$ es una matriz de peso de tamaño $(r \times m)$. $\mathbf{b}^{(x)}$ es un vector columna de longitud r que contiene los parámetros de tendencia, $j = 1, \dots, r$.

Los componentes principales no lineales \mathbf{z} están dados por:

$$\mathbf{z} = \mathbf{w}^{(x)} \mathbf{h}^{(x)} + \bar{\mathbf{b}}^{(x)} = \sum_{j=1}^r w_j^{(x)} h_j^{(x)} + \bar{b}^{(x)} \quad (5.6)$$

Entonces, la función de mapeo G está definido por:

$$G(\mathbf{x}) = \sum_{j=1}^r w_j^{(x)} h_j^{(x)} + \bar{b}^{(x)} \quad (5.7)$$

Luego, una función de transferencia $F1$ mapea a partir de \mathbf{z} hacia el final de la capa oculta $\mathbf{h}^{(z)}$, la cual es un vector columna de longitud r , con elementos:

$$h_j^{(z)} = F_1((w^{(z)} \mathbf{u} + \mathbf{b}^{(z)})_j) = F_1(w_j^{(z)} u + b_j^{(z)}) \quad (5.8)$$

Y la función de demapeo F está dada por:

$$F(z) = \sum_{j=1}^r v_{ji}^{(z)} h_j^{(z)} + \bar{b}_i^{(z)} \quad (5.9)$$

Las funciones de transferencia G_1 y F_1 son generalmente no lineales (usualmente son funciones tangente hiperbólicas, funciones sigmoides, la función exacta que se use no es crítica).

La salida de la red neuronal \hat{x} representa el vector de estimación de los datos originales x , la cual está dada por:

$$\hat{x} = F(k) = F(G(x)) \quad (5.10)$$

La función de costo $E = \|x - \hat{x}\|^2$ se minimiza al encontrar los valores óptimos de $V^{(x)}, b^{(x)}, w^{(x)}, \bar{b}^{(x)}, w^{(z)}, b^{(z)}, V^{(z)}$ y $\bar{b}^{(z)}$. El error medio al cuadrado (MSE) entre la salida de la red neuronal \hat{x} y los datos originales x es también minimizado.

5.4. Algoritmo de la Curva Principal

Hastie y Stuetzle [Has-89] generalizaron la propiedad de autoconsistencia de componentes principales e introdujeron el concepto de la curva principal. Considerando un vector de dimensión m y aleatorio $x = (x_1, \dots, x_m)$ y sea $F = (F_1(k), \dots, F_m(k))$ una curva suave en \mathfrak{R}^m parametrizada para $k \in \mathfrak{R}$. Para cualquier $x \in \mathfrak{R}^m$ donde $k_F(x)$ denota el valor del parámetro k para el cual se minimiza la distancia entre x y $F(k)$. Por la definición de Hastie y Stuetzle, $F(k)$ es una curva principal, si no se cruzan en sí misma y es auto consistente, tal que:

$$F(k) = \mathcal{E}(x | k_F(x) = k) \quad (5.11)$$

Donde k_F está definido como el índice de proyección de $\mathfrak{R}^m \rightarrow \mathfrak{R}$

$$k_F = \sup_k \left\{ k : \|x - F(k)\| = \inf_{\tau} \|x - F(\tau)\| \right\} \quad (5.12)$$

Autoconsistencia significa que cada punto de la curva G es la media en la distribución de x de todos los puntos que son proyectados ortogonalmente sobre la curva según se observa en la Figura 5.6. El enfoque de la curva principal ha sido presentado como una técnica de generalización dimensional no lineal de PCA. Un método de dos dimensiones o de superficie principal también ha sido derivado. Sin

embargo, por problemas de orden de dimensiones superiores, la parametrización de la hipersuperficie es mucho más compleja. Por consiguiente, la extensión del algoritmo de curva principal a altas problemas dimensionales fue propuesto por LeBlanc y Tibshirani [Leb-94]. En este caso, X puede ser expresado como:

$$X = F_1(T_1) + F_2(T_2) + \dots + F_l(T_l) + E \quad (5.13)$$

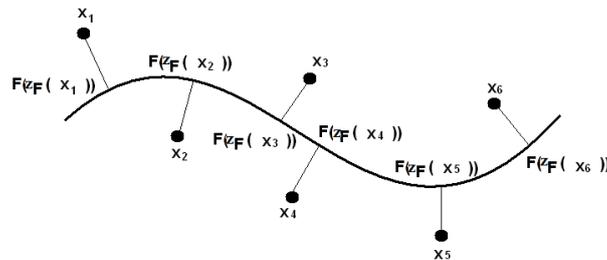


Figura 5. 6. Proyección de los puntos a la curva

El algoritmo de Hastie y Stuetzle está sesgado en lugares de gran curvatura y la convergencia no está asegurada. Sin embargo, ningún comportamiento de divergencia de este algoritmo se ha observado en la práctica. Otro enfoque [Keg-99] es definir curvas principales de longitud l en forma de curvas de longitud l que logren un mínimo esperado cuadrado de la distancia desde el punto a su proyección en la curva.

El algoritmo de la línea poligonal (PLA) encuentra curvas principales en términos de la definición más adelante. Recientemente, Verbeek et al. [Ver-02] proponen un algoritmo para encontrar las curvas principales mediante el uso de un modelo lineal local. Los modelos lineales locales se ajustan mediante un esquema iterativo de actualización. Después del montaje, los modelos lineales locales se combinan para formar unas líneas poligonales por medio de una búsqueda heurística.

5.5. Mapeo de la Red

Considere red ilustrada en la figura 5.7. La intención es usar esta red para definir la transformación dada por la ecuación:

$$G : \mathfrak{R}^m \rightarrow \mathfrak{R} \quad (5.14)$$

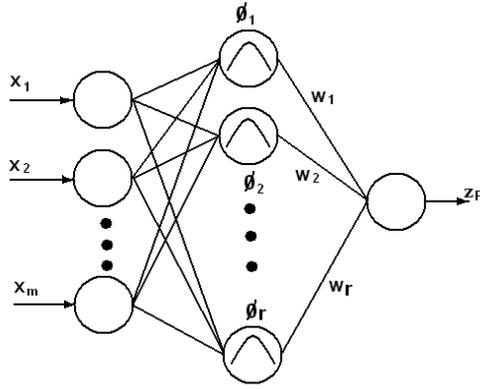


Figura 5. 7. Red neuronal para el mapeo

$$G(X) = \sum_{i=1}^r w_i \phi_i(x) \tag{5.15}$$

Donde $\{\phi_i, i = 1, \dots, r\}$ es una función no lineal de $x \in \mathfrak{R}^m$ y $\{w_i, i = 1, \dots, r\}$ denota un conjunto de parámetros de peso de la capa de salida. Las funciones de base gaussianas ϕ_i son definidas como:

$$\phi_i(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right) \tag{5.16}$$

Donde c_i y σ_i^2 denotan los centros y dispersiones respectivamente.

En el primer paso, los centros c_i son inicializados con la agrupación de las K-medias y las dispersiones σ_i^2 se determinan como la distancia entre c_i y el más cercano c_j ($j \neq i, j \in \{1, \dots, r\}$).

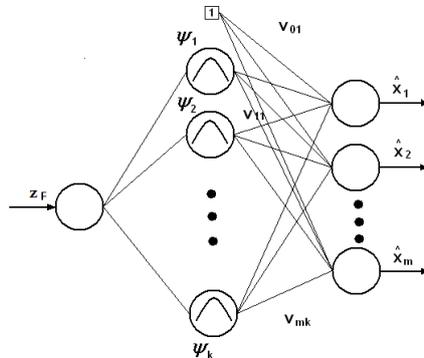


Figura 5. 8. Red neuronal para demapeo

Entonces, en los pasos siguientes, se realiza un descenso de gradiente de la función de error.

$$E = \frac{1}{2} \sum_{k=1}^N (z_k - G(x_k))^2 \quad (5.17)$$

Tomando la derivada de la función de error con respecto al centro c_i y dispersión σ_i se obtiene:

$$\frac{\partial E}{\partial c_i} = \sum_{k=1}^N (G(x_k) - t_k) \frac{\partial G(x_k)}{\partial c_i} \quad (5.18)$$

$$\text{Con } \frac{\partial G(x_k)}{\partial c_i} = w_i \frac{x_k - c_i}{\sigma_i^2} \phi_i(x_k) \quad \text{y}$$

$$\frac{\partial E}{\partial \sigma_i} = \sum_{k=1}^N (G(x_k) - t_k) \frac{\partial G(x_k)}{\partial \sigma_i}$$

$$\text{Con } \frac{\partial G(x_k)}{\partial \sigma_i} = w_i \frac{\|c_i - x_k\|}{\sigma_i^3} \phi_i(x_k). \text{ Estas dos derivadas son:}$$

$$w = (\Phi^T \Phi)^{-1} \Phi^T Z \quad (5.19)$$

Donde la i -ésima columna de Φ^T es $\phi(x_i)$ y $\phi(\cdot) = (\phi_1, \dots, \phi_r)$, $w^T = (w_1, \dots, w_r)$ y la i -ésima columna de Z^T es z_i .

5.6. Demapeo de la Red

El algoritmo de la sección anterior nos proporciona un método de proyección de datos en un espacio de dimensión reducida. Mediante el uso de la segunda red se define la proyección inversa de los datos originales, y por lo tanto determinar la forma real de la curva principal. Se define una transformación.

$$\hat{x} = F(z) = \sum_{j=1}^k v_j \psi_j(z) + v_o \quad (5.20)$$

Para algunos Kernels $\psi_j, (j=0, \dots, k)$, pesos $V = (v_o, \dots, v_k)$, donde k es el número de Kernels y $v_i \in \mathfrak{R}^l, (i=0, \dots, k)$. Se busca una solución para los pesos que minimizan.

$$E = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|^2 \quad (5.21)$$

Donde $\{x_i = i = 1, \dots, N\}$ es el conjunto de datos dado. Así \hat{x}_i es la proyección de x_i en la curva de ajuste (o de superficie) y se busca la curva que mejor ajuste usando mínimos cuadrados. Usando (5.17), la expresión (5.18) puede ser escrita como:

$$E = \frac{1}{N} \sum_{i=1}^N \left\| x_i - \sum_{j=0}^k v_j \psi_j(z_i) \right\|^2 \quad (5.22)$$

Dónde $\psi_0 = 1$. La solución para el argumento de la matriz $m \times (k + 1)$ que minimiza (5.19) está dada por:

$$V^T = (\Psi^T \Psi)^{-1} \Psi^T X \quad (5.23)$$

5.7. Detección de fallos múltiples usando NLPCA

El método propuesto consiste en tres pasos:

1) Encontrar la curva principal aplicando sucesivamente el algoritmo de curva principal a los datos observados y residuales [Has-89]. Luego, en el primer paso se puede utilizar Z_1 para representar el primer componente principal no lineal, por lo que: $X = F_1(Z_1) + E_1$, donde E_1 es el residual. Cuando más de un componente principal no lineal es necesario, se hace el mismo cálculo para los datos residuales como se muestra:

$$E_{i-1} = F_i(Z_i) + E_i \quad (5.24)$$

2) Entrenar una red neuronal para que relacione los datos originales y los transforme a componentes principales no lineales obtenidos del algoritmo de curva principal.

3) Entrenar una segunda red neuronal que relacione los componentes principales no lineales y los lleve a los datos originales.

El modelo de análisis de componentes principales no lineal se puede conseguir mediante el uso de dos redes neuronales para el mapeo y demapeo. En primer lugar, una red de tres capas se utiliza; la capa oculta se compone de neuronas de base radial realizando un mapeo no lineal del espacio de entrada en un espacio de dimensión inferior, de forma que las características no lineales son capturadas en el sentido de minimizar la suma de cuadrados de error de sus salidas. Al conservar la dimensión original de los datos, la segunda red trata de realizar la transformación inversa o reproducción de los datos originales, minimizando el error de predicción al cuadrado entre las muestras de datos originales y sus

salidas correspondientes. Las dos redes están formadas por separado y las salidas de la primera son las entradas de la segunda.

5.7.1. Detección y aislamiento de fallos múltiples

Las situaciones anormales que se producen debido a desviaciones en los sensores inducen cambios en las mediciones de los mismos. El NLPCA se utiliza para modelar el comportamiento normal de un proceso y los fallos son detectados al comparar las mediciones actuales del proceso contra el modelo.

Consideremos un nuevo vector de datos, $x(k)$ el cual puede ser representada como:

$$x(k) = \hat{x}(k) - e(k) \quad (5.25)$$

Donde $\hat{x}(k) = F(G(x(k)))$ es el vector de estimación, y el vector de estimación de error está dado por:

$$e(k) = x(k) - \hat{x}(k) \quad (5.26)$$

En el caso de cero fallos, la expectativa del vector residual es poco significativo o pequeño debido al ruido de medida y del comportamiento no modelado. En la presencia de fallos, la expectativa del vector residual ya no es poco significativo o pequeño y el fallo afecta a todos los componentes del vector de residuos. La detección de fallos en los sensores usando NLPCA se realiza mediante el control de los residuos. El error de predicción al cuadrado (SPE) es una estadística que mide la falta de ajuste del modelo a los datos del modelo NLPCA. En el instante k , el índice de detección SPE viene dado por:

$$SPE(k) = e^T(k)e(k) = \sum_{j=1}^m e_j^2(k) \quad (5.27)$$

Esta cantidad sugiere la existencia de una situación anormal en los datos cuando:

$$SPE(k) > \delta_\alpha^2 \quad (5.28)$$

Donde δ_α^2 es el límite de control para el SPE estimado usando los datos históricos. Para mejorar la calidad de la detección mediante la reducción de la tasa de falsas alarmas debido al ruido, se puede aplicar a los residuos un filtro de promedio móvil ponderado exponencialmente (EWMA). Así pues, los residuos filtrados se obtienen como:

$$\bar{e}(k) = (1 - \gamma)\bar{e}(k-1) + \gamma e(k) \quad (5.29)$$

Donde $\bar{e}(k)$ es el vector filtrado al tiempo k y γ es el factor de olvido escogido en un intervalo de $[0,1]$. Los límites de control para el SPE filtrado, \overline{SPE} , (δ^2_α) son calculados según propuesto por Dunia [Dun-96].

La figura 5.9 muestra los residuos obtenidos de un proceso usando el método PCA, mientras que la figura 5.10 muestra los residuos para los mismos datos pero usando NLPCA. La figura 5.11 muestra la comparación de los dos métodos.

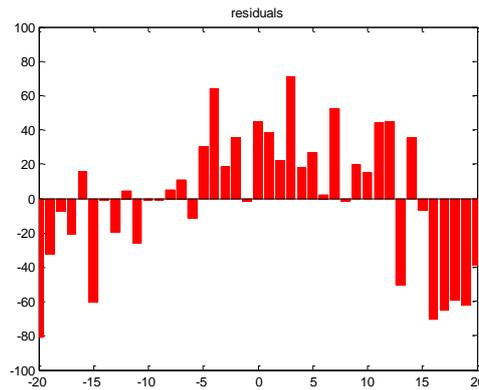


Figura 5. 9. Residuos usando PCA

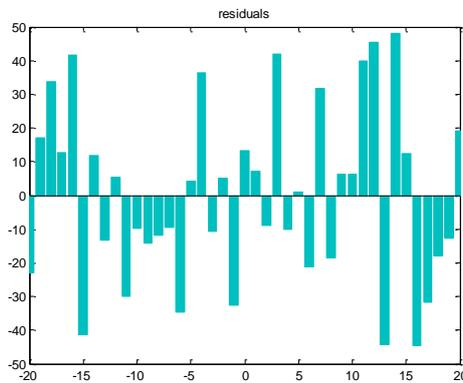


Figura 5. 10. Residuos usando NLPCA

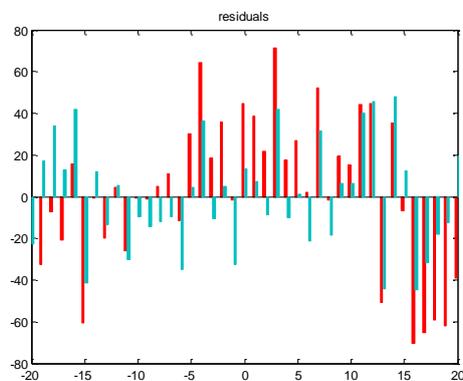


Figura 5. 11. Comparación de los residuos usando PCA versus NLPCA

5.8. Resultados:

A continuación se muestra la metodología del análisis de componentes principales no lineal para la detección y diagnóstico de fallos múltiples en una turbina de gas SR-30 ubicada en la Universidad Interamericana de Puerto Rico, Recinto de Bayamón [Rin-13].

5.8.1. Detección de fallos múltiples usando NLPCA.

Para demostrar el método propuesto se llevaron a cabo medidas de los diferentes sensores de la turbina SR-30 a diferentes condiciones de operación: 55.000, 65.000 y 75.000 revoluciones por minutos.

Inicialmente se obtuvieron medidas en los sensores de temperatura, presión, rpm y torque en la turbina de gas SR-30 sin la presencia de fallos. Hay que destacar que para el análisis no se consideran las medidas de temperatura y presión de entrada en el compresor por ser constantes.

Todas las mediciones se estandarizan con relación a la varianza y media de los datos. Luego aplicando el algoritmo de la curva principal para los datos tomados a 55.000, 65.000 y 75.000 respectivamente se obtuvieron las gráficas 5.12 y 5.13. Las graficas muestran los dos primeros componentes principales no lineales y la curva mostrada minimiza la distancia de los NLPCA.

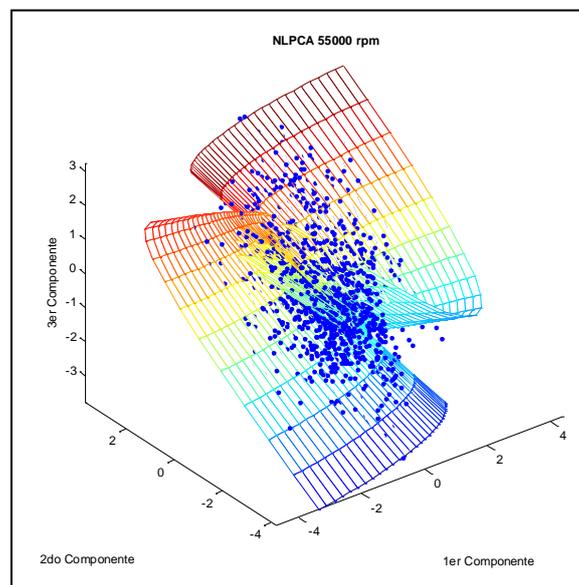


Figura 5. 12. NLPCA aplicando la curva principal a 55000 rpm.

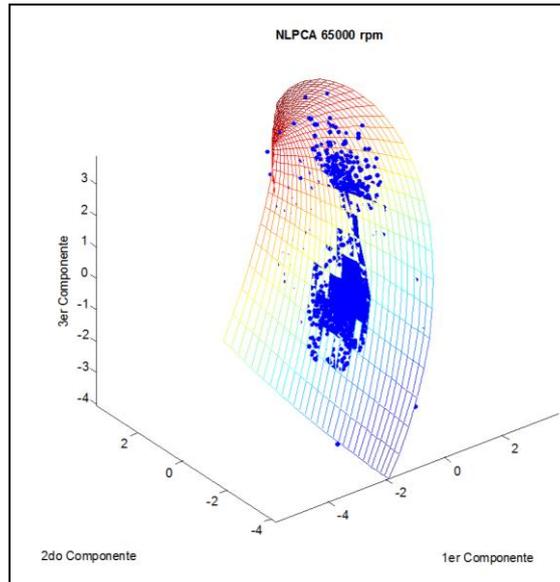


Figura 5. 13. NLPCA aplicando la curva principal a 65000 rpm.

Al aplicar el algoritmo de la curva principales se obtienen componentes principales con una distancia mínima a una curva determinada. Con la red neuronal de mapeo se obtienen los dos primeros componentes principales y mediante el uso de el error de predicción al cuadrado se puede realizar la detección de fallos múltiples según lo muestran las siguientes figuras.

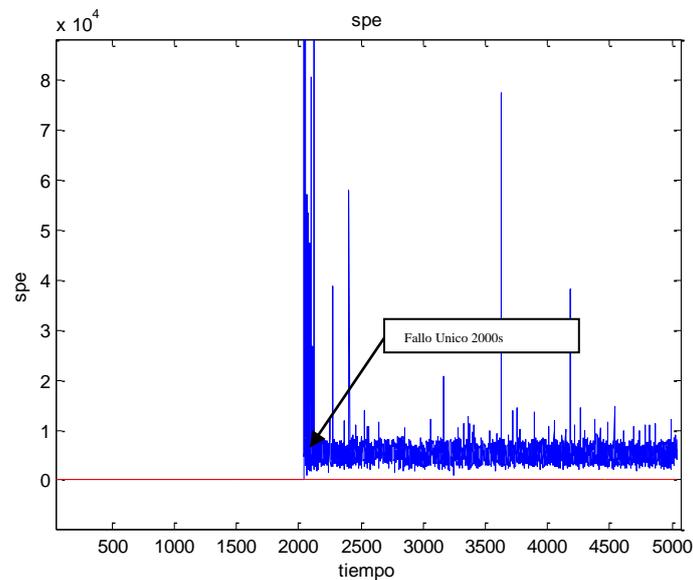


Figura 5. 14. SPE para un fallo de 10% ET

La figura 5.14 muestra el error de predicción al cuadrado SPE para cuando existe en el sistema un fallo de 10% en la eficiencia de la turbina. El fallo típico en las turbinas de gas que hacen que disminuya la eficiencia generalmente son por desgaste en los alabes.

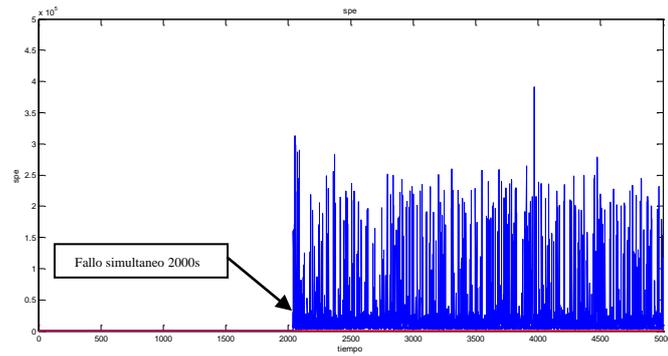


Figura 5. 15. SPE para fallos múltiples simultaneo de 10%ET y 8%EC

La figura 5.15 muestra el error de predicción al cuadrado SPE para cuando existe en el sistema un fallo de 10 % en la eficiencia de la turbina y simultáneamente a los 2000 segundos de presenta un fallo de 8% en la eficiencia del compresor.

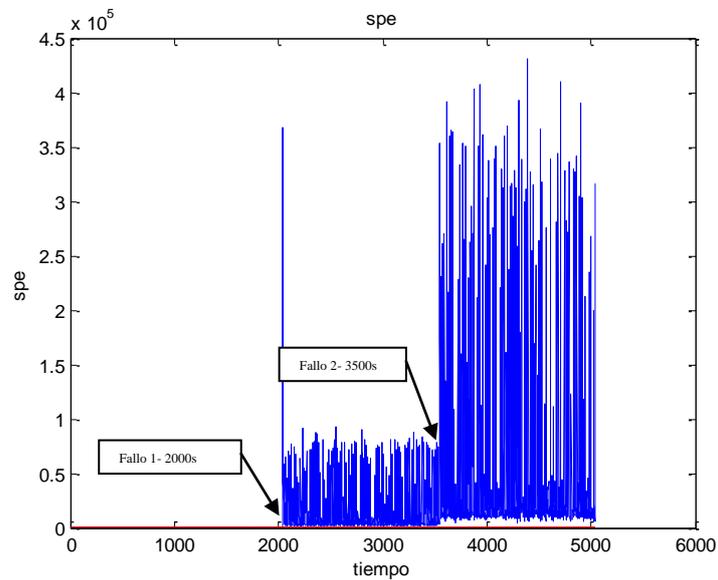


Figura 5. 16. SPE para fallos múltiples secuenciales de 10%ET y 8%EC

La figura 5.16 muestra el error de predicción al cuadrado SPE para cuando existe en el sistema un fallo de 10% en la eficiencia de la turbina y simultáneamente se presenta a los 2000 segundos un fallo de 8% en la eficiencia del compresor.

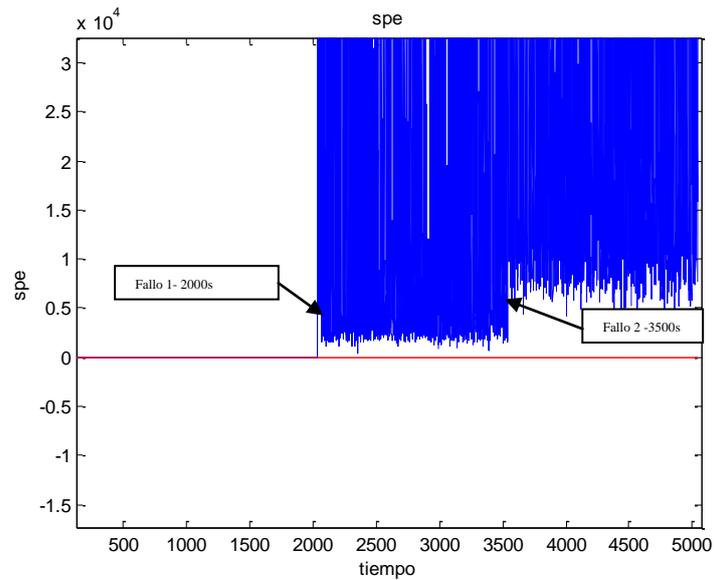


Figura 5. 17. SPE para fallos múltiples secuenciales de 10%ET y 8%EC

La figura 5.16 muestra el error de predicción al cuadrado SPE para cuando existe en el sistema un fallo de 10% en la eficiencia de la turbina y secuencialmente se presenta a los 3500 segundos un fallo de 8% en la eficiencia del compresor.

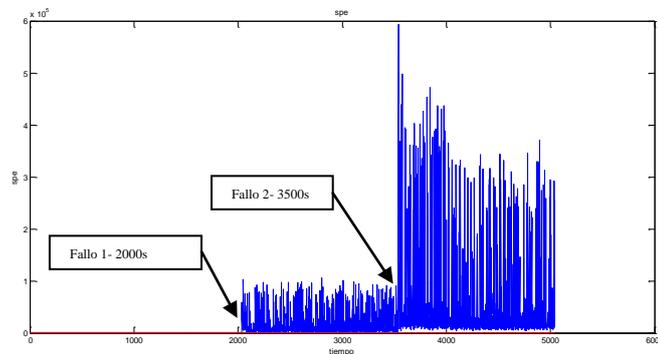


Figura 5. 18. SPE para fallos múltiples secuenciales de 10%ET, 8%EC y 5%WT

La figure 5.18 muestra el error de predicción al cuadrado SPE cuando se presentan tres fallos secuenciales en el sistema. El primer fallo de 10% en la eficiencia en la turbina se presenta a los 2000 segundos, el segundo fallo de 8% en la eficiencia en el compresor se presenta a los 3500 segundos y el tercer y último fallo de un 9% en la masa de combustible en la turbina se presenta a los 4500 segundos.

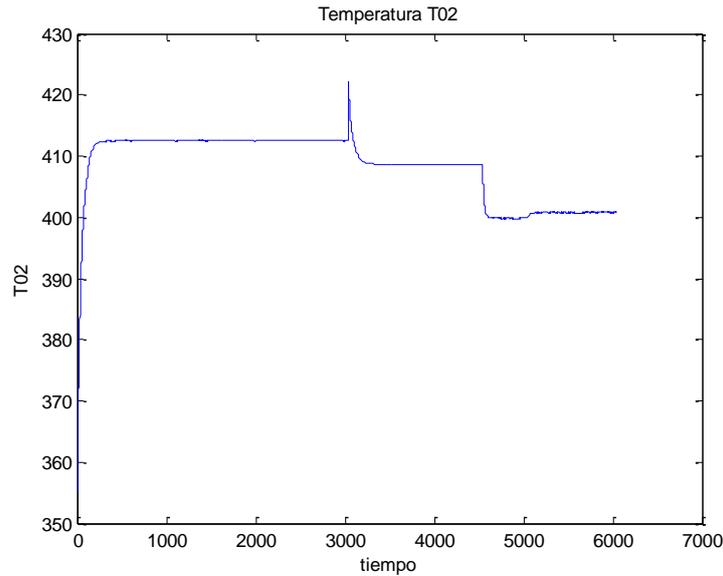


Figura 5. 19. Cambio en T02 cuando existen 3 fallos secuenciales

Las figuras 5.19 y 5.20 muestran el comportamiento de la variable temperatura y RPM cuando ocurren los tres fallos secuenciales expresados en la figura 5.18.

Cuando se presentan fallos en sensores solamente se nota un cambio en las medidas del sensor o sensores afectados mientras que el resto de los sensores sin fallo continúan midiendo sin cambio alguno. Para el caso de fallos en componentes y actuadores las medidas de los sensores cambian todas y es más difícil detectar a simple vista el fallo.

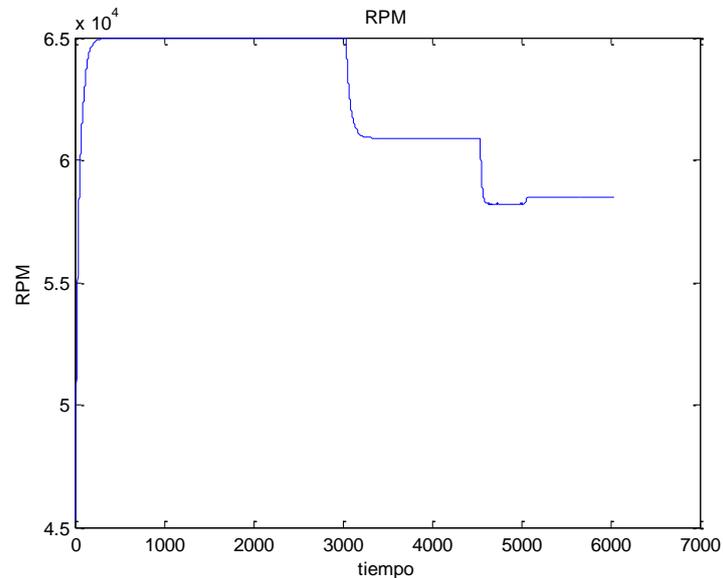


Figura 5. 20. Variación de las rpm cuando existen tres fallos secuenciales

En esta sección solo de detalle en los resultados la parte de detección de fallos mediante el uso de la primera red neuronal de mapeo en la topología propuesta [Figura 5.7]. El capítulo 6 detalla el uso de

las segunda red neuronal de demapeo se aplica para poder generar los residuos en la fase de aislamiento de fallos simples y múltiples.

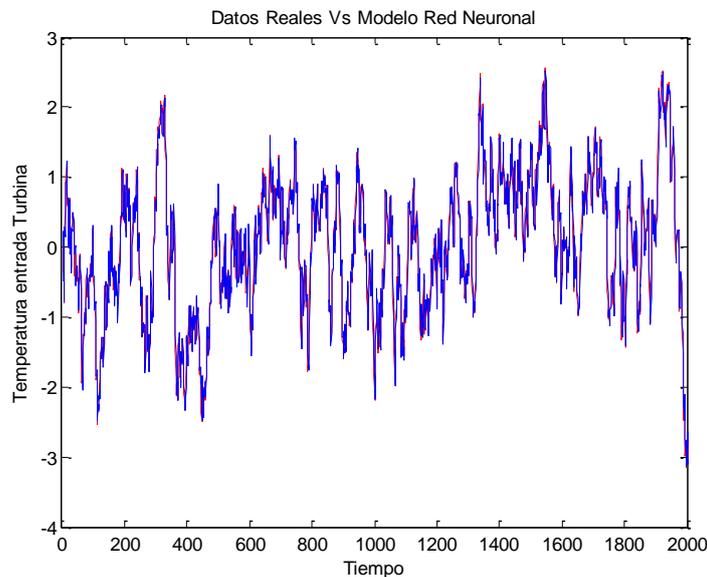


Figura 5. 21. Validación de la red neuronal de mapeo

La figura 5.21 muestra la validación de la red neuronal de demapeo usada para reproducir el comportamiento de la turbina sin fallo. Para la red se usaron datos de entrenamiento y luego se verifico mediante datos de validación, dando un muy buen resultado.

5.11. Conclusiones

En el presente estudio, se propuso un método original para reducir la dimensión del gran conjunto de datos de sensores en una turbina de gas y la obtención de los componentes principales de ellos. Se utilizó una red neuronal en cascada con un cuello de botella estructurado para obtener la reducción de dimensión, dando los componentes principales de la variabilidad de los datos. La misma estructura también se utiliza para ampliar los datos obtenidos de los componentes principales.

El método NLPCA encuentra y elimina las correlaciones no lineales en los datos. Análogamente al análisis de componentes principales, este método se puede utilizar para reducir la dimensionalidad de los datos mediante la eliminación de información redundante. En teoría, este método puede eliminar cualquier tipo de correlación no lineal que se produzca en los datos, ya que las funciones de base utilizadas pueden modelar cualquier función continua y con precisión arbitraria.

El método se aplica para estudiar la detección de fallos múltiples en sistemas dinámicos, en este caso una turbina de gas.

PCA no lineal se puede aplicar a los mismos problemas que el PCA convencional: la reducción de datos y visualización, la validación de sensores, detección de fallos, control de calidad, regresión de componentes principales, etc. Debido a la capacidad de NLPCA para describir los datos no lineales más eficientemente que el PCA es un método en vías de buscar una mayor aplicación en la industria.

Dichas las ventajas del método es bien importante destacar que computacionalmente esta es una metodología que requiere mayor requerimiento computacional y por lo tanto para sistemas o procesos lineales o no lineales que se pueden linealizar el PCA es mejor opción.

En cuanto a costos de aplicación y desarrollo de las técnicas PCA y NLPCA se destaca que la técnica NLPCA es mucho más costosa y requiere de mayor esfuerzo computacional y técnico.

5.12. Bibliografía

- [Don-94] Dong, D. y McAvoy, T. (1994). Nonlinear principal component analysis - based on principal curves and neural networks. Proceedings of the American Control Conference, Baltimore, Maryland, pp. 1284-1288.
- [Dun-96] Dunia, R., Qin, S. y Edgar, T. (1996). Identification of faulty sensors using principal component analysis. AIChE Journal, vol. 42, N. 10, pp. 2797-2812, October 1996.
- [Has-89] Hastie, T. y Stuetzle, W. (1989). Principal curves. Journal of the American Statistical Association, vol. 84, N. 406, pp. 502-516.
- [Kra-91] Kramer, M. (1991). Nonlinear principal component analysis using auto-associative neural networks. AIChE Journal, vol. 37, N. 2, pp. 233-243.
- [Kre-91] Kresta, J., MacGregor, J. y Marlin, T. (1991). Multivariate statistical monitoring of process operating performance. Canadian J. Chem. Eng., vol. 69, N. 1, pp. 35-47.
- [Leb-94] LeBlanc, M. y Tibshirani, R. (1994). Adaptive principal surfaces. Journal of American Statistical Association, vol. 89, N. 425, pp. 53-64.
- [Tan-95] Tan, S. y Mavrovouniotis, M. (1995). Reduction data dimensionality through optimizing neural network inputs. AIChE Journal, vol. 41, N. 6, pp. 1471-1480.
- [Ton-95] Tong, H. y Crowe, C. (1995). Detection of Gross errors in data reconciliation by Principal Component Analysis. Process systems engineering, vol. 41, N. 7, July 1995.
- [Rin-13] Rincón, A. Quevedo, J. (2013). Multiple Fault Detection and Diagnosis in Gas Turbine Using Nonlinear Principal Component Analysis and Structured Residuals. ASME 2013 International Mechanical Engineering Congress & Exposition.
- [Ver-02] Verbeek, J., Vlassis, N. y Krose, B. (2002). A k-segments algorithm for finding principal curves. Pattern Recognition Letters, vol. 23, pp. 1009-1017.
- [Web-96] Webb, A. (1996). An approach to nonlinear principal component analysis using radially symmetric kernel functions. Statist. Comput., vol. 6, pp. 159-168.

CAPITULO 6:

DIAGNOSTICO DE FALLOS MULTIPLES USANDO RESIDUOS ESTRUCTURADOS

Contenido

- 6.1. Introducción
- 6.2. Diagnóstico de fallos
- 6.3. Redundancia analítica
- 6.4. Evaluación de Residuos
- 6.5. Localización de los fallos mediante residuos estructurados
- 6.6. Metodología de diagnóstico
- 6.7. Aplicación a una Turbina de Gas
- 6.8. Resultados
- 6.9. Conclusiones
- 6.10. Bibliografía.

6.1. Introducción

Todo sistema sin excepción está sujeto a cambios en sus parámetros, los cuales modifican el comportamiento para el cual fueron diseñados. Estos cambios pueden ser debidos a efectos de la temperatura, desgastes ocasionados por la fricción, el envejecimiento de los componentes, etc. Los cambios en los parámetros del sistema a zonas fuera de los límites establecidos de acuerdo a criterios de ingeniería son considerados como fallos. También puede haber fallos en los sensores y actuadores que son de tipo aditivo.

La rápida detección de la presencia de fallos en los sistemas puede ayudar a tomar acciones correctivas y de este modo reducir el daño potencial que esta falla puede ocasionar al sistema. Al procedimiento anterior se le conoce con el nombre de reconfiguración.

Tradicionalmente los mecanismos para la detección de fallos son basados en el concepto de redundancia y más específicamente en el uso de redundancia física, es decir, en el uso de elementos repetidos en el sistema. Estos elementos nos permiten, por medio de comparaciones del funcionamiento, tomar decisiones sobre la presencia de fallos y sobre posibles acciones correctivas. Los métodos basados en redundancia física son confiables y permiten una rápida corrección de posibles fallos. Sin embargo debe tenerse claro que la utilización de elementos repetidos en un sistema no puede ser llevada a la práctica siempre. Bajo ciertas condiciones, como por ejemplo, el costo, el tamaño o el peso de los dispositivos, el uso de este tipo de redundancia está limitado.

Técnicas alternativas tuvieron que ser desarrolladas para enfrentar los problemas antes descritos. A principios de la década de los 70 fue introducido el concepto de redundancia analítica, el cual complementa los resultados disponibles de los métodos basados en redundancia física. La redundancia analítica está basada en el conocimiento del modelo matemático del sistema en cuestión así como de las señales de entrada y de salida del sistema.

Actualmente existe una gran variedad de métodos en la literatura, basados en modelos matemáticos del sistema monitoreado y en la teoría de control. En el presente capítulo se presenta la teoría propuesta para el diagnóstico de fallos múltiples usando residuos estructurados.

6.2. Diagnóstico de Fallos

El concepto de diagnóstico de fallos se refiere tanto a la detección como a la localización de uno o varios fallos, es decir, además de poder determinar si uno o varios fallos están presentes se requiere saber que componente o componentes los están ocasionando. Esto último es indispensable para poder determinar una acción correctiva. La tarea de diagnóstico puede realizarse en dos pasos principales: generación de residuos y evaluación de residuos [Fra-90].

Generación de Residuos: Este paso consiste en la obtención de señales que contienen información sobre los fallos solamente. Estas señales son llamadas residuos y en el caso ideal los residuos son cero pero en realidad nunca se puede considerar que un residuo es cero y se adoptará un valor pequeño que supere el efecto del ruido de medida y dinámica no modelizada en el residuo cuando no hay fallos y difieren de cero en la presencia de uno o varios fallos.

Evaluación de Residuos: Este paso pretende la extracción de la información contenida en los residuos tratando de determinar el origen del fallo o fallos múltiples. La evaluación proporciona información sobre el tiempo en el que un fallo ocurrió y sobre el elemento en el cual este fallo está actuando.

6.2. Redundancia analítica

La detección de fallos mediante redundancia analítica, a grandes rasgos, consiste en la comparación de mediciones disponibles del sistema con información de su comportamiento nominal proporcionada por el modelo matemático del sistema o usando medidas reales de un proceso. Debido a que este método de diagnóstico de fallos se basa en el modelo de la planta o proceso, la detección y localización de fallos usando redundancia analítica es también llamada FDI basada en modelo [Sal-04].

Los fallos se detectan estableciendo un umbral fijo o variable sobre una cantidad residual resultante de la diferencia entre mediciones del sistema y los valores estimados de los mismos obtenidos del modelo matemático del sistema. Se puede diseñar un determinado número de residuos de tal manera que cada uno de ellos tenga especial sensibilidad a fallos específicos en distintas partes del sistema. El análisis subsecuente de cada residuo, una vez que un umbral es sobrepasado, determina la localización del fallo.

El diagnóstico de fallos basado en modelo comprende de dos etapas principales, la generación de residuos y la evaluación de residuos o toma de decisiones. En la figura 6.1 se ilustra la estructura básica de un sistema de FDI. Como se observa, la generación de residuos se realiza considerando al sistema en lazo abierto, exista o no un lazo de control para el sistema monitoreado. Esto se debe a que la información necesaria para la FDI basada en modelo son las entradas y salidas del sistema en lazo abierto. De este modo, no es necesario considerar al controlador (cuando este exista) en el diseño de un esquema de diagnóstico de fallos.

La detección de fallos basada en modelos es un método enfocado principalmente al diagnóstico en línea, en el cual, dicho diagnóstico es obtenido durante la operación del sistema. Esto se debe a que la información requerida de entrada y salida del sistema está disponible solo cuando el sistema se encuentra en operación.

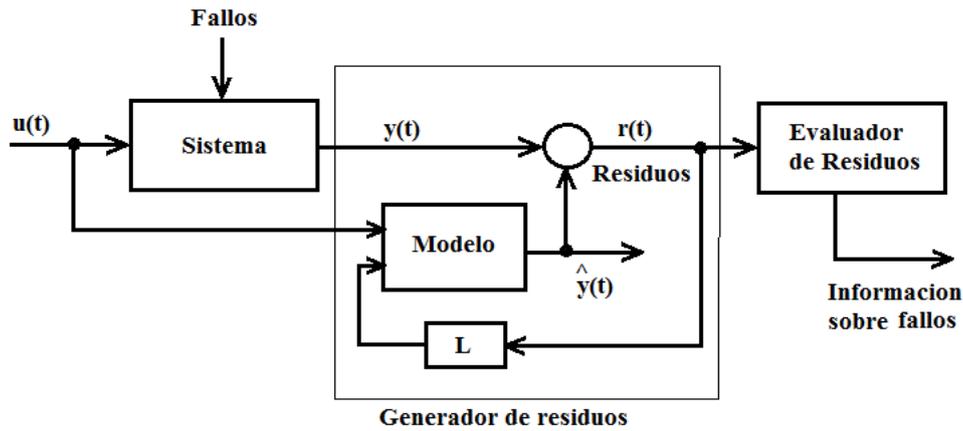


Figura 6. 1. Esquema básico de detección de fallos usando redundancia analítica

6.3. Generación de residuos

Un residuo es una señal generada a partir de mediciones disponibles de entrada y salida del sistema monitoreado, la cual debe ser portadora de información indicativa de fallos. Una forma común de generar residuos es estimar el vector de salida del sistema y . Después la salida estimada \hat{y} es sustraída del valor real medido de y , de este modo cada residuo resulta ser:

$$r_i = y_i - \hat{y}_i \quad i \in \{1, 2, \dots, p\}; y, \hat{y}, r \in \mathcal{R}^p. \quad (6.1)$$

Otra forma similar de generar el vector de residuos es a partir de la diferencia del valor nominal de un determinado vector de parámetros Θ_{nom} y su estimado $\hat{\Theta}$, así el vector de residuos es:

$$r_{\Theta} = \Theta_{nom} - \hat{\Theta}. \quad (6.2)$$

Es conveniente utilizar estos residuos cuando los fallos se pueden caracterizar como una variación no tolerada en uno de los parámetros.

Los residuos se deben diseñar de tal modo que sean iguales o converjan a cero ($r_i \approx 0$) cuando no existan fallos presentes, y desviarse significativamente de cero ($|r_i| > T_i > 0$, donde $T_i \in \mathcal{R}$ denota un umbral) bajo la presencia de fallos. Idealmente, un residuo solo debe llevar información relacionada con fallos y al mismo tiempo ser insensible ante incertidumbre del modelo matemático.

Existen varias técnicas para generar el vector de residuos r , se pueden destacar los métodos basados en estimación de parámetros, ecuaciones de paridad y observadores.

6.3.1. Generación de Residuos mediante estimación de parámetros

Este método se basa en la suposición de que los fallos se reflejan en los parámetros físicos del sistema (fricción, masa, viscosidad, resistencia, etc.). Los parámetros del proceso son continuamente estimados en línea y comparados con los parámetros del modelo de referencia, es decir, en ausencia de fallos, por lo que cualquier discrepancia indica la presencia de un fallo (Blá-03).

$$y(t) = f(\theta, u(t)) \quad (6.3)$$

Este método normalmente utiliza el modelo matemático de entrada - salida de un sistema en la forma expresada en la ecuación 6.1, donde θ es el vector de coeficientes del modelo que está directamente relacionado con los parámetros físicos del sistema, P_j , con $j = 1 \dots n$, siendo n el número de parámetros físicos del sistema. La función $f(.,.)$ puede ser lineal o bien no lineal [Sod-89], [Lju-99], [Hab-99], [Tan-02]. El procedimiento básico para llevar a cabo la detección y diagnóstico de fallos mediante el método de estimación de parámetros, de acuerdo a [Blá-03], es el siguiente:

- Establecer el modelo del proceso $y(t) = f(\theta, u(t))$.
- Determinar las relaciones entre los coeficientes del modelo θ_i y los parámetros físicos del proceso P_j .
- Estimar los coeficientes del modelo nominal, θ_i a partir de las medidas de *entrada - salida* del proceso, $u(t)$ e $y(t)$, mediante algún método de identificación.
- Calcular los parámetros físicos del proceso nominal.
- Determinar los cambios de los parámetros que ocurren por los diferentes casos de fallo $P_j = f^{-1}(\theta_i)$.

Para llevar a cabo el último paso, para fallos conocidos, se puede construir una base de datos de fallos y sus síntomas durante la operación del sistema, los coeficientes del modelo del sistema se identifican periódicamente a partir de las entradas y salidas medidas, y se comparan con los parámetros del modelo nominal y con fallos.

Para generar residuos mediante este método, se debería utilizar un algoritmo en línea de identificación paramétrica. Si se llama a la estimación de los coeficientes del modelo en el instante $k-1$ como θ_{k-1} , el residuo se puede definir en alguna de las formas expresadas en las ecuaciones 6.4, donde θ_o son los coeficientes del modelo nominal.

$$r(k) = \theta_k - \theta_o \tag{6.4}$$

$$r(k) = y(k) - f(\theta_{k-1}, u(k))$$

No es fácil conseguir el aislamiento del fallo mediante el método de estimación de parámetros. Esto es por qué los parámetros que se identifican son parámetros del modelo que no siempre pueden convertirse en los parámetros físicos del sistema [Ise-84]. Sin embargo los fallos se representan mediante variaciones en los parámetros físicos. En [Dor-96) se propone el método de la matriz de influencia para superar la dificultad del aislamiento. La idea es identificar la influencia de cada parámetro físico en el residuo.

Básicamente existen dos enfoques para modelar el comportamiento de entrada salida del sistema monitoreado: el método de ecuación de error y el método de error de salida, los cuales se describen a continuación:

Método de ecuación de error.

El modelo de tiempo discreto de un proceso SISO lineal, invariante en el tiempo, de orden n se escribe en forma de regresor:

$$y(t) = \Psi^T \Theta \tag{6.5}$$

Donde:

$$\Theta^T = [a_1 \dots a_n, b_1 \dots b_n] \tag{6.6}$$

Es el vector de parámetros y:

$$\Psi^T = [y(t-1) \dots y(t-n), u(t-1) \dots u(t-n)] \tag{6.7}$$

Es el vector de datos en tiempo discreto, también llamado regresor. El error se define entonces por:

$$e(t) = y(t) - \Psi^T \Theta \tag{6.8}$$

O bien, si

$$\frac{Y(z)}{U(z)} = \frac{B(z)}{A(z)} \quad (6.9)$$

Es la función de transferencia del proceso, la ecuación de error empleando transformada z es:

$$E(z) = \hat{B}(z)U(z) - \hat{A}(z)Y(z) \quad (6.10)$$

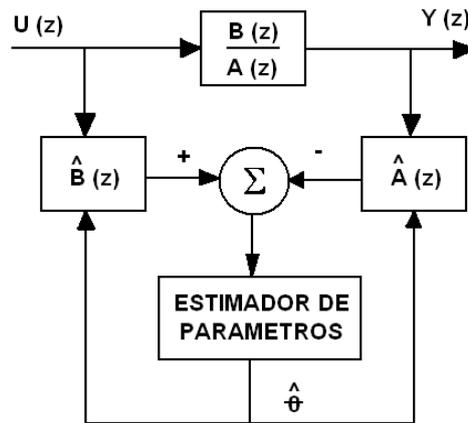


Figura 6. 2. Estimación de parámetros ecuación del error

Método de error de salida

En este caso, a diferencia del anterior, se emplea la salida del modelo como se observa en la figura 6.3. Una desventaja de este método es que no es posible el cálculo directo del parámetro estimado Θ , dado que $e(t)$ no es lineal en los parámetros [Sim-03].

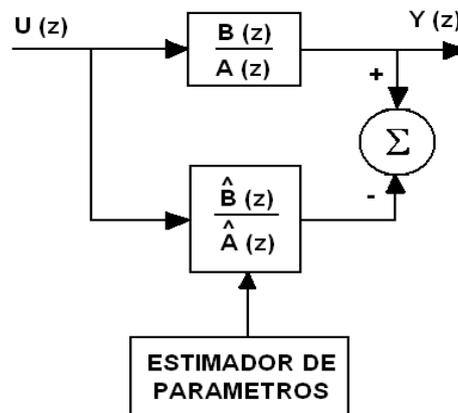


Figura 6. 3. Estimación de Parámetros. Error de salida.

6.3.2. Ecuaciones de paridad

Las ecuaciones de paridad, introducidas por [Will-76], se basan en comprobar la paridad (coherencia) de los modelos de la planta con las salidas de sensores (mediciones) y las variables del proceso. Bajo condiciones ideales, es decir, en el sistema libre de fallos, el residuo o valor de las relaciones de paridad son cero. Sin embargo, en condiciones reales, el ruido, las inexactitudes del modelo, los fallos, etc., producen una salida distinta de cero, [Che-99], por lo que se debe reorganizar la estructura del modelo, a fin de obtener el mejor aislamiento de fallos.

Considerando un sistema en variables de estado, como el expresado en la siguiente ecuación:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t)\end{aligned}\tag{6.11}$$

Asumiendo que $D = 0$, con $y \in R^n$ y $x \in R^m$, si se cumple que $n > m$, hay redundancia analítica directa, es decir, el número de medidas es mayor que el número de variables a considerar, entonces las salidas vendrán dadas por $y(t) = Cx(t)$, con $C \in R^{n \times m}$ cuando no hay fallos, y cuando ocurre un fallo, la representación vendrá dada por $y(t) = Cx(t) + f(t)$, entonces el espacio de paridad puede construirse a partir de la elección adecuada de una matriz $V \in R^{(n-m) \times n}$ que satisfaga:

$$\begin{aligned}VC &= 0 \\ V^T V &= I_n - C(C^T C)^{-1} - C^T\end{aligned}\tag{6.12}$$

Obteniéndose un vector de paridad:

$$p(t) = Vy(t) = VCx(t) + Vf(t) = Vf(t)\tag{6.13}$$

Cuando el sistema está libre de fallos $p(t) = 0$. Para un fallo simple del i -ésimo sensor tendremos: $f = [0 \ 0 \ 0 \ \dots \ f_i \ \dots \ 0]^T$ y $V_f = f_i$, determinando V las direcciones asociadas a las direcciones de los fallos de esos sensores, e indicando que el residuo solo contiene información del fallo y es independiente del estado no medible $x(t)$.

En el caso de redundancia temporal el sistema es representado en el dominio de la frecuencia, donde se encuentran las ecuaciones diferenciales o las ecuaciones en diferencia y se aplica la transformada Z para obtener las relaciones.

Una forma más general de representar las relaciones de paridad tanto para redundancia directa como temporal es la propuesta en [Cho-84], donde se considera el modelo de espacio de estados definido por:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{6.14}$$

Siendo la salida en el instante $t+1$:

$$y(t+1) = CAx(t) + CBu(t) + Du(t+1) \tag{6.15}$$

Para algún $s > 0 \in s = 1, \dots, n_1 \leq n$, $y(t+s)$ tomara la forma:

$$y(t+s) = CA^s x(t) + CA^{s-1} Bu(t) + \dots + CBu(t+s-1) + Du(t+s) \tag{6.16}$$

Que de forma compacta puede expresarse como:

$$Y(t) = Qx(t - n_1) + RU(t) \tag{6.17}$$

Siendo:

$$Y(t) = \begin{pmatrix} y(t - n_1) \\ y(t - n_1 + 1) \\ y(t - n_1 + 2) \\ \vdots \\ y(t) \end{pmatrix} \quad U(t) = \begin{pmatrix} u(t - n_1) \\ u(t - n_1 + 1) \\ u(t - n_1 + 2) \\ \vdots \\ u(t) \end{pmatrix}$$

Q y R son matrices que se construyen como:

$$Q = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n_1} \end{pmatrix} \quad R = \begin{pmatrix} D & 0 & \dots & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ CA^{n_1-1}B & CA^{n_1-2}B & \dots & \dots & D \end{pmatrix}$$

Realizando la multiplicación de esa relación por un vector w^T de dimensiones adecuadas, obtenemos:

$$w^T Y(t) = w^T Qx(t - n_1) + w^T RU(t) \tag{6.18}$$

En general, la relación contiene las variables de entrada, salida y de estado desconocidas, que será una relación de paridad únicamente si desaparecen los estados, por lo que se cumplirá:

$$w^T Q = 0 \quad (6.19)$$

En consecuencia tendremos un conjunto de ecuaciones lineales homogéneas, y si el sistema es observable, las ecuaciones serán independientes, por lo que se obtendrán los mismos resultados para redundancia directa o temporal al fijarse los objetivos de diseño para la obtención de fallos [Ger- 91].

Las ecuaciones de paridad también se pueden construir utilizando un modelo *entrada - salida* de la transformada Z , o una representación de matriz de transferencia discreta, tal como se expresa en las siguientes ecuaciones:

$$y(k) = G_u(z)u(k) \quad (6.20)$$

$$Q(z)y(k) = N(z)u(k) \quad (6.21)$$

Las relaciones entre las matrices de transferencia discretas vienen expresadas como:

$$G_u(z) = C(zI - A)^{-1}B + D \quad (6.22)$$

$$N(z) = C[\text{adjunta}(I - z^{-1}A)]z^{-1}B + D[\det(I - z^{-1}A)] \quad (6.23)$$

$$Q(z) = [\det(I - z^{-1}A)]I \quad (6.24)$$

A partir de las relaciones (6.20) y (6.21), los vectores de residuo son expresados como:

$$r(k) = y(k) - G_u(z)u(k) \quad (6.25)$$

$$r(k) = Q(z)y(k) - N(z)u(k) \quad (6.26)$$

Cuando se realiza la transformada Z de los polinomios se debe conseguir una estructura en las ecuaciones de paridad que cumpla satisfactoriamente con los objetivos de diagnóstico, cuyas condiciones son:

$$J(z)\psi(k) = 0 \quad (6.27)$$

$$\psi(k) = \begin{pmatrix} u(k) \\ y(k) \end{pmatrix} \quad J(k) = \begin{pmatrix} N(k) & -Q(k) \end{pmatrix}$$

En el caso real los residuos no son nulos, por lo que en la práctica, mediante técnicas de identificación, se obtiene $\psi_{med}(k)$ que es una estimación del modelo para los valores medidos, siendo el vector de residuos una combinación de ruido, fallos aditivos, fallos multiplicativos y errores de modelado, que es representado como:

$$r(k) = \hat{J}(z)\psi_{med}(k) \quad (6.28)$$

En forma general, para resolver el problema definido por el modelo (6.27), se define la matriz de incidencia Π , que es una matriz booleana:

$$\Pi = Inc[J(z)] \quad (6.29)$$

Cuyos elementos toman el valor 0 ó 1 según la regla:

$$\Pi_{ij} = \begin{cases} 1 & \text{si } j_{ij} \neq 0 \\ 0 & \text{si } j_{ij} \cong 0 \end{cases} \quad (6.30)$$

De acuerdo a [Blá-03], si $J(z)$ es la matriz del modelo *entrada - salida* definido por la ecuación (6.26), entonces $\Pi_{ij} = 1$ significa que una desviación en la medida de la variable j , originada por ruido o por un fallo, influye en el residuo i , mientras que $\Pi_{ij} = 0$ significa lo contrario.

Si consideramos que cada fallo está asociado directamente a una variable j , lo cual es evidente para los fallos en sensores y actuadores, el elemento Π_{ij} indica si el fallo j será detectado por el residuo r_i .

El distinguir si dicha desviación es producto del ruido o realmente de un fallo es labor del módulo de decisión. Se suele decir que un residuo es significativo cuando el citado módulo decide que es realmente distinto de cero. Utilizando la matriz de incidencia Π se pueden definir los siguientes conceptos:

- El sistema definido por la matriz $J(z)$ tiene una estructura sensible al conjunto de fallos asociados a las variables $\Psi(k)$ si ninguna de las columnas de Π tiene todos sus elementos nulos.
- Dicho sistema tiene una estructura débilmente aislables si todas las columnas de Π son distintas.
- La estructura será fuertemente aislable si todas las columnas tienen al menos dos elementos distintos entre sí.

El conjunto de ecuaciones escogidas para describir el sistema reflejará en su estructura las características de aislabilidad y robustez, que si son aceptables permitirán el diagnóstico adecuado del sistema.

6.3.3. Observadores

Los observadores son elementos que permiten estimar las variables de un sistema en base a las mediciones de las señales de salida y entrada del sistema en estudio. Estos tienen la ventaja de ser robustos a las perturbaciones, variaciones de los parámetros y el ruido del sistema. Cuando todas las variables de estado de un sistema están disponibles podemos construir un observador de orden completo y cuando solo disponemos de algunas construimos un observador de orden reducido.

Las aplicaciones de los observadores son muy variadas, y en el caso en que son utilizados para los esquemas de detección y diagnóstico de fallos, se diseña el error de estimación del observador para generar el residuo o un conjunto de ellos, que detecten e identifiquen de manera exclusiva los diferentes fallos.

Cuando utilizamos un determinado observador debemos considerar los tipos de fallos que se van a detectar (componentes, actuadores o sensores), los criterios de diseño (robustez, aislabilidad, etc.) y la estructura del sistema (observabilidad, controlabilidad, linealidad, etc.). En cuanto al aislamiento de fallos, se pueden utilizar observadores para construir residuos estructurados, que se conoce como esquema de observadores dedicado y esquema de observador generalizado [Fran-87], [Fra-90].

La dinámica que se propone para un observador lineal, aplicada para reconstruir los estados de un sistema lineal está dada por la siguiente relación:

$$\begin{aligned}\dot{\hat{x}} &= A\hat{x} + L(y - \hat{y}) + B_c U_c \\ \hat{y} &= C\hat{x}\end{aligned}\tag{6.31}$$

Donde \hat{x} son los estados estimados, \hat{y} es la salida estimada, U_c es una señal de control, L y B_c son vectores de ganancia a ser diseñados de tal manera que sea posible obtener un estimador asintótico de los estados del sistema lineal. Para ello, es necesario definir el error de observación, el cual viene dado por la diferencia entre el valor real del vector de estado del sistema linealizado y el valor estimado de tal vector de estado. Esta definición, por supuesto, se extiende a los valores iniciales de tales estados, aunque el valor inicial del error, en un cierto t_o , es desconocido, y probablemente será distinto de cero [Sir-05].

$$e_x = x - \hat{x}, \quad e_x(t_o) = x(t_o) - \hat{x}(t_o) \quad (6.32)$$

Siendo e_x el error de observación y su dinámica, considerando un sistema en variables de estado como el escrito en la ecuación 6.11 y suponiendo $D = 0$, entonces:

$$\dot{e}_x = \dot{x} - \dot{\hat{x}} = Ax + Bu - A\hat{x} - L(y - C\hat{x}) - B_c u_c \quad (6.33)$$

El elegir B_c y u_c apropiadamente permite que el error de observación evolucione según la dinámica lineal y autónoma dada por:

$$\dot{e}_x = (A - LC)e_x \quad (6.34)$$

Entonces para el diseño del vector L se debe cumplir que el error de estimación sea asintóticamente estable a cero, siendo condición necesaria y suficiente que los autovalores de la matriz $(A - LC)$ sean negativos, ubicando los polos del sistema lineal (representado por el error de observación) en el semiplano izquierdo del plano complejo.

En la práctica muchos sistemas no lineales no se pueden representar mediante modelos lineales, en particular cuando dichos sistemas no operan en un punto fijo de operación nominal. Algunas técnicas que han sido desarrolladas con observadores no lineales permiten establecer mejores estrategias para generar residuos más robustos, donde la representación matemática general es expresada como:

$$\begin{aligned} \dot{x}(t) &= g(x(t), u(t), f(t), d(t)) \\ y(t) &= h(x(t), u(t), f(t), d(t)) \end{aligned} \quad (6.35)$$

Siendo $x(t)$ el vector de estado, $y(t)$ el vector de salida, $u(t)$ el vector de entrada, $f(t)$ el vector del fallo, $d(t)$ el vector de perturbaciones y $g(., ., ., .)$ y $h(., ., ., .)$ representan las funciones no lineales.

El problema del diagnóstico de fallos es generar un vector de residuos $r(t)$ a partir de la dinámica del error de observación ($\dot{\xi}(t)$) utilizando la estructura del observador:

$$\begin{aligned} \dot{\xi}(t) &= g_r(\xi(t), u(t), y(t)) \\ r(t) &= h_r(\xi(t), u(t), y(t)) \end{aligned} \quad (6.36)$$

El residuo debe satisfacer la siguiente condición:

$$\|r(t)\| \begin{cases} \approx 0, & \text{cuando } f(t) = 0 \\ \geq 0, & \text{cuando } f(t) \neq 0 \end{cases} \quad (6.37)$$

Estas estructuras generales han definido diversos observadores no lineales, que pueden ser estudiadas con mayor profundidad en [Che-99].

6.4. Evaluación de residuos

Una vez que los residuos son generados, la siguiente etapa es analizarlos, determinar si contienen información acerca de posibles fallos, y proporcionar información adecuada en caso de que exista alguna. A esta etapa se le llama evaluación de residuos o toma de decisiones. El proceso de evaluación puede consistir en una simple prueba de umbral sobre el valor instantáneo o la medida variable de los residuos, o puede consistir en métodos estadísticos. La etapa de evaluación de residuos se vuelve relativamente sencilla si estos son bien diseñados.

La evaluación requiere determinar si los residuos sobrepasan algún valor de umbral determinado. El valor de umbral es necesario para evitar falsas alarmas debidas a condiciones iniciales, pequeñas perturbaciones o dinámicas no modeladas.

Generalmente la evaluación se realiza obteniendo una medida del residuo. Esta medida puede ser estadística o determinística. La primera está basada en cálculo de la desviación estándar, varianza, media, etc. Mientras que la segunda puede ser basada en la noción de distancias definidas en los espacios vectoriales, tales como normas, siendo las más utilizadas la norma 2 o euclidiana y la norma infinito.

La monitorización con estadísticas univariantes se usan para determinar los umbrales de cada variable observada de forma independiente. Los umbrales definen el límite para las operaciones en control y el traspaso de estos umbrales por los datos en línea indican la presencia de un fallo. Los valores límites de control superior e inferior son críticos para minimizar las falsas alarmas y el número de alarmas no detectadas [Fue-06]. Para detectar los fallos se pueden utilizar las gráficas de Shewhart. Cuando se monitoriza una variable continua de un proceso en condiciones libres de fallo, dicha variable tendrá una distribución normal alrededor de la media μ . Y la dispersión de los valores puede medirse en términos de la desviación estándar σ .

En procesos bajo control, normalmente el 99.7% de las medidas están entre los límites $(\mu \pm 3\sigma)$, este hecho impone los umbrales para cada variable. En consecuencia la representación se hace con una gráfica de control, que consiste, para el caso univariado, en tomar medidas de una variable, calcular su

media y representarla en una gráfica donde está la media de esa variable, así como sus límites inferior y superior de control, $\mu \pm 3\sigma$, figura 6.4.

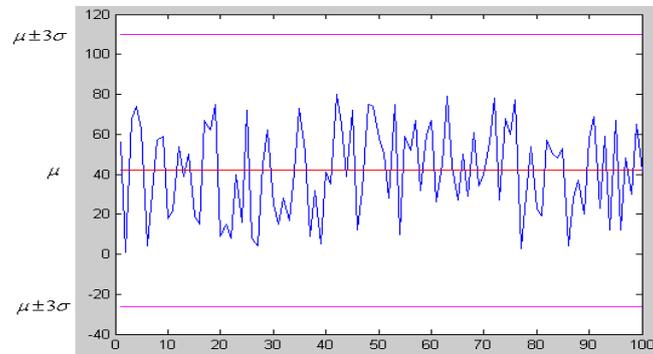


Figura 6. 4. Grafica de control

El gráfico indica que si la media de la variable esta fuera de los límites establecidos, el sistema está en presencia de un fallo. Este sistema tiene un inconveniente, y es que al trabajar con un proceso real, no se toman medidas individuales de cada variable y se representan en la distribución normal del proceso, ya que esto aumentaría la variabilidad del mismo, sino que se toman muestras de n datos cada una y se calcula su media y este valor es el que se representa en la gráfica de control. En este caso la desviación estándar de las muestras se llama error estándar, SE , y se calcula como: $SE = \sigma / \sqrt{n}$. Por tanto ahora, los límites superior e inferior de control son $\mu \pm 3SE$, es decir, $\mu \pm 3\sigma / \sqrt{n}$ [Fue-06].

Además, si se utilizan muestras de datos de tamaño n , en vez de datos individuales, el teorema del límite central, dice que aunque la distribución de las medidas de la variable no sea normal, cuanto mayor sea n , la distribución de las medias de las muestras se aproxima más a una distribución normal con media μ y error estándar σ / \sqrt{n} , y por tanto los límites utilizados son válidos.

Este método no es adecuado para detectar cambios pequeños, se puede mejorar si se usan datos que incorporen medidas tomadas en múltiples consecutivos instantes de tiempo, como el método de la Suma Acumulativa o CUSUM, y la media móvil pesada exponencialmente (EWMA). En [Leg-98] se examina la viabilidad de utilizar la suma acumulativa (CUSUM) gráficas de control y las redes neuronales artificiales de base radial (RBF) en conjunto para la detección y diagnóstico de fallos.

En [Fue-06] se señala que su aplicación es limitada, ya que el método no considera las posibles correlaciones de las variables del proceso, por lo que no se podrán detectar diferentes tipos de fallos presentes en los procesos industriales y serán poco robustos.

En el análisis de datos multivariantes existen técnicas para reducción de la dimensionalidad, basadas en que es posible describir con precisión m variables con un subconjunto $r < m$ de ellas, reduciendo la dimensión del problema a costa de una pequeña pérdida de información.

Las variables medidas en un proceso están correlacionadas tanto con otras variables del proceso como con otras medidas de la misma variable tomadas en tiempos distintos. Dichas correlaciones son más significativas cuando los intervalos de muestreo son pequeños. Estas características han permitido el desarrollo de técnicas de monitorización para detección y diagnóstico de fallos basadas en técnicas estadísticas multivariantes.

6.5. Localización de fallos mediante residuos estructurados

Para lograr la tarea de localización de fallos el diseño de residuos estructurados resulta un método útil. Cada residuo se debe diseñar para ser sensitivo a un fallo o un subconjunto de fallos, mientras aparezca insensible a los fallos restantes. El conjunto de residuos que tiene la sensibilidad requerida a fallos específicos e insensibilidad al resto de los fallos es conocido como *conjunto de residuos estructurados*. La ventaja de un vector de residuos estructurados es que la evaluación se simplifica a determinar cuál de los residuos se desvía de cero.

Si todos los fallos posibles pueden ser localizadas se puede diseñar un conjunto de residuos de acuerdo a las siguientes condiciones de sensibilidad de fallos.

$$r_i(t) = R(f_i(t)); i \in \{1, 2, \dots, g\} \quad (6.38)$$

Donde $R(\cdot)$ denota una relación funcional, f_i denota el i -ésimo fallo y g es el número de fallos a detectar y ubicar. Los residuos resultantes son llamados *conjunto dedicado de residuos*. Se puede aplicar una sencilla lógica de umbral para determinar la aparición de un fallo específico.

$$r_i > T_i \Rightarrow f_i(t) \neq 0; i \in \{1, 2, \dots, g\} \quad (6.39)$$

Donde T_i ($i = 1, 2, \dots, g$) son umbrales. Esta estructura es muy simple y se puede detectar todas las fallas simultáneamente, sin embargo es muy difícil de diseñar en la práctica. Aun cuando este conjunto de residuos puede ser diseñado, normalmente no queda libertad de diseño para alcanzar otras características deseables como robustez ante errores de modelado. Un esquema mejor y más usado consiste en hacer cada residuo sensitivo a todos menos a un fallo, esto es:

$$r_1(t) = R(f_2(t), \dots, f_g(t)) \quad (6.40)$$

$$r_i(t) = R(f_1(t), \dots, f_{i-1}(t), f_{i+1}(t), \dots, f_g(t)) \quad (6.41)$$

$$r_g(t) = R(f_1(t), \dots, f_{g-1}(t)) \quad (6.42)$$

El cual es definido como conjunto generalizado de residuos. La localización se puede realizar de nueva cuenta usando sencillas pruebas de umbral de acuerdo a la siguiente lógica.

En la figura 6.5 se ilustran ambos conjuntos de residuos para el caso de tres fallos distintos. Los fallos se pueden localizar usando cualquiera de los dos métodos.

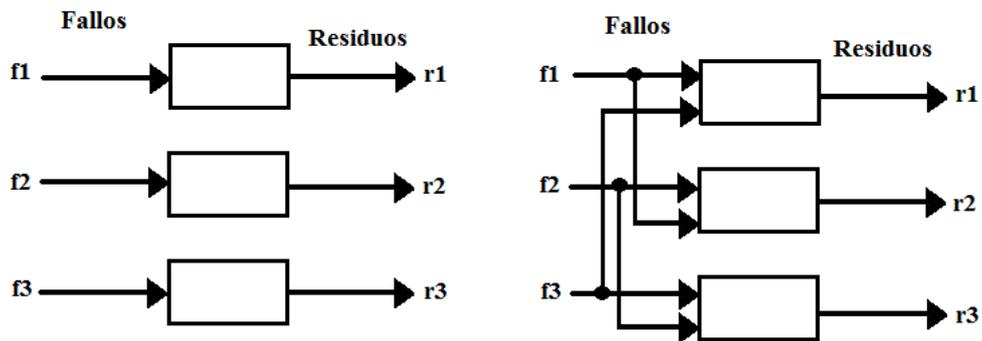


Figura 6. 5. Conjunto de residuos estructurados a) Dedicado b) Generalizado

6.6. Metodología de Diagnóstico propuesta

La metodología de diagnóstico de fallos que se utiliza en esta tesis se fundamenta en la teoría de diagnóstico basado en modelos que se describe en [Ger-98]. Dicha metodología parte de la utilización de un conjunto de indicadores de la existencia de fallos, denominados residuos $r(k)$. Dichos residuos se obtienen a partir de las medidas de las entradas $u(k)$ y salidas $y(k)$ del sistema monitorizado y las relaciones analíticas existentes entre ellas obtenidas a partir del modelo del sistema:

$$r(k) = \psi(y(k), u(k)) \quad (6.43)$$

Siendo Ψ la función generadora de residuos que depende del tipo de estrategia de detección utilizada (observador o ecuación de paridad). La etapa de detección en tiempo real consiste en evaluar en cada instante de tiempo la expresión de los residuos, utilizando las medidas de las entradas y salidas. Idealmente, dichos residuos deberían ser cero en caso de que no exista fallo, mientras que en caso de fallo serían diferentes de cero. En la práctica, debido a la existencia de ruido e incertidumbre (imprecisión) en el modelo se utiliza un umbral (determinado por ejemplo mediante técnicas estadísticas) a partir del cual se decide la existencia o no del fallo. El resultado de la evaluación de cada residuo $r_i(k)$ frente a su test de detección permite obtener un conjunto de firmas observadas de fallo: $\phi(k) = [\phi_1(k), \phi_2(k), \dots, \phi_n(k)]$, donde cada indicador de fallo es:

$$\phi_i(k) = \begin{cases} 0, & \text{si } |r_i(k)| \leq \tau_i \\ 1, & \text{si } |r_i(k)| > \tau_i \end{cases} \quad (6.44)$$

La firma observada del fallo la proporciona el módulo de aislamiento del fallo. Este módulo es capaz de aislar el fallo puesto que dispone de la relación existente entre el conjunto de fallos considerado $f(k) = \{f_1(k), f_2(k), \dots, f_n(k)\}$ y el conjunto posibles firmas de fallo observadas $\phi_i(k)$. Dichas relaciones se almacenan en una matriz denominada matriz teórica de firmas de fallos (FSM). Cada elemento FSM_{ij} de dicha matriz es igual a 1 en caso de que el fallo $f_j(k)$ afecte al residuo $r_i(k)$ de tal manera que el indicador de fallos $\phi_i(k)$ será igual a 1 cuando dicho fallo aparezca en el sistema monitorizado. En caso contrario, el elemento FSM_{ij} valdrá cero.

6.6.1. Análisis de sensibilidad

El problema que presenta la metodología de diagnóstico clásica, presentada en la sección anterior, es que pueden existir varios fallos que presenten la misma firma binaria teórica de fallos. En ese caso, para poder separar dichos fallos se debe realizar un análisis más exhaustivo, utilizando otras

informaciones complementarias asociadas a la relación entre los residuos y los fallos como el signo, la sensibilidad, el orden o el tiempo activación de los indicadores de fallo [Pui-06]. En esta tesis, se propone la utilización de la información proporcionada por la sensibilidad del residuo frente al fallo para el diseño del diagnosticador. La sensibilidad, tal como la define Gertler se presenta en la ecuación siguiente:

$$S_f = \frac{\partial r}{\partial f} \quad (6.45)$$

La ecuación 6.45 es una función de transferencia que describe el efecto sobre el residual, r , de un fallo f dado. Esta sensibilidad proporciona una información cuantitativa del efecto del fallo sobre el residuo y una información cualitativa sobre su sentido de variación (signo). La utilización de esta información en la etapa de diagnóstico permitirá separar fallos que aun presentando la misma firma de fallos teórica binaria, presentan sensibilidades cualitativamente o cuantitativamente diferentes. Para ello, el algoritmo de diagnóstico propuesto utilizará una *matriz teórica de firmas sensibilidad FSMsensit* de los residuos frente a los fallos en lugar de una tabla de firmas de fallo binaria, de forma que en tiempo real se evaluará la sensibilidad observada frente a la teórica de cara a aislar el fallo. La sensibilidad observada se puede calcular teniendo en cuenta que el valor del residuo $r_i(k)$ cuando aparezca el fallo $f_j(k)$ vendrá dado por el valor de la sensibilidad $S_{r_i f_j}$:

$$r_i(k) = S_{r_i f_j} f_j(k) \quad (6.46)$$

La matriz de sensibilidad S_f establece la capacidad residual para detectar fallos, su importancia en todo el proceso de diagnóstico de fallos es crucial. La matriz FSMsensit se define como una matriz cuyos elementos FSMsensit_{ij} establecen la facilidad con la que un fallo f_j hará que el r -ésimo residual $r_i(k)$ pueda superar su umbral adaptativo asociado. La siguiente ecuación describe como calcular las FSMsensit_{ij}.

$$FSMsensit_{ij} = \begin{cases} \frac{S_{f_{ij}}(q^{-1})\eta(k-t_0)}{|\bar{r}_i^o(k)|} & \text{if } r_i^o(k) \geq 0, k \geq t_0 \\ \frac{S_{f_{ij}}(q^{-1})\eta(k-t_0)}{|\bar{r}_i^o(k)|} & \text{if } r_i^o(k) < 0, k \geq t_0 \\ 0 & \text{if } k < t_0, \text{ or } S_{f_{ij}}(q^{-1}) = 0 \end{cases}$$

Donde $\eta(k)$ es un escalón unitario, $S_{f_{ij}}(q^{-1})$ es la sensibilidad asociada con el residual nominal $r_i^o(k)$ con respecto a la hipótesis de fallo $f_j(f_{yj}, f_{uj}, f_{aj})$ y t_0 es el instante de tiempo de ocurrencia del fallo.

Cuando t_0 es desconocido, debe ser estimado usando el instante de tiempo de ocurrencia k_{ϕ_i} de la primera señal de fallo observada $\phi_i(k)$. Como consecuencia de la dependencia del tiempo de la sensibilidad del residuo a fallos, $FSMsensit_{ij}$ evoluciona dinámicamente desde la ocurrencia del fallo en el instante de tiempo t_0 . Esta función se corrige utilizando el umbral de detección más grande posible, y así se podrá observar las señales de fallo más difíciles.

6.6.2. Modelización del efecto de los fallos.

Mientras que los fallos de sensores y actuadores pueden ser manejadas por una simple extensión del modelo de sistema sin fallos, los fallos de proceso requieren modelado explícito. Esto se puede hacer sobre el sistema físico, por emulación de fallos, o en un simulador. En cualquier caso, un fallo f_j se introduce (con ubicación y tamaño conocido) y el vector resultante $x(t | f_j)$ es observado (este vector es normalizado usando la media y la desviación estándar de los datos de entrenamiento libre de fallos). Este puede ser utilizado como la ganancia (sensibilidad) de la respuesta, por la ampliación de $x(t | f_j)$ a un fallo de tamaño unidad.

A continuación se hará uso del vector de sensibilidad transformado $S_{\bullet j} = [S_{1j} \dots S_{(n-m+k)j}]'$, $j = 1 \dots k$, donde:

$$S_{ij} = q_i' x(t | f_j) / f_j \quad i = 1 \dots n - m + k \quad (6.47)$$

Hay que tener en cuenta que es necesario simular cada fallo con un número de diferentes tamaños, para comprobar la linealidad de la respuesta o evaluar su desviación de la linealidad, debido a la no linealidad del sistema.

6.6.3. Monitorización

En la fase de monitorización, medidas $x(t)$ son tomadas del sistema físico o simulado y normalizado con el promedio y desviación estándar del set de entrenamiento libre de fallos. Estos datos normalizados son usados para obtener los residuos para el diagnóstico de fallos.

Los residuos pueden evaluarse desde un punto de vista direccional o estructurado. El punto de vista direccional es complejo y no es discutido en esta tesis. En un análisis estructurado, el hecho de que la dimensión del vector residual $p(t)$ es igual al número de fallos, permite el diseño de una estructura diagonal.

$$r(t) = S^{-1}p(t) \quad (6.48)$$

Obedeciendo a una estructura diagonal: cada residuo responde a un fallo en específico. Esta estructura permite el aislamiento de múltiples fallos simultáneos.

6.7. Aplicación a fallos múltiples de una turbina de gas

Para ilustrar la propuesta metodología de diagnóstico de fallos múltiples basados en residuos estructurados, se ha utilizado el modelo presentado en el capítulo 3 para la turbina de gas SR-30 desarrollado en Simulink y donde se pueden emular fallos en componentes de diferente magnitud y comportamiento. Las ecuaciones matemáticas, tipos de fallos en la turbina y el modelo en Simulink también fueron explicados en el capítulo 3 [Rin-12][Rin-13].

6.7.1. Descripción de los fallos a diagnosticar.

Los fallos considerados en este ejemplo, se describen en la Tabla 6.1.

Tabla 6. 1. Descripción de los fallos en componentes de la turbina

Fallo	Descripción
Fallo (1) PRC	Obstrucción o desgaste en el compresor
Fallo (2) EC	Rotura o desgaste mecánico en los alabes del compresor
Fallo (3) WC	Fuga de masa de aire en compresor
Fallo (4) PRT	Obstrucción o desgaste en la turbina
Fallo (5) ET	Rotura o desgaste mecánico en los alabes de la turbina
Fallo (6) WT	Fuga en la cámara de combustión
Fallo (7) WF	Obstrucción en la válvula de combustible

La simulación de cada escenario de fallo descrito en la tabla 6.1 se logra mediante la adaptación del modelo desarrollado en el capítulo 3. Las variables de salida o variables medidas del sistema y,

consecuentemente, generadoras de residuos tales como: masa de aire de compresor, masa de aire combustible en la cámara de combustión, eficiencia del compresor, eficiencia de la turbina y flujo de combustible.

La tabla 6.2 muestra las variables medidas en el sistema y con las cuales podemos obtener la matriz de residuos estructurados.

Tabla 6. 2. Variables medidas en la turbina

SENSOR	VARIABLE MEDIDA
s1	Temperatura Estática de entrada del Compresor (T01)
s2	Temperatura de Estancamiento de salida de la etapa de compresor (T02)
s3	Temperatura del Estancamiento de entrada de la etapa de la turbina (T03)
s4	Temperatura de Estancamiento de salida de la etapa de la turbina (T04)
s5	Temperatura de Estancamiento de salida de la tobera (T05)
s6	Presión estática de entrada del compresor (P01)
s7	Presión de estancamiento de salida de la etapa de compresor (P02)
s8	Presión de la cámara de combustión (P03)
s9	Presión de estancamiento de salida de la turbina (P04)
s10	Presión de estancamiento de la salida de la tobera (P05)
s11	Revoluciones del eje de la turbina. (RPM)
S12	Torque generado por la turbina (Tr)

6.7.2. Matriz de sensibilidad

Es claro notar que debido a la no linealidad del sistema, la validez de la matriz de sensibilidad S está restringida a valores cercanos al punto de operación en la que se calcula. La matriz de sensibilidad se obtuvo para fallos de 1% para que la matriz de sensibilidad obtenida fuese unitaria y de esta manera se refleje el porcentaje del fallo de los componentes sin tener que hacer más cálculos matemáticos de conversión y calibración. Los residuos estructurados se consideran anormales por encima de +/- 2% de la variable medida. Se realizó un estudio de ruido en las medidas de los sensores y un análisis de incertidumbre del modelo y así delimitar el umbral de detección.

Los residuos se generan a partir de las variables observables en la turbina de gas, es decir, de los valores nominales de las entradas. Idealmente, los residuos sólo deben verse afectados por los fallos. Sin embargo, la presencia de errores de perturbaciones, de ruido y de modelado también hace que los residuos sean distintos de cero y por lo tanto interfieren con la detección de fallos. Por lo tanto el generador de residuos debe ser diseñado de modo que sea máximamente afectado por estas entradas molestas, lo que significa que sea robusto frente a errores de perturbación, ruido y modelado. Los residuos estructurados están diseñados de modo que cada residual responda a un subconjunto diferente de fallos e insensible a los otros. Cuando se produce un fallo en particular, algunos de los residuos deben reaccionar y otros no. Entonces, el patrón del conjunto de respuestas, firma fallos o código de fallos, es característica de un fallo en particular.

Al examinar el siguiente ejemplo de una firma de fallos:

	<i>R1</i>	<i>R2</i>	<i>R3</i>
<i>L1</i>	1	0	0
<i>L2</i>	0	1	0
<i>L3</i>	0	0	1

La firma de fallos anterior implica que *L1* sólo afecta al residual *R1* así como *L2* afecta *R2* y *R3* afecta a *L3*. Con el fin de mejorar la detección y el aislamiento del conjunto de fallos, los residuos estructurados se pueden utilizar. La firma de fallos describe el subconjunto de los residuos que reaccionan a cada fallo. Dado que las diferentes variables del sistema se supone que son medibles, habrá siete residuos correspondientes a cada uno de los sensores. Cuando hay uno o varios fallos, todos los residuos se vean afectados. Por una simple revisión de los residuos es posible predecir el cambio en el comportamiento del sistema de normal. Pero no es posible identificar la ubicación de los fallos. Así que el residual tiene que ser transformado para mejorar el aislamiento.

Dejando que la salida de la planta esté dada por:

$$Y_p = G(s)U(s) + G_F(s)L(s) \tag{6.49}$$

Donde: $G(s)$ es la función de transferencia en condiciones normales.

$G_F(s)$ es la función de transferencia con fallo.

$$Y_M(s) = G(s)U(s) + G_F(s)L(s) \tag{6.50}$$

Donde:

$$U(s) = \begin{bmatrix} \text{fin}_1(s) \\ 0 \\ \text{fin}_3(s) \end{bmatrix} \quad L(s) = \begin{bmatrix} L_1(s) \\ L_2(s) \\ L_3(s) \end{bmatrix}$$

El residuo $R(s)$ está definido como la diferencia entre la salida del proceso y la salida del modelo.

$$R(s) = Y_p(s) - Y_M(s) \quad (6.51)$$

Sustituyendo las expresiones para $Y_p(s)$ y $Y_M(s)$ en la ecuación (6.51)

$$R(s) = G_F(s)L(s) \quad (6.52)$$

$$\begin{bmatrix} R_1(s) \\ R_2(s) \\ R_3(s) \end{bmatrix} = \begin{bmatrix} G_{F11} & G_{F12} & G_{F13} \\ G_{F21} & G_{F22} & G_{F23} \\ G_{F31} & G_{F32} & G_{F33} \end{bmatrix} \begin{bmatrix} L_1(s) \\ L_2(s) \\ L_3(s) \end{bmatrix} \quad (6.53)$$

$$R_1(s) = G_{F11}(s)L_1(s) + G_{F12}(s)L_2(s) + G_{F13}(s)L_3(s) \quad (6.54)$$

$$R_2(s) = G_{F21}(s)L_1(s) + G_{F22}(s)L_2(s) + G_{F23}(s)L_3(s) \quad (6.55)$$

$$R_3(s) = G_{F31}(s)L_1(s) + G_{F32}(s)L_2(s) + G_{F33}(s)L_3(s) \quad (6.56)$$

Para transformar residual crudo $R(s)$ en forma estructurada $R_t(s)$, se multiplica $R(s)$ con la matriz de ponderación $W(s)$.

$$R_t(s) = W(s)R(s) \quad (6.57)$$

La matriz de ponderación se selecciona de acuerdo con:

$$W(s) = Z(s)G_F^{-1}(s) \quad (6.58)$$

Sustituyendo $W(s)$ en la ecuación (6.57)

$$R_t(s) = [Z(s)G_F^{-1}(s)][R(s)]$$

$$R_t(s) = [Z(s)G_F^{-1}(s)][G_F L(s)]$$

$$R_t(s) = Z(s)IL(s) \quad (6.59)$$

Donde I es la matriz de identidad.

$$R_t(s) = Z(s)L(s) \quad (6.60)$$

Si $Z(s)$ es la matriz diagonal,

$$R_{1t}(s) = Z_{11}(s)L_1(s) \quad (6.61)$$

$$R_{2t}(s) = Z_{22}(s)L_2(s) \quad (6.62)$$

$$R_{3t}(s) = Z_{33}(s)L_3(s) \quad (6.63)$$

La ecuación anterior es válida sólo en la ausencia de un desajuste del modelo de la planta y en la ausencia de ruido de la medición. De las ecuaciones anteriores, es evidente que la presencia de fallo afectará a todos residuos. Es difícil de identificar la ubicación del fallo mediante el control de los residuos. Por lo tanto para mejorar el aislamiento de fallos se requiere transformar los residuos.

La matriz de transformación fue construida calculando los residuos basados en los 10 sensores que trae instalado de fábrica la turbina real SR-30 y que se muestra en la tabla 6.2. Para poder obtener una matriz diagonal y cuadrada fue necesario descartar 3 de los 10 sensores. Al reducir el número de sensores se obtuvieron las siguientes matrices de residuos para 55000, 60000 y 65000 rpm respectivamente, tal como se muestran en las tablas 6.3 a la 6.5.

Tabla 6. 3. Matriz de Residuos para 55000 rpm

	Fallo PRC	Fallo EC	Fallo WC	Fallo PRT	Fallo ET	Fallo WT	Fallo WF
(T02)	1.5333	-0.4529	0.8543	-3.2106	1.5730	-0.5533	1.9080
(T03)	3.8415	-2.5997	-2.3879	4.2282	-1.9787	0.7088	0.4291
(T05)	3.8470	-2.6976	-2.6817	5.1458	-2.4160	0.8644	0.0076
(P03)	-0.3787	0.8108	1.9285	-1.9763	0.9233	-0.3316	1.7980
(P04)	-0.3595	0.7961	1.9108	-0.3595	0.9066	0.3256	1.7947
(RPM)	0.1117	0.07034	0.3561	-0.1762	0.0822	0.6319	0.2831
(FG)	-2.58171	1.5181	0.8896	-3.2762	1.5300	-0.5490	1.8764

Tabla 6. 4. Matriz de Residuos para 60000 rpm

	Fallo PRC	Fallo EC	Fallo WC	Fallo PRT	Fallo ET	Fallo WT	Fallo WF
(T02)	0.1528	-0.0843	0.0106	-0.0484	0.0285	-0.0094	0.0254
(T03)	0.2950	-0.2355	-0.2098	0.3539	-0.2074	0.0684	0.0219
(T05)	0.3764	-0.3187	-0.3231	0.5654	-0.3311	0.1092	0.0164
(P03)	0.0811	-0.0139	0.0738	0.0131	-0.0074	0.0025	0.0253
(P04)	0.0817	-0.0146	0.0728	0.0817	-0.0082	0.0028	0.0253
(RPM)	0.0731	-0.0036	0.0075	0.0075	-0.0070	0.1080	0.0114
(FG)	-0.0329	0.0257	-0.0492	-0.0492	0.0286	-0.0095	0.0254

Tabla 6. 5. Matriz de Residuos para 65000 rpm

	Fallo PRC	Fallo EC	Fallo WC	Fallo PRT	Fallo ET	Fallo WT	Fallo WF
(T02)	0.1079	-0.0692	0.0080	-0.0418	0.0318	-0.0104	0.0256
(T03)	0.1766	-0.1672	-0.1227	0.1950	-0.1469	0.0483	0.0162
(T05)	0.1752	-0.1771	-0.1488	0.2494	-0.1878	0.0618	0.0082
(P03)	0.0442	-0.0003	0.0557	-0.0054	0.0052	-0.0016	0.0255
(P04)	0.0447	-0.0009	0.0552	0.0447	0.0045	-0.0014	0.0255
(RPM)	0.0224	0.0079	0.0083	-0.0103	0.0066	0.0454	0.0127
(FG)	-0.0333	0.0297	0.0065	-0.0436	0.0324	-0.0107	0.0255

A partir de las matrices de residuos se obtuvieron las diferentes matrices de sensibilidad según se muestran en las tablas 6.6 a la 6.8.

Tabla 6. 6. Matriz de Sensibilidad para 55000 rpm

	Fallo PRC	Fallo EC	Fallo WC	Fallo PRT	Fallo ET	Fallo WT	Fallo WF
(T02)	-0.7115e3	4.1263e3	-3.9726e3	-0.0313e3	0.0250e3	0.0044e3	-0.1966e3
(T03)	-1.4947e3	5.6631e3	-6.3426e3	-0.0658e3	0.0525e3	0.0092e3	-0.4169e3
(T05)	0.4881e3	-2.5318e3	2.7250e3	0.0226e3	-0.0171e3	-0.0030e3	0.1358e3
(P03)	0.0001e3	-0.0011e3	0.0010e3	-0.0006e3	0.0006e3	0.0000e3	0.0000e3
(P04)	0.0001e3	0.0023e3	-0.0026e3	-0.0020e3	0.0013e3	0.0005e3	-0.0001e3
(RPM)	0.0155e3	-0.0899e3	0.0666e3	0.0003e3	-0.0005e3	0.0014e3	0.0040e3
(FG)	0.0036e3	-0.0215e3	0.0209e3	0.0004e3	-0.0001e3	-0.0000e3	0.0013e3

Tabla 6. 7. Matriz de Sensibilidad para 60000 rpm

	Fallo PRC	Fallo EC	Fallo WC	Fallo PRT	Fallo ET	Fallo WT	Fallo WF
(T02)	-1.6310	5.1171	-3.3967	-0.0296	0.0220	0.0013	-0.5885
(T03)	-2.7595	8.6293	-5.7280	-0.0500	0.0373	0.0022	-0.9834
(T05)	1.0773	-3.3817	2.2436	0.0300	-0.0147	-0.0009	0.3799
(P03)	0.0001	-0.0004	0.0002	-0.0149	0.0149	-0.0000	0.0000
(P04)	0.0433	-0.01066	0.0687	-0.0357	0.0253	0.0031	0.0134
(RPM)	0.8119	-2.5458	1.6896	0.0113	-0.0103	0.0088	0.2911
(FG)	0.4948	-1.5554	1.0350	0.0157	-0.0066	-0.0004	0.2109

Tabla 6. 8. Matriz de Sensibilidad para 65000 rpm

	Fallo PRC	Fallo EC	Fallo WC	Fallo PRT	Fallo ET	Fallo WT	Fallo WF
(T02)	-0.2491	0.8640	-0.7290	-0.0038	0.0030	0.0004	-0.0643
(T03)	-0.3533	1.2220	-1.0310	-0.0054	0.0042	0.0006	-0.0900
(T05)	0.1625	-0.5639	0.4756	0.0040	-0.0020	-0.0003	0.0407
(P03)	0.0001	-0.0004	0.0003	-0.0020	0.0020	-0.0000	0.0000
(P04)	0.0185	-0.0607	0.0508	-0.0037	0.0025	0.0007	0.0045
(RPM)	0.1303	-0.4519	0.3812	0.0015	-0.0015	0.0019	0.0329
(FG)	0.0771	-0.2679	0.2265	0.0022	-0.0009	-0.0001	0.0228

Al analizar las matrices de sensibilidad se pudo observar que éstas son diferentes y que para cada punto de operación establecido en la turbina es necesario usar la correspondiente matriz. Por la razón anterior, para la implementación de un sistema de detección y diagnóstico en la turbina de gas sería necesario tener un banco de matrices y usar la correspondiente de acuerdo a las condiciones de operación.

Los residuos estructurados se calculan usando la siguiente ecuación:

$$R_s = S^{-1}[R]$$

6.8. Resultados

En el presente apartado se presentan los resultados obtenidos para fallos únicos, dobles y triples para el sistema propuesto, en condiciones de operación de 55000 rpm:

6.8.1. Fallos Simples

Las figura 6.6 a la 6.9 muestran los residuos obtenidos cuando ocurre un fallo. A partir de la matriz de sensibilidad que se muestra en la tabla 6.6 se obtiene de la matriz R_s , donde el valor de cada fila corresponde al porcentaje de fallo en cada componente. Si el valor de la primera fila es de 3, esto significa que hay un error de 3% en WC o fuga de masa de aire en el compresor. Si hay un valor de 5 en la fila 5 de la matriz R_s esto significa que hay un fallo de 5% en la eficiencia de la turbina.

La presencia de valores por encima del umbral de detección en diferentes filas de la matriz Rs es indicio de fallos múltiples.

Las figuras 6.6 a la 6.9 muestran los residuos que se presentan cuando hay fallos únicos en el sistema en estudio. Se pudieron detectar fallos por encima de 2% en todos los componentes.

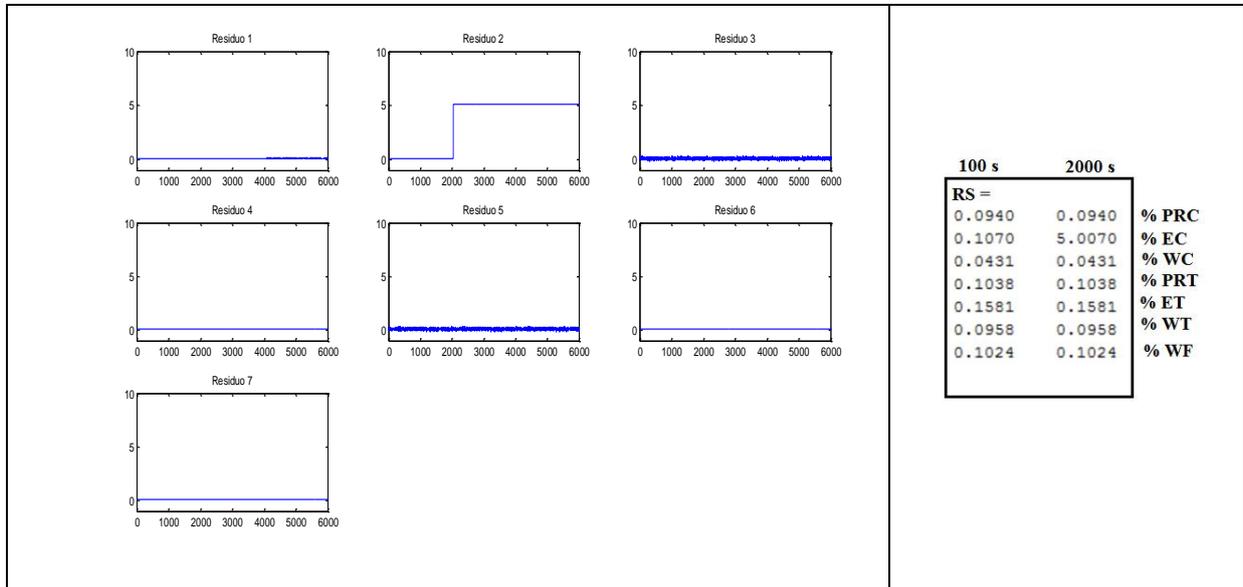


Figura 6. 6. Fallo de 5% en la eficiencia del compresor

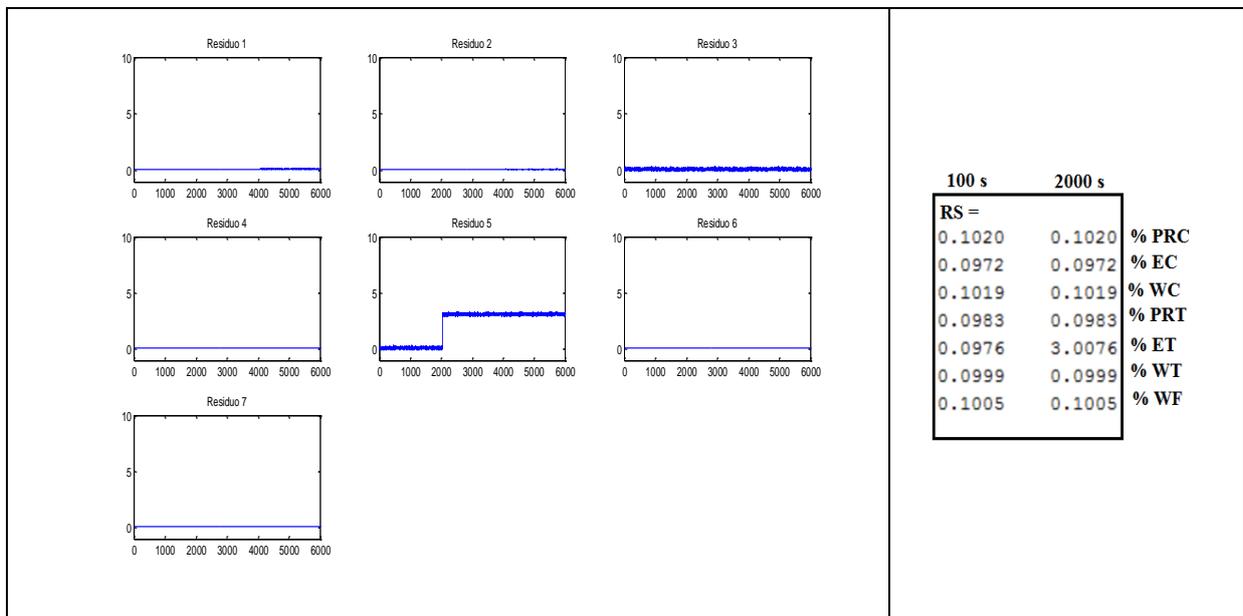


Figura 6. 7. Fallo de 3% en la eficiencia de la turbina

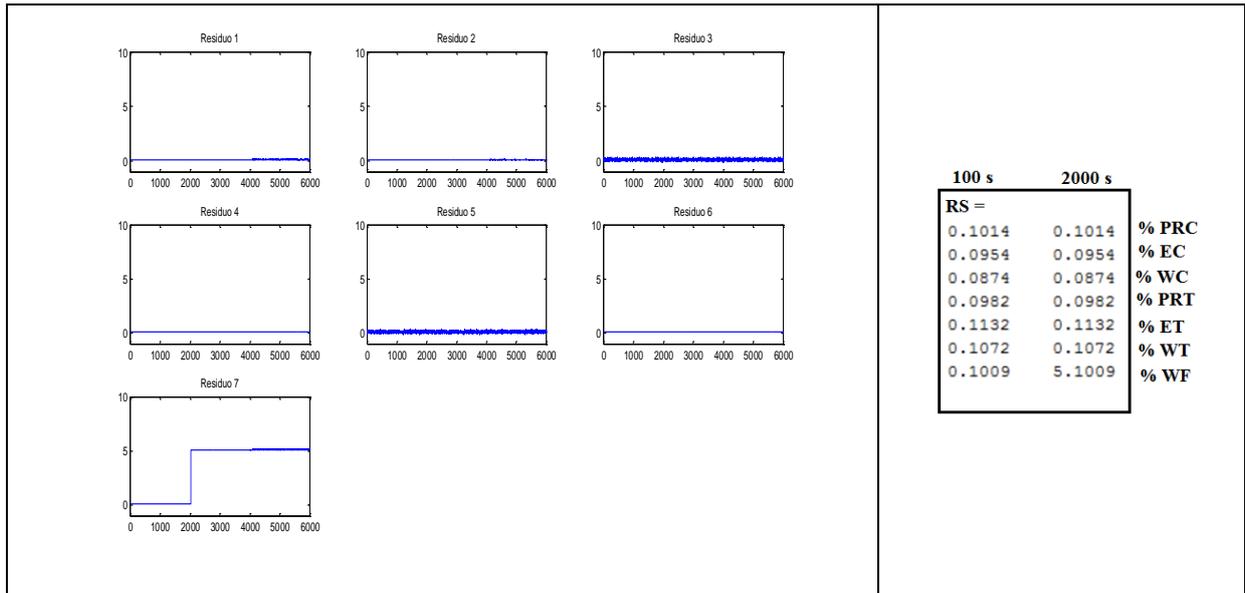


Figura 6. 8. Fallo de 5% en WF

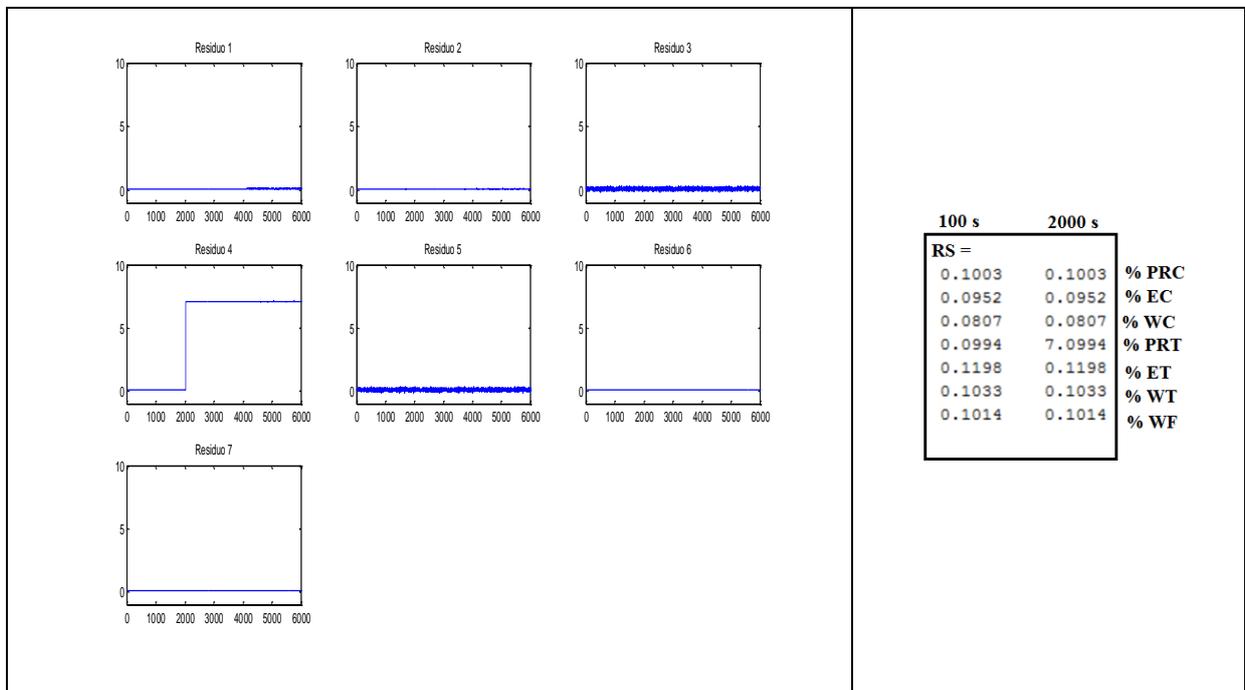


Figura 6. 9. Fallo de 7% de PRT

6.8.2. Fallos Dobles.

Para mostrar el funcionamiento de la metodología de NLPCA y residuos estructurados se simularon fallos dobles simultáneos y secuenciales. Las figuras 6.10 a la 6.14 muestran los residuos de la matriz Rs cuando se presentan fallos simultáneos y secuenciales.

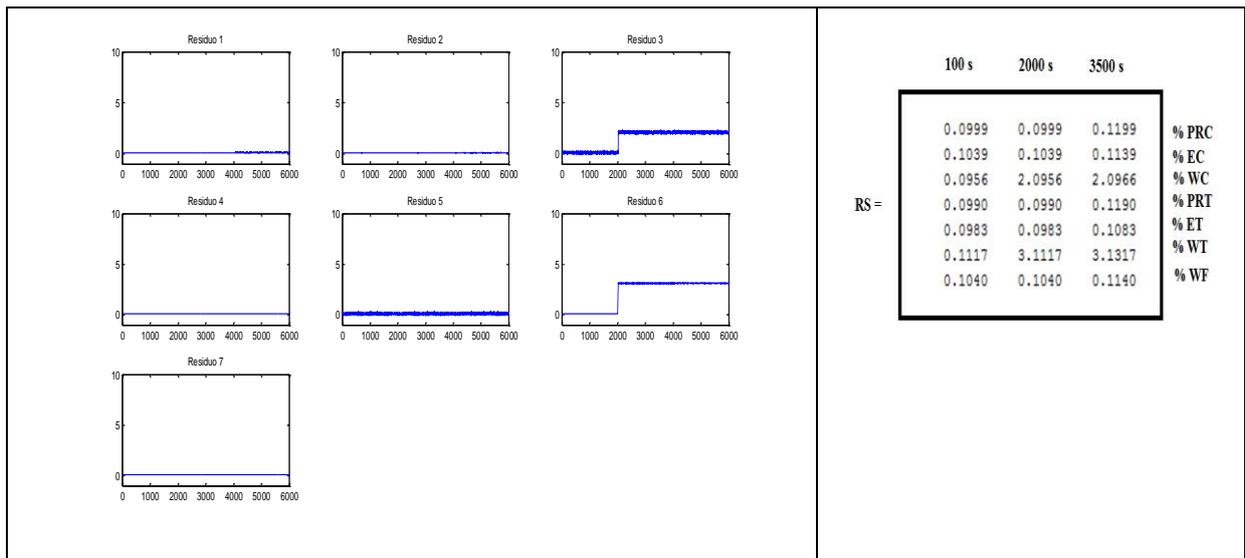


Figura 6. 10. Dos Fallos simultáneos del 2% WC y 3% WT

Al analizar la figura 6.10 se puede ver que los residuos R1, R2, R4, R5 y R7 permanecen en cero, mientras que los residuos R3 y R6 presentan valores de 2 y 3 respectivamente. Lo anterior indica la presencia de fallos múltiples simultáneos de 2% de fuga de aire en el compresor y 3% de fuga de masa de combustible en la turbina.

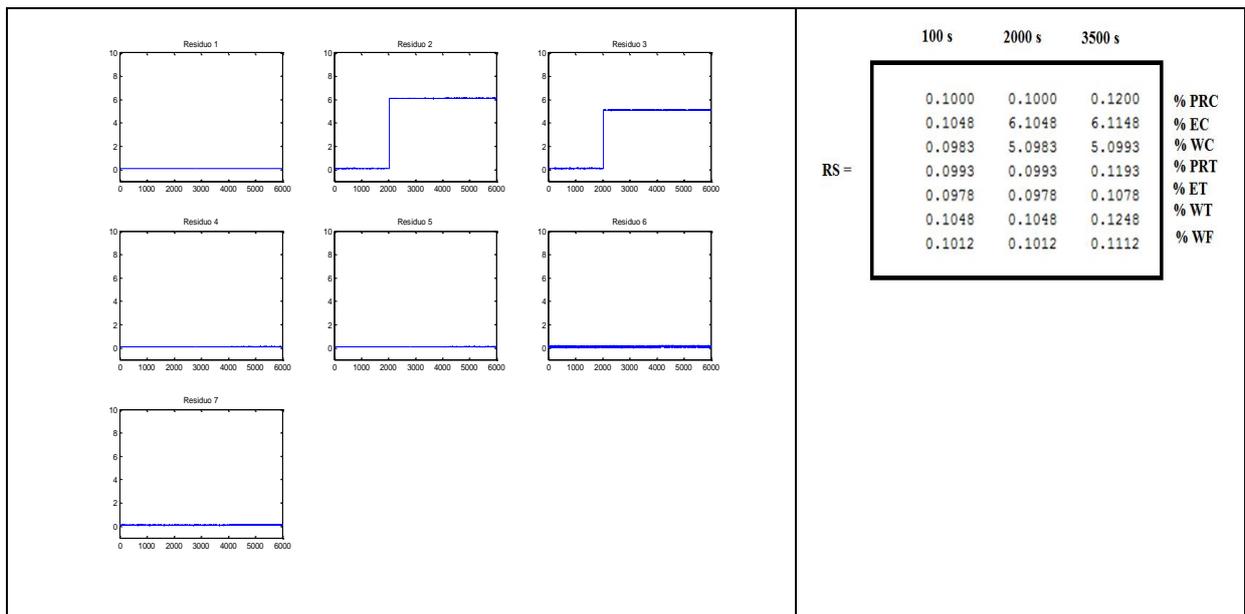


Figura 6. 11. Dos Fallos simultáneos del 5% WC y 6% EC

En la figura 6.11 se puede ver que los residuos R1, R4, R5, R6 y R7 permanecen en cero, mientras que los residuos R2 y R3 presentan valores de 5 y 6 respectivamente. Lo anterior indica la presencia de fallos múltiples simultáneos de 5% de fuga de aire en el compresor y 6% de reducción en la eficiencia del compresor.

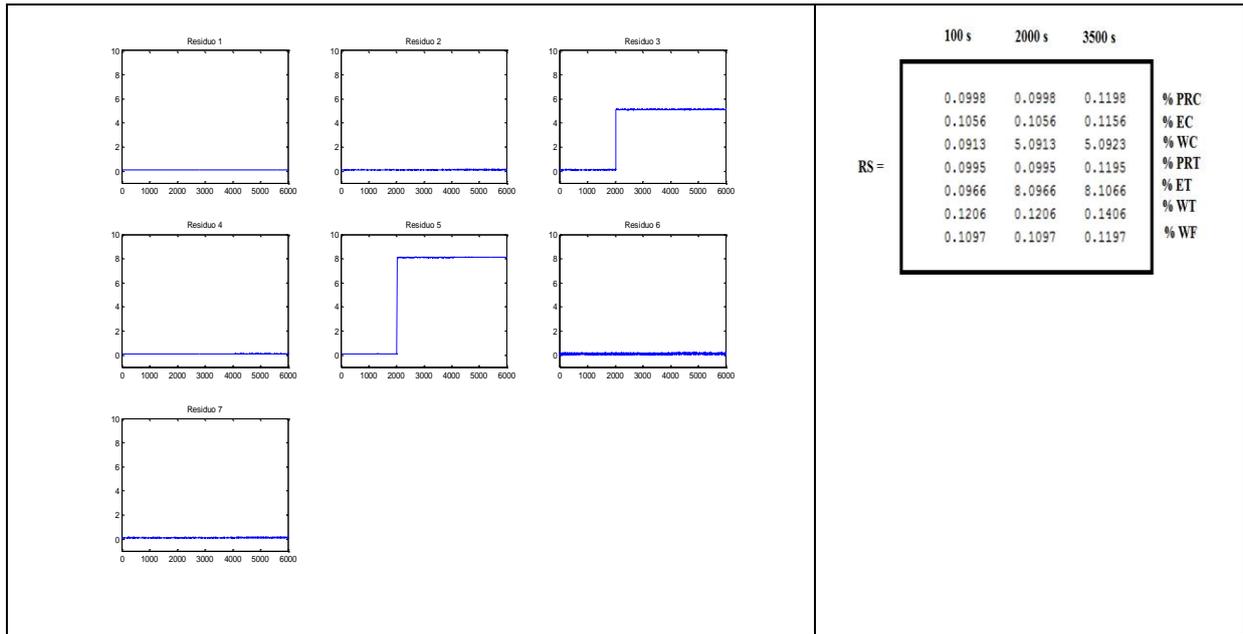


Figura 6. 12. Dos Fallos simultáneos 5% WC y 8% ET

La figura 6.12 muestra que los residuos R1, R2, R4, R6 y R7 permanecen en cero, mientras que los residuos R3 y R5 presentan valores de 5 y 8 respectivamente. Lo anterior indica la presencia de fallos múltiples simultáneos de 5% de fuga de aire en el compresor y 8% de reducción en la eficiencia de la turbina.

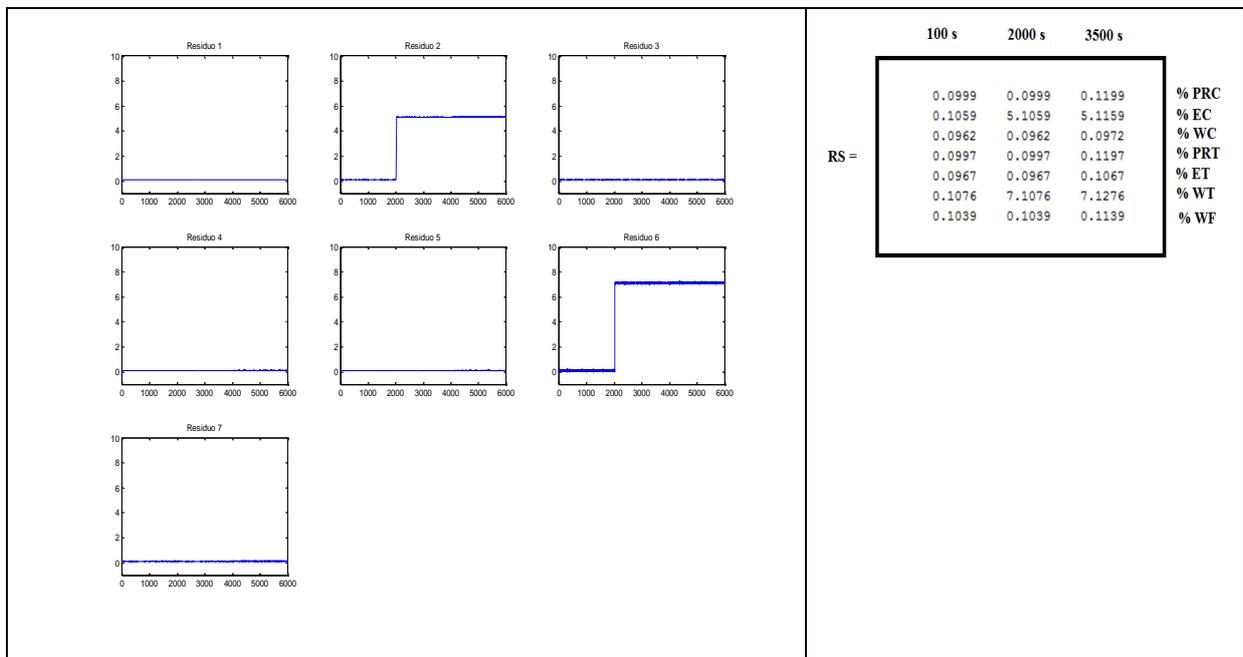


Figura 6. 13. Dos Fallos simultáneos 7% WT y 5% EC

En la figura 6.13 se puede observar que los residuos R1, R3, R4, R5 y R7 permanecen en cero, mientras que los residuos R2 y R6 presentan valores de 7 y 5 respectivamente. Lo anterior indica la presencia de fallos múltiples simultáneos de 7% de fuga de combustible en la turbina y 5% de reducción en la eficiencia del compresor.

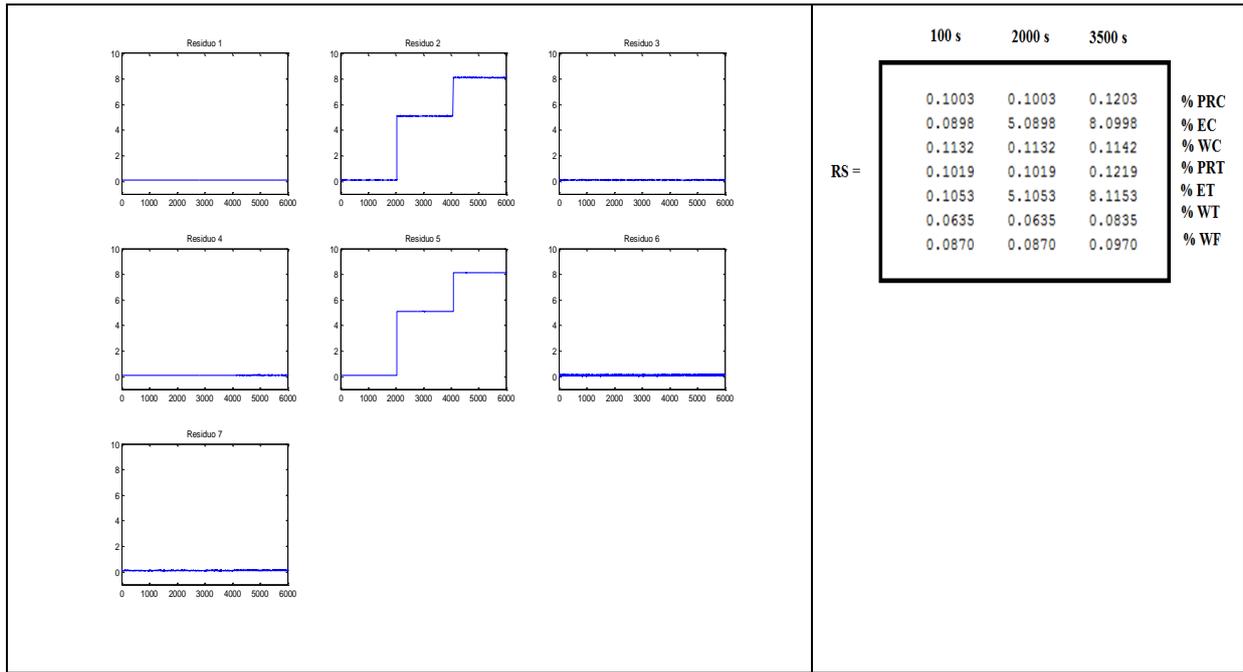


Figura 6. 14. Dos Fallos secuenciales 5%EC 8%EC y 5%ET 8%ET

En la figura 6.14 se muestra que los residuos R1, R3, R4, R6 y R7 permanecen en cero, mientras que los residuos R2 y R3 presentan valores de 5 y 5 a 2000 segundos y luego cambia a 8 y 8. Lo anterior indica la presencia de fallos múltiples secuenciales de 5% de fuga de aire en el compresor y 5% de reducción en la eficiencia del compresor en el instante 2000 segundos y luego a los 3500 segundos un fallo de 8% en la eficiencia del compresor y 8% de fuga de aire en el compresor.

6.8.3. Fallos Triples

Para mostrar el funcionamiento de la metodología de NLPCA y residuos estructurados se simularon fallos triples simultáneos y secuenciales. Las figuras 6.15 a la 6.18 muestran los residuos de la matriz Rs cuando se presentan fallos simultáneos y secuenciales.

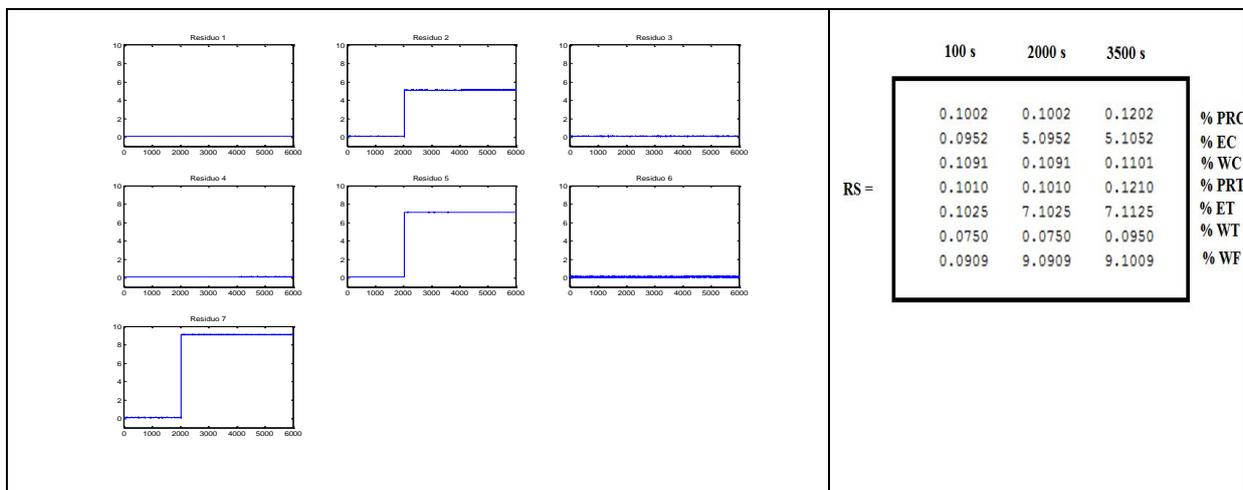


Figura 6. 15. Tres Fallos simultáneos 5%EC, 7%ET y 9%WF

En la figura 6.15 se puede observar que los residuos R1, R3, R6 y R7 permanecen en cero, mientras que los residuos R2, R5 y R7 presentan valores de 5, 7 y 9 respectivamente. Lo anterior indica la presencia de fallos múltiples simultáneos de 5% en la eficiencia del compresor, 7% en la eficiencia de la turbina y 9% en la válvula de combustible.

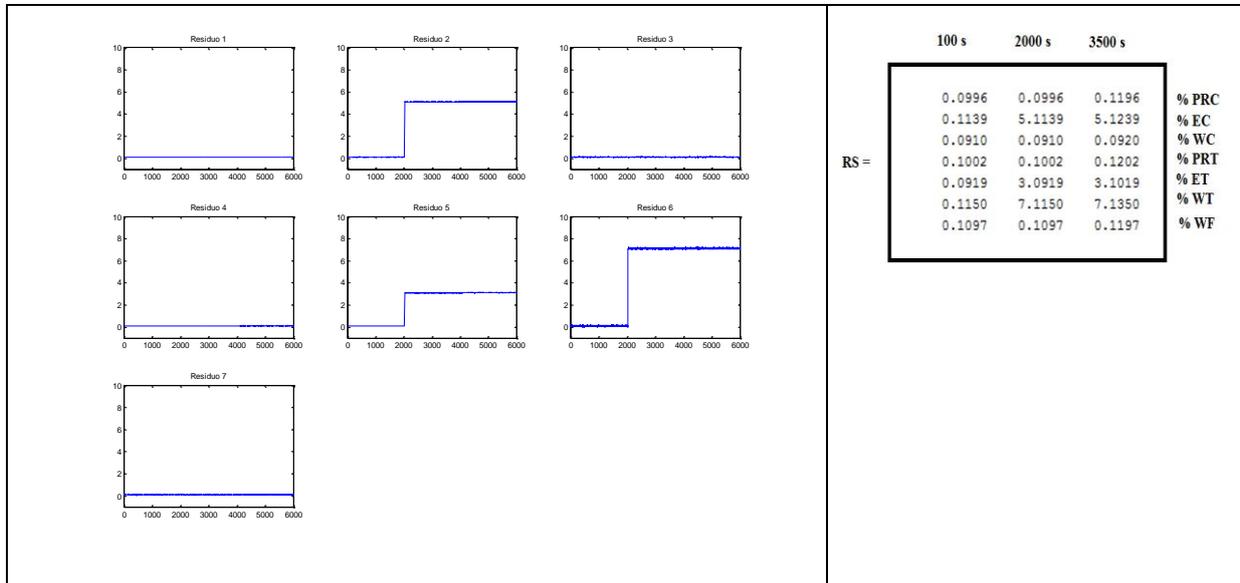


Figura 6. 16. Tres Fallos simultáneos 5%EC, 7%WT y 3%ET

En la figura 6.16 se puede observar que los residuos R1, R2, R4 y R7 permanecen en cero, mientras que los residuos R3, R5 y R6 presentan valores de 5, 7 y 3 respectivamente. Lo anterior indica la presencia de fallos múltiples simultáneos de 5% en la eficiencia del compresor, 3% en la eficiencia de la turbina y 7% en la masa de combustible en la turbina.

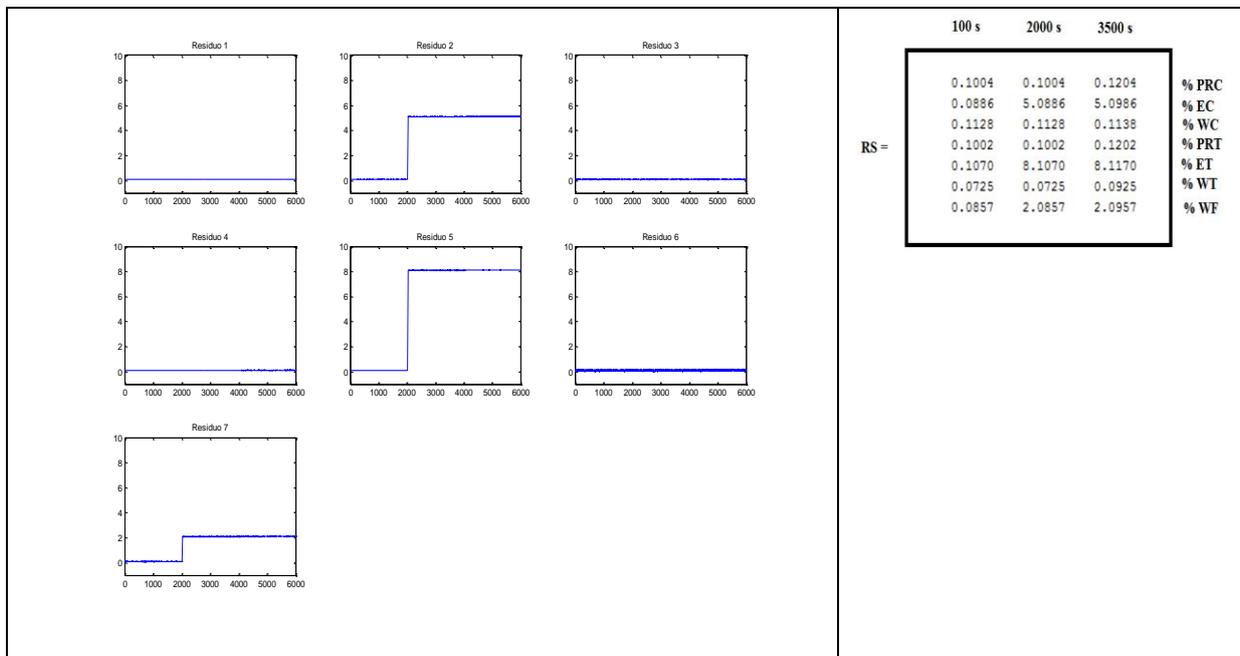


Figura 6. 17. Tres Fallos simultáneos 5%EC, 8%ET y 2WF

La figura 6.17 muestra que los residuos R1, R3, R4 y R7 permanecen en cero, mientras que los residuos R2, R5 y R6 presentan valores de 5, 8 y 2 respectivamente. Lo anterior indica la presencia de fallos múltiples simultáneos de 5% en la eficiencia del compresor, 8% en la eficiencia de la turbina y 2% en la válvula de combustible.

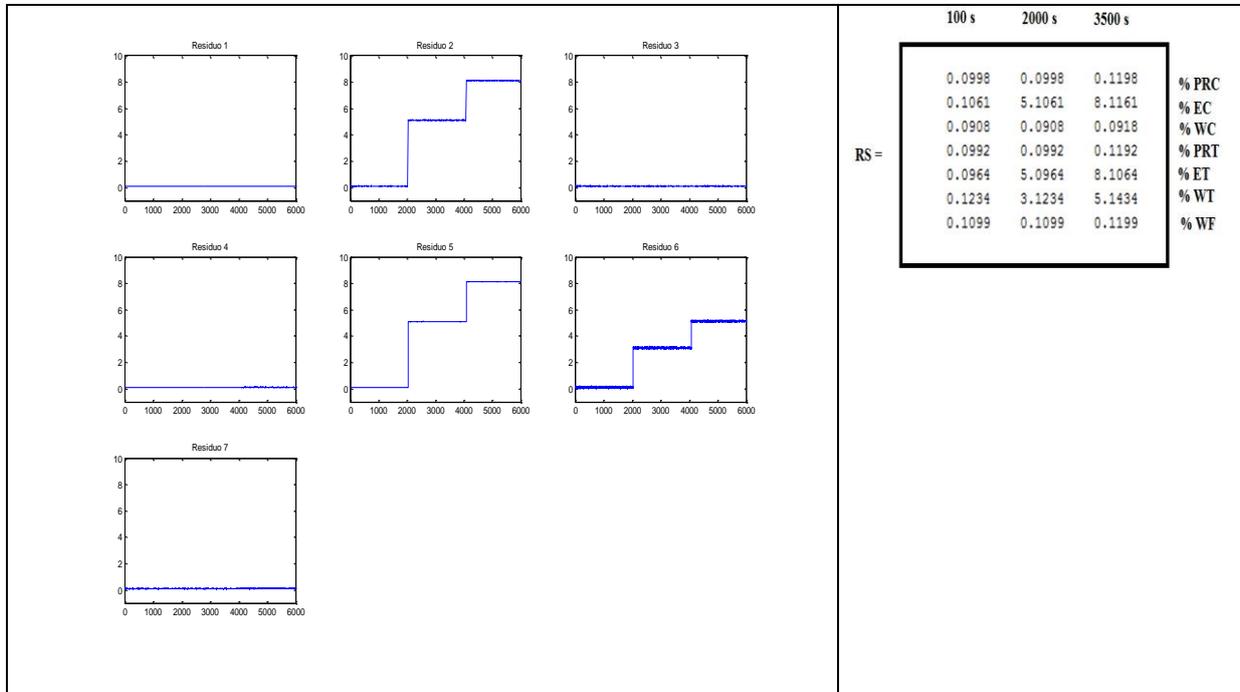


Figura 6. 18. Tres Fallos secuenciales 5%EC 8%EC, 5%ET 8%ET y 3%WT 5%WT

Finalmente en la figura 6.18 se puede observar que los residuos R1, R3, R4 y R7 permanecen en cero, mientras que los residuos R2, R5 y R6 presentan valores de 5, 5 y 3 para el instante de tiempo igual a 2000 segundos y 8, 5 y 5 para el instante de tiempo de 4000 segundos. Lo anterior indica la presencia de fallos múltiples simultáneos.

6.9. Conclusiones

Un método de detección de fallos múltiples y un esquema de aislamiento basado en la generación de residuos estructurados se ha estudiado para detectar y aislar los fallos en componentes y actuadores en una turbina de gas SR-30. Además un modelo en Simulink de la turbina de gas SR-30 se utilizó para generar residuos estructurados para el aislamiento de fallos.

Se estudiaron siete escenarios de fallos: Obstrucción o desgaste en el compresor, Rotura o desgaste mecánico en los alabes del compresor, Fuga de masa de aire en compresor, Obstrucción o desgaste en la turbina, Rotura o desgaste mecánico en los alabes de la turbina, Fuga en la cámara de combustión y Obstrucción en la válvula de combustible.

Los resultados de simulación muestran el buen comportamiento de la detección de fallos y sistema de aislamiento. Sin embargo, este esquema de detección y aislamiento de fallos no puede aislar los fallos

simultáneos para diferentes puntos de operación y cuando existen transitorios en las lecturas de los sensores, y esta es la línea de investigación abierta para el desarrollo de este método.

Al analizar las matrices de sensibilidad se observó que éstas son diferentes y que para cada punto de operación establecido en la turbina es necesario usar la correspondiente matriz. Por la razón anterior, para la implementación de un sistema de detección y diagnóstico en la turbina de gas sería necesario tener un banco de matrices y usar la correspondiente de acuerdo a las condiciones de operación.

6.10. Bibliografía

- [Fra- 90] Frank, P. (1990). Fault diagnosis in dynamics systems using analytical and knowledge based redundancy a survey. Automática.
- [Ger-07] Gertler, J. (2007). Fault Detection and Diagnosis in Engineering Systems". Marcel Dekker, 1998. Oscar Camacho. Fault Diagnosis based on multivariate statistical techniques. Rev. Téc. Univ. Zulia. Vol. 30, No 3, pp. 253-262.
- [Ger-05] Gertler, J. (2005) Design of optimal structured residuals from partial principal component models for fault diagnosis in linear systems. El Sevier. Journal of Process Control. 15 2005, pp. 585-603.
- [Hos-12] Hossein, M. y Poshtan, J. (2012). Fault detection and isolation using unknown input observers with structured residual generation.
- [Pui-06] Puig, V., Quevedo, J., Escobet, T. y Meseguer, J. (2006). Towards a Better Integration of passive Robust Interval-Based FDI Algorithms. 6th IFAC SAFEPROCESS.
- [Rin-12] Rincón, A. (2012). Multiple fault detection and diagnosis in a Gas Turbine using principal component analysis and structured residuals. 20th Mediterranean Conference on Control & Automation (MED 2012).
- [Rin-13] Rincón, A. Quevedo, J. (2013). Multiple Fault Detection and Diagnosis in Gas Turbine Using Nonlinear Principal Component Analysis and Structured Residuals. ASME 2013 International Mechanical Engineering Congress & Exposition.
- [Sal-04] Saldaña, M. (2004). Modelado y detección de fallas incipientes en transformadores usando redundancia analítica. Tesis de maestría Universidad Autónoma de Nuevo León.
- [Web-96] Webb, A. (1996). An approach to nonlinear principal component analysis using radially symmetric kernel functions. Statistic. Comput., vol. 6, pp. 159-168.
- [Wei-01] Weihua, L. y Sirish, S. (2001). Structured residual vector-based approach to sensor fault detection and isolation. El Sevier. Journal of Process Control 12. Pp. 429-443.

CAPITULO 7:

CONCLUSIONES Y TRABAJOS FUTUROS

Contenido

- 7.1. Introducción
- 7.2. Conclusiones
- 7.3. Trabajos Futuros

7.1. Introducción.

En este capítulo se presentan las conclusiones y principales aportes en el desarrollo de este trabajo. Se realiza una comparación de los diferentes métodos utilizados en base a las características deseables de un esquema de detección y diagnóstico de fallos múltiples. Adicionalmente, se incluyen las líneas futuras de investigación.

7.2. Conclusiones

Las aplicaciones de un esquema de detección y diagnóstico de fallos deben estar ajustadas al tipo de sistema: sea lineal, no lineal, tipo de información disponible, estudio de las limitaciones y restricciones de los sistemas simulados y/o sistemas reales, y además de considerar si la aplicación se hará en el ámbito académico o industrial.

En general hay diferentes técnicas y combinaciones de ellas utilizadas en la detección y diagnóstico de fallos múltiples, pero eso dependerá de la naturaleza del problema para seleccionar la técnica que tiene el mejor rendimiento. Todos estos métodos no son excluyentes entre sí, sino que son complementarios. La complejidad de la tarea para la que se diseñan hace que no tenga que ser abordada desde una única perspectiva.

Considerando que la mayoría de los sistemas complejos son complicados de modelar matemáticamente o mediante modelos de gráficos de unión, se ha enfocado esta tesis al uso de análisis de las señales, específicamente análisis de componentes principales no lineal como método de detección de fallos múltiples y los residuos estructurados como método de diagnóstico.

En un sistema complejo como por ejemplo lo es una turbina de gas, consistente de múltiples componentes interconectados, un fallo simple en un componente puede producir una respuesta anómala que puede servir como una excitación de entrada a otros componentes en buen estado y degradar el desempeño del sistema completo. En general, los fallos en una turbina de gas se clasifican en tres categorías principales: Pérdida de eficiencia en subsistemas; Degradación de los sensores; Fallos del actuador.

De lo analizado anteriormente resulta aconsejable en los casos en que las variables estén expresadas en diferentes unidades de medida trabajar con la matriz de correlaciones, para aislar así los efectos nocivos de las distintas unidades de medida.

Por otro lado, en los casos en que las variables se encuentren en la misma unidad de medida, debe asegurarse que la información que proveen las diferencias en las varianzas al análisis estadístico sea un elemento útil al propósito del análisis, para usar el enfoque de la matriz de varianzas y covarianzas. Esto es así, ya que al usar la matriz de correlación no solo se anulan los efectos de unidad de medida

si no toda otra información referente a la diferencia de varianza que, en algunos casos, podría aportar riqueza al estudio bajo análisis y su no inclusión podría distorsionar el significado y en análisis de las componentes.

Adicionalmente merece la pena destacar dos conclusiones importantes: En primer lugar, si las varianzas de las variables son mayores que uno, utilizar la matriz de correlaciones, que indirectamente estandariza las varianzas de todas las variables a uno, tiende a incrementar la proporción de la varianza total explicada por la primer componente principal.

En segundo lugar, si las varianzas son menores que uno (hecho que puede suceder cuando las variables están expresadas en porcentajes tal y como sucede en el ejemplo que ilustraremos al final de trabajo), utilizar la matriz de covarianzas reduce la proporción explicada por la primer componente dificultando así la validez del método, el proceso de selección del número de componentes y fundamentalmente su interpretación.

La aplicación de métodos estadísticos multivariantes en procesos con un gran número de variables donde la reducción del sistema es posible, permite simplificar los algoritmos matemáticos y obtener representaciones más sencillas que traen como ventaja en sistemas motorizados un mejor desempeño de los equipos de supervisión.

En los procesos industriales donde se tiene suficiente información de las variables es posible aplicar técnicas estadísticas, sin embargo, lograr una representación con menor número de variables depende de las propiedades estadísticas presentes en el proceso.

Alcanzar una representación en un espacio de dimensión reducida con la técnica de análisis de componentes principales requiere que exista correlación en algunas de sus variables, aquellas variables que están altamente correlacionadas tienen la misma representación. Si las mismas no están correlacionadas no es posible reducir la representación y esta técnica de análisis no es adecuada para el proceso que se evalúe.

Cuando las variables de un sistema están altamente correlacionadas la reducción de la dimensionalidad es posible. Por lo que la aplicación de las técnicas PCA y NLPCA tiene una mejor representación si algunas de las variables medidas están correlacionadas.

Para determinar los modelos PCA e NLPCA es necesario seleccionar un conjunto de datos adecuado. En el caso de PCA, la técnica es óptima en la captura de la variabilidad de las señales, por lo que considerar un conjunto de datos que presenten cambios en su punto de operación es más adecuado para la construcción de los modelos. Obteniéndose respuestas más robustas del esquema de

diagnóstico y detección de fallos ante diferentes conjuntos de señales, que operen en un punto nominal sin cambios o que estén cambiando su punto de operación.

Existen muchos criterios para determinar la cantidad de componentes principales lineales y no lineales por lo que en este trabajo se utilizó el criterio de varianza total, que para los casos de estudio proporcionó buenos resultados de forma muy simple. Sin embargo, se debe considerar que todos los criterios de selección del número de componentes introducen errores, por lo que es recomendable utilizar la base de la experimentación para determinar las mejoras cuantitativas en las estadísticas con más o menos componentes.

La aplicación de un esquema de detección y diagnóstico de fallos (DDF) debe estar ajustado al tipo de sistema lineal o no lineal, tipo de información disponible, estudio de las limitaciones y restricciones de los sistemas simulados y sistemas reales, y además considerar si la aplicación se hará en el ámbito académico o industrial.

Los métodos analíticos requieren un conocimiento a priori de las entradas y salidas del sistema, y de todas las formas de interacción con el ambiente para definir adecuadamente el número de fallos presentes y el diagnóstico. Sin embargo, en la práctica suele haber poca información del efecto del ruido y las perturbaciones externas. Esto en general, es una limitación de todos los métodos analíticos. Adicionalmente, a nivel industrial las aplicaciones de estos métodos han tenido poco impacto, dado que la mayoría de los procesos son no lineales, mientras que la teoría de los métodos analíticos está basada en el diseño lineal, y por ende no presenta ventajas ante otros métodos que se basan en análisis más simples.

Este estudio se aplicó en el ámbito académico a una turbina de gas con dinámica multivariable no lineal y de segundo orden, donde se obtuvieron buenos resultados. Aunque el proceso de optimización con la utilización del programa Simulink fue complejo y difícil de implementar, desde el punto de vista matemático, se dedujo que para un proceso con un orden superior el nivel de dificultad aumenta para implementar la solución. En los esquemas propuestos, estos métodos presentaron un buen desempeño y robustez ante los efectos de las no linealidades del sistema y el ruido.

No obstante, los requerimientos computacionales son más altos y el nivel de dificultad es mayor en el desarrollo matemático de la técnica NLPCA que en la técnica de PCA.

Se comprobó que para sistemas dinámicos no lineales las redes neuronales multicapa del NLPCA logran los objetivos de diseño con estructuras que requieren pocas neuronas en cada capa. Adicionalmente al utilizar el análisis de componentes principales no lineales con el diseño de los residuos estructurados se obtiene un diseño más eficiente y con mejores resultados cuando se quiere trabajar en procesos en línea.

Un método de detección de fallos múltiples y un esquema de aislamiento basado en la generación de residuos estructurados se ha estudiado para detectar y aislar los fallos en componentes y actuadores en una turbina de gas SR-30. Además un modelo en Simulink de la turbina de gas SR-30 se utilizó para generar residuos estructurados para el aislamiento de fallos.

Los resultados de simulación muestran el buen comportamiento del sistema de aislamiento mediante el análisis de sensibilidad de los residuos. Sin embargo, este esquema de detección y aislamiento de fallos no puede aislar los fallos simultáneos para diferentes puntos de operación y cuando existen transitorios en las lecturas de los sensores, y esta es la línea de investigación abierta para el desarrollo de este método.

En este estudio se aplicó la teoría de aislamiento mediante residuos estructurados y análisis de sensibilidad para el caso de fallos múltiples con resultados satisfactorios en una turbina real ubicada en la Universidad Interamericana de Puerto Rico.

7.3. Trabajos Futuros.

Las líneas futuras de investigación que se proponen para el futuro son:

Estudiar la adaptabilidad de los algoritmos propuestos al ámbito industrial ya que es de suma importancia que las metodologías propuestas sean fáciles de implementar y tengan poco costo computacional. La industria de hoy buscan soluciones eficientes y económicas, por lo tanto metodologías complejas y costosas no son atractivas.

Aplicar los algoritmos de detección y diagnóstico a sistemas más complejos resulta ser una de los retos que enfrentan la mayoría de los investigadores que se dedican al estudio de metodologías de detección y diagnóstico de fallos.

Investigar nuevas técnicas de detección y diagnóstico de fallos múltiples considerando el número de sensores óptimos y la posición adecuada de estos en un sistema dinámico.

Los métodos estadísticos son los más fáciles de implementar y con buenos resultados. Por lo que sería interesante estudiar variaciones en estos métodos estadísticos que se puedan utilizar para la tarea de detección y diagnóstico de fallos múltiples.

Investigar nuevas formas de diagnóstico de fallos que no necesiten datos del comportamiento de fallos para ser identificados, ya que estos datos no suelen estar disponibles en un sistema real.

Por último se recomienda investigar esquemas de detección y aislamiento de fallos que puedan aislar fallos simultáneos para diferentes puntos de operación y cuando existan transitorios en las lecturas de los sensores.

8. BIBLIOGRAFIA

Capítulo 1:

- [Ben-00] Benoff, D. (2000). Reliability Centered Maintenance Gives a new Meaning to the Phrase: If it can't Broke don't fix it, Business & Commercial Aviation Magazine.
- [Bla-02] Blázquez, L. (2002). Diagnóstico Automático de Fallos para Sistemas Dinámicos No Lineales. Publicación del Departamento de Ingeniería Eléctrica y Electrónica de la Universidad de León.
- [Bla-03] Blanke, M., Kinnaert, M. y Lunze, J.(2003). Diagnosis and Fault Tolerant Control. Springer.
- [Cam-07] Camacho, O. (2007). Fault Diagnostic base on multivariate statistical techniques. Revista técnica de la Universidad de Zulia. Vol. 30, No 3, pp. 253-262.
- [Cas-04] Casavola, A., Famularo, D. y Franze, G. (2004). Robust Multiple-fault detection and isolation: A gradient flow approach. Proceeding of the 2004 American Control Conference.
- [Dai-07a] Daigle, M. Koutsoukos, X. y Biswas, G. (2007). A Quantitative Approach to Multiple Isolation in Continuous Systems. Institute for Software Integrated Systems (ISIS). AAAI'07 Proceedings of the 22nd national conference on Artificial intelligence - Vol 1.
- [Dai-07b] Daigle, M. Koutsoukos, X. y Biswas, G. (2007). Multiple Diagnoses in Complex Physical Systems. Institute for Software Integrated Systems (ISIS).
- [Dav-92] Davis, R. y Hamscher, W. Model-based reasoning: Troubleshooting, in Hamscher, W., Console, L. y Kleer, J. (Eds.). (1992). Readings in Model-Based Diagnosis, Morgan Kaufmann Publishers, San Mateo, CA, pp. 3–24..
- [Fra-90] Frank, P. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy – A survey and some new results, Automática 26(3): pp.459–474.
- [Fra-96] Frank, P. (1996) Analytical and qualitative model-based fault diagnosis – A survey and some new results, European Journal of Control 2(1): pp. 6–28.

- [Gar-11] García, D., Fuente, M. (2011). Estudio Comparativo de detección de fallos basadas en Análisis de Componentes Principales. Elsevier.
- [Gas-02] Gastineau, Z. (2002). Propulsion Technology Planning for Engine Health Management, Controls and Engine Health Management, Turbine Engine Division Propulsion Directorate, Air Force Research Laboratory.
- [Ger-98] Gertler, J. (1998). Fault Detection and Diagnosis in Engineering Systems, Marcel Dekker, Inc. New York- Basel- Hong Kong.
- [Ger-10] Gertler, J., Romera, J., Puig, V. y Quevedo, J. (2010). Leak Detection and Isolation in water distribution network using Principal Component Analysis and Structured Residuals. Conference on Control and Fault Tolerant Systems.
- [Ham-92] Hamscher, W., Console, L. y Kleer, J. (1992) . Readings in Model-Based Diagnosis, Morgan Kaufmann, San Mateo, CA.
- [Ise-06] Isermann, R. (2006). Fault Diagnosis Systems. An Introduction from Fault Detection to Fault tolerance. Springer, Verlag. Germany.
- [Kor-04] Korbicz, J., Koscielny, J., Kowalczyk, Z. y Cholewa, W. (2004). Fault Diagnosis. Models, Artificial Intelligence, Applications, Springer-Verlag, Berlin.
- [Kos-04a] Koscielny, J. (2004). Methodology of process diagnosis, in Korbicz, J., Koscielny, J., Kowalczyk, Z. y Cholewa. Fault Diagnosis. Models, Artificial Intelligence, Applications, Springer-Verlag, Berlin, pp. 57–114. 2004.
- [Kos-04b] Koscielny, J. (2004). Models in process diagnosis, in Korbicz, J., Koscielny, J., Kowalczyk, Z. y Cholewa., Fault Diagnosis. Models, Artificial Intelligence, Applications, Springer-Verlag, Berlin, pp. 29–43.
- [Kem-06] Kempowsky, T., Subias, A., Aguilar, J. (2006) Process situation assessment: from a fuzzy partition to a finite state machine. Engineering Applications of Artificial Intelligence. Vol.19, N°5, pp.461-477.
- [Lig-08] Ligeza, J., Koscielny, J. (2008). A New Approach to Multiple Diagnosis: A Combination of Diagnostic matrices, Graphs, Algebraic and Rule-based Models. The case of two layer models. Int. Journal Appl. Math. Comput. Sci., Vol. 18, No.4, pp. 465-476.
- [Mor-07] Mora, J., Barrera, V. y Carrillo, G. (2007). Fault Location in Power Distribution

Systems Using a Learning Algorithm for Multivariable Data Analysis. IEEE Transactions on Power Delivery, VOL. 22, No. 3.

- [Nie-10] Nieto, J., Castañón, L. y Morales, R. (2010). Multiple Fault Diagnosis in Electrical Power Systems with Dynamic Load Changes Using Probabilistic Neural Networks. *Computación y Sistemas*. Vol. 14 No. 1.
- [Pui-06] Puigjaner, L., Ollero, P., de Prada, C. y Jiménez, L. (2006). Estrategias de modelado, simulación y optimización de procesos químicos. Editorial Síntesis.
- [Reg-83] Reggia, J., Nau, D. y Wang, P. (1983) Diagnostic expert system based on a set covering model, *International Journal on Man-Machine Studies* 19(5): pp. 437–460.
- [Reg-85] Reggia, J., Nau, D. y Wang, P. (1985). A formal model of diagnostic inference. Problem formulation and decomposition, *Information Sciences* 37(1–3): pp.227–256.
- [Rei-87] Reiter, R. (1987). A theory of diagnosis from first principles, *Artificial Intelligence* 32(1): pp. 57–95.
- [Tza-89a] Tzafestas, S. (1989). System fault diagnosis using the knowledge-based methodology, in Tzafestas S. *Knowledge-Based System Diagnosis, Supervision and Control*, Plenum Press, New York, NY/London.
- [Tza-89b] Tzafestas, S. (1989). *Knowledge-Based System Diagnosis, Supervision and Control*, Plenum Press, New York, NY/London.

Capítulo 2:

- [Bre-01] Bremer, C. y Kaplan, D. (2001). Markov chain and Monte Carlo estimation of nonlinear dynamics from time series. *Phys D*.
- [Bro-74] Broersen P. (1974). Estimation of parameters of nonlinear dynamical systems. *Int. J. Non-linear Mech.*
- [Bop-96] Boppana, V., Hartanto, I. y Fuchs, W. (1996). Fault diagnosis using state information. *Proceeding of the 26th annual international symposium on fault tolerant computing*.
- [Chi-05] Chin, S. y Ray, A. (2005). Symbolic time series analysis for anomaly detection: a comparative evaluation. *Signal process*.
- [Chi-06] Ching, J., Beck J. y Porter, K. (2006) Bayesian state and parameter estimation of uncertain dynamical systems. *Probab. Eng. Mech.*

- [Cor-04] Correcher, A. y García, E. (2004). Diagnóstico de Fallos Intermitentes basado en modelos de eventos discretos. WODES04.
- [Cor-10] Correcher, A. y García, E. (2010). Factory Automation. Chapter 29. Diagnosis of Intermittent Fault Dynamics. InTech.
- [Dav-02] David, B. y Bastin, G. (2002). Parameter estimation in nonlinear system with auto and cross correlated noise. Automatic.
- [Din-01] Ding, Y., Wu, Z. y Zhang, Y. (2001). Multi-fault diagnosis method based on a joint estimation of states and estimation parameters. J. Tsinghua Univ.
- [Fao-05] Faouzi, M., Mourot, G. y Ragot, J. (2005). Sensor Fault and Isolation of an Air Quality Monitoring Network using Nonlinear Principal Component Analysis. Proceeding of the 16th IFAC World Congress.
- [Fao-03] Faouzi, M., Mourot, G. y Ragot, J. (2003). Nonlinear PCA Combining Principal curves and RBF-Networks for Process Monitoring. Proceeding of the 42nd IEEE Conference on Decision and Control.
- [Gar-11] García, D. y Fuente, M. (2011). Estudio Comparativo de detección de fallos basadas en Análisis de Componentes Principales. Elsevier.
- [Ger-98] Gertler, J. (1998). Fault Detection and Diagnosis in Engineering Systems, Marcel Decker,(Inc. New York- Basel- Hong Kong.
- [Ger-99] Gertler, J., Li, W., Huang, Y. y McAvoy. (1999). Isolation Enhanced Principal Component Analysis. AIChE Journal, 45, pp. 323-334.
- [Ger-10] Gertler, J., Romera, J., Puig, V. y Quevedo, J. (2010). Leak Detection and Isolation in water distribution networks using Principal Component Analysis and Structured Residuals. Conference on Control and Fault Tolerant Systems (Systol).
- [Gha-01] Ghanem, R. y Romero, E. (2001). A wavelet-based approach for model and parameter identification of nonlinear systems. Int. J. Non-linear Mech.
- [Gup-07a] Gupta, S. y Ray, A. (2007). Fatigue damage monitoring by ultrasonic measurements: a symbolic dynamic approach. Int. J. Fatigue.
- [Gup-07b] Gupta, S. y Ray, A. (2007). Ray. Real time fatigue life estimation in mechanical systems. Meas. Sci. Technol.

- [Gaw-00] Gawthrop, P. (2000). Sensitivity bond graphs. *J. Franklin Inst.*, pp. 907-922.
- [Hof-02] Hoffman, A. y Van der Merwe, N. (2002). The application of neural network to vibrational diagnostics for multiple fault conditions. *Comput. Stand. Interfaces*.
- [Kor-04] Korbicz, J., Koscielny, J., Kowalczyk y Cholewa, W. (2004). *Fault Diagnosis. Models, Artificial Intelligence, Applications*, Springer-Verlag, Berlin.
- [Kem-06] Kempowsky, T., Subias, A. y Aguilar, J. (2006). Process situation assessment: from a fuzzy partition to a finite state machine” *Engineering Applications of Artificial Intelligence*. Vol.19, N°5, pp.461-477.
- [Kos-03] Koscielny, J. y Bartys, M. (2003). Multiple Faults Isolation in Diagnostic of Industrial Processes. *European Control Conference*.
- [Kos-09a] Koscielny, J. y Bartys, M. (2009). Application of the Method of Dynamic Decomposition for Recognition of Multiple Faults in the Large Scale systems. *7th Workshop on Advanced Control and Diagnostic ACD 2009 (a)*.
- [Kos-09b] Koscielny, J., Bartys, M. y Syfert M. (2009). Method of Multiple Isolation in Diagnostics of Large Systems. *7th Workshop on Advanced Control and Diagnostic ACD 2009 (b)*.
- [Kre-91] Kresta, J., , MacGregor, J. y Marlin, T.E. (1991). Multivariate Statistical Monitoring of processes, “*Can. J. Chem. Eng.*”, pp. 69- 35.
- [Lig-08] Ligeza, J. y Koscielny, J. (2008) A New Approach to Multiple Diagnosis: A Combination of Diagnostic matrices, Graphs, Algebraic and Rule-based Models. The case of two layer models. *Int. Journal Appl. Math. Comput. Sci.*, Vol. 18, No.4, 465-476.
- [Moh-07] Mohamed, H., Salah, D., Noureddine, D. y Mohamed, B. (2007). Sensor Fault Detection, Isolation and Reconstruction Using Nonlinear Principal Component Analysis. *International Journal of Automation and Computing*.
- [Mor-07] Mora-Flórez, J., Barrera-Núñez, V. y Carrillo-Caicedo, G. (2007). Fault Location in Power Distribution Systems Using a Learning Algorithm for Multivariable Data Analysis. *IEEE Transactions on Power Delivery*, VOL. 22, No. 3.

- [Moh-12] Mohamed, H. y Messaoud, B. (2012). Nonlinear Principal Component Analysis combining TT Net and RBF neural networks for process monitoring. Proceeding of the 20th Mediterranean conference on Control and Automation.
- [Mos-99] Montermán, P. y Biswas, G. (1999). Diagnosis of continuous valued systems in transients operating regions. IEEE Transaction on systems, Man and Cybernetics, Part A.
- [Oga-02] Ogaji, S., Singh, R. y Probert, S. (2002). Multiple-Sensor fault diagnoses for a 2 shaft stationary gas turbine.
- [Per-09] Pérez, R., Puig, V., Pascual J., Peralta A., Landeros, E. y Jordanas, Ll. (2009). Pressure sensor distribution for leak detection in Barcelona water distribution network". Water Science & Technology, Vol. 9, No 6, pp. 715–721.
- [Pul-05] Pulido, B., Puig, V., Escobet, T. y Quevedo, J. (2005). A new fault algorithm that improves the integration between fault detection and localization in dynamic systems. Decision and control 2005. European Control Conference.
- [Rei-87] Reiter, R. (1987). A theory of diagnosis from first principles, Artificial Intelligence 32(1): pp. 57–95.
- [Sch-12] Scholz, M. (2012). Validation of nonlinear PCA. Neural Proceeding Letters, Volume 36, Number 1. pp. 21-30.
- [Sch-02] Scholz, M. y Vigario, R. (2002). Nonlinear PCA: a new hierarchical approach. In M. AVerleysen, editor, Proceedings ESANN.
- [Sch-07] Scholz, M., Fraunholz, M. y Selbig, J. (2007). Nonlinear Principal Component analysis: Neural Network Model and Applications. In Principal Manifolds for Data Visualization and Dimension Reduction. Volume 58 of LNCSE, pp. 44-67. Springer Berlin Heidelberg.
- [Son-03] Song, H. y Zhang, H. (2003). Approach for multiple faults diagnosis based on parity equations and parameter estimation. Control Decis.
- [Sim-03] Simani, S., Fantuzzi, C. y Patton, R. (2003). Model based fault diagnosis in dynamical systems using identification techniques.
- [Tol-06] Tolani, D., Yasar, M. y Ray, A. (2006). Anomaly detection in aircraft gas turbine engines. J. Aerosp. Comput. Info. Commun.

- [Van-01] Van, L. y Witteveen, F. (2001). Multiple nonlinear parameter estimation using PI feedback control. Control engineering pract.
- [Ven-03] Venkatasubramanian, V. (2003). A review of process fault detection and diagnosis. El Sevier, Computer and Chemical Engineering 27 pp. 327-346.
- [Wis-96] Wise, B. y Gallagher, N. (1996). "The Process Chemometrics Approach to Chemical Process Fault Detection and Supervision," Journal of Process Control, 6(6), pp. 329-348.
- [Wan-00] Wan, E. y and Van der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. In proceedings of the IEEE symposium.
- [Wan-06] Wang, Y. y Geng, I. (2006). Bayesian network based fault section estimation in power systems. In proceeding of the IEEE region 10th annual international conference TENCON.
- [Yao-94] Yao, I. y Sethares, W. (1994). Nonlinear parameter estimation via the genetic algorithm. IEEE Trans. Signal Process.

Capítulo 3:

- [Ail-01] Ailer, P. y Santa, I. (2001). Nonlinear Model Building of a Low Power Gas. Periódica Polytechnica SER. Transp. eng. vol. 29, No. 1-2, pp. 117-135.
- [Cam-07] Camacho, O. (2007). Fault Diagnostic base on multivariate statistical techniques. Revista técnica de la Universidad de Zulia. Vol. 30, No 3, pp. 253-262.
- [Fre-03] French, K. (2003). Recycled Fuel Performance in the SR-30 Gas Turbine. Proceedings of the 2003 American Society for Engineering Education Annual Conference and Exposition.
- [Kim-00] Kim, S., Pilidis, P. y Yin, J. (2000). Gas Turbine Dynamic simulation using simulink?. Department of propulsion Engineering and power. Cranfield University, United Kingdom. 2000-01-3647. Society of Automotive Engineers.
- [Laz-01] Lazzaretto, A. y Toffolo, A. (2001). Analytical and Neural Network Model for Gas Turbine Design and Off-Design simulation. Int. J. Applied Thermodynamics. Vol. 4 (No. 4) pp173-182, December.
- [Leo-08] Leonard, O., Thomas, J. y Borguet, S. (2008). Ten Years of Experience with a Small Jet Engine as a Support for Education. Proceedings of ASME Turbo Expo 2008: Power for Land, Sea and Air GT2008, June 9-13, Berlin, Germany.
- [Per-03] Perez-Blanco, H. (2003). Activities around the SR-30 Minilab at PSU. Proceeding of the 2003 American Society for Engineering Education Annual Conference and Exposition copyright. American Society for engineering education.

- [Rah-11] Rahman, M., Ibrahim, T. y Abdalla, A. (2011). Thermodynamic performance analysis of gas-turbine power-plant. *International Journal of the Physical Sciences* Vol. 6(14), pp. 3539-3550, 18 July.
- [Sar-01] Saravanamuttoo, H. (2001). *Gas Turbine Theory*. Fifth Ed. Pearson Prentice Hall.
- [Wat-04] Watanabe, A., Ölçmen, S., Leland, R. y Whitaker K. (2004). Soft Computing Applications on SR-30 Turbojet Engine. 1st AIAA Intelligent Systems AIAA 2004-6444, Technical Conference Chicago, IN, September 20-22.
- [Wit-03] Witkowski, T., White, S. y Ortiz, C. (2003). Characterizing the performance of the SR30 Turbojet Engine. *Proceeding of the 2003 American Society For Engineering education Annual Conference and exposition*.

Capítulo 4:

- [Cam-07] Camacho, O. (2007). Fault Diagnostic base on multivariate statistical techniques. *Revista técnica de la Universidad de Zulia*. Vol. 30, No 3, pp. 253-262.
- [Chi-00] Chiang, L. (2000). *Fault detection and diagnosis in Industrial Systems*, Springer, Nueva York.
- [Doa-04] Doan, T., Khiang, L., y Liu, J. (2004). Comparative study of pca approaches in process monitoring and fault detection. *The 30th annual conference of the IEEE industrial electronics society*, pages 2594-2599.
- [Gar-11] García, D. y Fuentes, M. (2011). Estudio comparativo de técnicas de detección de fallos basadas en el análisis de componentes principales. *Revista Iberoamericana de Automática e Informática Industrial*. Elsevier.
- [Kre-91] Kresta, J. (1991). *Multivariate Statistical Monitoring of Process Operating Performance*. *The Canadian Journal of Chemical Engineering*.
- [Jac-91] Jackson, J. (1991). *A user's guide to principal components* . Wiley.
- [Kui-95] Ku, W., Storer, H. y Geogakis, Ch. (1995). *Disturbance Detection and Isolation by Dynamic Principal Component Analysis*. *Chemometrics and Intelligent Laboratory Systems*.
- [Mas-02] Mason, R., Young, J. (2002). *Multivariate Statistical Process Control with Industrial Applications*. ASA SIAM Series.
- [Min-06] Mina, J. y Verde, C. (2006). Proceso de Estandarización Adaptable para Detección de Fallas en Sistemas MIMO. *Congreso Nacional de Control Automático*. UNAM México DF.
- [Pen-02] Peña, D. (2002). *Análisis de Datos Multivariantes*. Mc Graw Hill.
- [Shl-05] Shlens, J. (2005). *A Tutorial on principal component analysis*. La Jolla, CA 92037: Salk Institute for Biological Studies.
- [Wei-01] Weighell, M. (2001). The statistical monitoring of a complex manufacturing process. *Journal of applied statistics*.

Capítulo 5:

- [Don-94] Dong, D. y McAvoy, T. (1994). Nonlinear principal component analysis - based on principal curves and neural networks. Proceedings of the American Control Conference, Baltimore, Maryland, pp. 1284-1288.
- [Dun-96] Dunia, R., Qin, S. y Edgar, T. (1996). Identification of faulty sensors using principal component analysis. AIChE Journal, vol. 42, N. 10, pp. 2797-2812, October 1996.
- [Has-89] Hastie, T. y Stuetzle, W. (1989). Principal curves. Journal of the American Statistical Association, vol. 84, N. 406, pp. 502-516.
- [Kra-91] Kramer, M. (1991). Nonlinear principal component analysis using auto-associative neural networks. AIChE Journal, vol. 37, N. 2, pp. 233-243.
- [Kre-91] Kresta, J., MacGregor, J. y Marlin, T. (1991). Multivariate statistical monitoring of process operating performance. Canadian J. Chem. Eng., vol. 69, N. 1, pp. 35-47.
- [Leb-94] LeBlanc, M. y Tibshirani, R. (1994). Adaptive principal surfaces. Journal of American Statistical Association, vol. 89, N. 425, pp. 53-64.
- [Tan-95] Tan, S. y Mavrovouniotis, M. (1995). Reduction data dimensionality through optimizing neural network inputs. AIChE Journal, vol. 41, N. 6, pp. 1471-1480.
- [Ton-95] Tong, H. y Crowe, C. (1995). Detection of Gross errors in data reconciliation by Principal Component Analysis. Process systems engineering, vol. 41, N. 7, July 1995.
- [Ver-02] Verbeek, J., Vlassis, N. y Krose, B. (2002). A k-segments algorithm for finding principal curves. Pattern Recognition Letters, vol. 23, pp. 1009-1017.
- [Web-96] Webb, A. (1996). An approach to nonlinear principal component analysis using radially symmetric kernel functions. Statist. Comput., vol. 6, pp. 159-168.

Capítulo 6:

- [Fra- 90] Frank, P. (1990). Fault diagnosis in dynamics systems using analytical and knowledge based redundancy a survey. Automática.
- [Ger-07] Gertler, J. (2007). Fault Detection and Diagnosis in Engineering Systems". Marcel Dekker, 1998. Oscar Camacho. Fault Diagnosis based on multivariate statistical techniques. Rev. Téc. Univ. Zulia. Vol. 30, No 3, pp. 253-262.
- [Ger-05] Gertler, J. (2005) Design of optimal structured residuals from partial principal component models for fault diagnosis in linear systems. El Sevier. Journal of Process Control. 15 2005, pp. 585-603.
- [Hos-12] Hossein, M. y Poshtan, J. (2012). Fault detection and isolation using unknown input observers with structured residual generation.

- [Pui-06] Puig, V., Quevedo, J., Escobet, T. y Meseguer, J. (2006). Towards a Better Integration of passive Robust Interval-Based FDI Algorithms. 6th IFAC SAFEPROCESS.
- [Rin-12] Rincón, A. (2012). Multiple fault detection and diagnosis in a Gas Turbine using principal component analysis and structured residuals. 20th Mediterranean Conference on Control & Automation (MED 2012).
- [Rin-13] Rincón, A. Quevedo, J. (2013). Multiple Fault Detection and Diagnosis in Gas Turbine Using Nonlinear Principal Component Analysis and Structured Residuals. ASME 2013 International Mechanical Engineering Congress & Exposition.
- [Sal-04] Saldaña, M. (2004). Modelado y detección de fallas incipientes en transformadores usando redundancia analítica. Tesis de maestría Universidad Autónoma de Nuevo León.
- [Web-96] Webb, A. (1996). An approach to nonlinear principal component analysis using radially symmetric kernel functions. *Statistic. Comput.*, vol. 6, pp. 159-168.
- [Wei-01] Weihua, L. y Sirish, S. (2001). Structured residual vector-based approach to sensor fault detection and isolation. *El Sevier. Journal of Process Control* 12. Pp. 429-443.

APENDICES:

Apéndice A. Programas en Matlab para procesar datos de la turbina.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%PROGRAMA PARA CALCULAR LAS COMPONENTES PRINCIPALES DE DATA OBTENIDA DEL
%SIMULADOR TURBINA DE GAS SR-30
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear
clc
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%CARGA DE LOS ARCHIVOS MAT 65000 RPM SIN FALLO
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
load SR30_SINFALLO_65000_2014.mat
data1=SR30_SINFALLO_65000_2014'; %Datos a 65000 rpm sin fallo
data1(1:1000,:)=[]; % Eliminando las primeras 1000 filas
data1(:,1)= []; % Eliminando la columna de tiempo
data1(:,1)= []; % Eliminando la columna de temperatura amb
data1(:,5)= []; % Eliminando la columna de presion amb
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%ESTANDARIZACION DE LOS DATOS SIN FALLO
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[n1 m1]=size(data1); % Tamaño de los datos sin fallo
data1mean=mean(data1); % Media de los datos sin fallo
Zdata1=data1-ones(n1,1)*data1mean; % Datos con media cero
data1std=std(data1); % Desviacion estandar de los datos sin fallo
Sdata1=Zdata1./(ones(n1,1)*data1std); % Estandarización

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%CONSTRUCCION DEL MODELO
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%a-Calculo de la matriz de covarianza
C=Sdata1'*Sdata1/(n1-1);%Matriz de Covarianza
%b-Calculo de los autovalores y Autovectores de la matriz de covarianza
[eigvec,lamda]=eigs(C,2);%Lamda es la matriz de autovalores decrecientes
npc=2; %en este caso, dos componentes principales para 90%
Vk=eigvec(:,1:npc);%Matriz de transformacion Vk
sigma=lamda(1:npc,1:npc);%sigma lamdak
%c-Calculo del limite de confidencialidad, Hotelling, SPE y scores.
T2lim=npc*(n1-1)/(n1-npc)*finv(0.95,npc,n1-npc);%alfa=0.95 niv. confidencia
T2lim_1=sqrt(sigma(1,1))*tinvt(0.95,n1-1);%Para el primer pca
zeta1=lamda(2,2);%segundo autovalor
zeta2=lamda(2,2)^2;
zeta3=lamda(2,2)^3;
h0=1-2*zeta1*zeta3/3/zeta2^2;
ca=norminv(0.95,0,1);% Valor de la distribucion normal 0.95 confidencia
SPElim=zeta1*(ca*h0*sqrt(2*zeta2)/zeta1+1+zeta2*h0*(h0-1)/zeta1^2)^(1/h0);
%d-Calculo de SCORES1
Y1=Sdata1*eigvec;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%e-Calculo de Hotelling
for k=1:n1
T11(k)=sum((Sdata1(k,:)*Vk*inv(sigma)*Vk'*Sdata1(k,:)',2);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%f-Calculo de SPE
Vb=Vk*Vk';
[nvk,mvk]=size(Vb);%Tamaño de Vb
A=(eye(nvk,mvk)-Vb);
for l=1:n1
r1=A*Sdata1(l,:);
SPE1(l)=r1'*r1;
end

```

Apendice B. Programa en Matlab para la regresiones

```

%PROGRAMA PARA LA REGRESION FOR MC, MT, EC, ET                                     %
%-----%
%                               Init Model                                       %
%-----%
clc
clear
format long

RHOof = 0.797;           % Fuel density, kg/lt
Tamb = 27+273.15;      % Ambient temperature, oK
Pamb = 101325;         % Ambient pressure, Pa

%----- DATA FILE LOAD -----%
%load datageib2.mat
%load datasubidajunio1.mat
load datareal.mat

time = data(:,1)';      % Time, seconds
dp1  = data(:,2)';      % Compressor inlet dynamic pressure, PSIG
pg2  = data(:,3)';      % Compressor exit gauge pressure, PSIG
pg3  = data(:,4)';      % Turbine inlet gauge pressure, PSIG
pg4  = data(:,5)';      % Turbine exit gauge pressure, PSIG
pg5  = data(:,6)';      % Nozzle exit gauge pressure, PSIG
wf   = data(:,7)';      % Fuel flow, Gal/hr
n    = data(:,8)';      % Rotational speed, rpm
ft   = data(:,9)';      % Thrust, Lbf
t1   = data(:,10)';     % Compressor inlet stagnation temperature, C
t02  = data(:,11)';     % Compressor exit stagnation temperature, C
t03  = data(:,12)';     % Turbine inlet stagnation temperature, C
t04  = data(:,13)';     % Turbine exit stagnation temperature, C
t05  = data(:,13)';     % Nozzle exit stagnation temperature, C
[n1 m1]=size(n);
x=[1:m1];
y=0*x+1;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%CONVERSION TO KPA, KG/S, K DEGREE.

Patm = Pamb*y;         % Ambient pressure, Pa
DP1  = (dp1*6894.75729); % Delta compressor inlet dynamic pressure,
Pa
P1   = (DP1+Patm);     % Compressor Inlet Pressure, pa
PG2  = (pg2*6894.75729)+Patm; % Compressor exit gauge pressure, Pa
PG3  = (pg3*6894.75729)+Patm; % Turbine inlet gauge pressure, Pa
PG4  = (pg4*6894.75729)+Patm; % Turbine exit gauge pressure, Pa
PG5  = (pg5*6894.75729)+Patm; % Nozzle exit gauge pressure, Pa
WF   = wf.*RHOof*3.78541178/3600; % Fuel flow, Kg/s
N    = n;              % Rotational speed, rpm
FT   = ft*4.44822;    % Thrust, N
T1   = t1 + 273.15;   % Compressor inlet stagnation temperature,
K
T02  = t02 + 273.15; % Compressor exit stagnation temperature,
K
T03  = t03 + 273.15; % Turbine inlet stagnation temperature, K
T04  = t04 + 273.15; % Turbine exit stagnation temperature, K
T05  = t05 + 273.15; % Nozzle exit stagnation temperature, K

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%COMPRESSOR PARAMETERS
WFC = Patm.*(WF./P1).*(sqrt(T1/Tamb));           %Corrected Fuel Flow. Kg/s
den=(P1)/(286.9*T1);                             %Air Density, Kg/m3 with
R=286.9 Nm/Kgk
vel=((2*dp1*6894.75)./den).^0.5;                 %Air Velocity, m/s
Ain=0.0031;                                       %Compressor inlet Area, m^3
Maire=den*vel*Ain;                               %Compressor Mass Air, Kg/s
Mairec=(Patm.*(Maire./P1).*(sqrt(T1/Tamb)));    %Compressor Mass Air Corrected
Ncorr=N./(sqrt(T1/Tamb));                       %Rotational Speed Corrected

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%ISENTROPIC EFFICIENCY OF THE COMPRESSOR
PRC=PG2./P1;                                     %Compressor Pressure Ratio (P2/P1)
T2s=T1.*(PRC.^(0.4/1.4));                       %Isentropic T2
EC=(T2s-T1)./(T02-T1);                         %Isentropic Efficiency of
the Compressor
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%TURBINE PARAMETERS
MTC=Mairec+WFC;                                 %Turbine Mass, Kg/s
%MTc=(MT.*sqrt(T03))./PG3;                     %Turbine Mass corrected
%ISENTROPIC EFFICIENCY OF THE TURBINE
PRT=(pg3+14.7)./(pg4+14.7);                   %Turbine Pressure Ratio
(P3/P4)
T4s=T03./(PRT.^(0.4/1.4));                     %Isentropic T4
ET=(T03-T04)./(T03-T4s);                     %Isentropic Efficiency of
the Turbine
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%REGRESION MC,N,PRC
x1=Ncorr';
x2=PRC';
y1=Mairec';
X = [ones(size(x1)) x1 x2 x1.*x2 x2.^2 x1.*x2.^2];
alpha = regress(y1,X);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%Plot the data and the model:
figure(1)
scatter3(x1,x2,y1,'filled')
hold on
x1fit = min(x1):1000:max(x1);
x2fit = min(x2):0.01:max(x2);
[X1FIT,X2FIT] = meshgrid(x1fit,x2fit);
YFIT = alpha(1) + alpha(2)*X1FIT + alpha(3)*X2FIT +
alpha(4)*X1FIT.*X2FIT+alpha(5)*X2FIT.^2+alpha(6)*X1FIT.*X2FIT.^2;
mesh(X1FIT,X2FIT,YFIT)
Mairecal=alpha(1) + alpha(2)*x1 + alpha(3)*x2 +
alpha(4)*x1.*x2+alpha(5)*x2.^2+alpha(6)*x1.*x2.^2;
xlabel('RPM')
ylabel('PRC')
zlabel('MC')
alpha
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%REGRESION EC,N
x11=Ncorr';

```

```

x22=Mairec';
y11=EC';
X11 = [ones(size(x11)) x11 x22 x11.*x22 x22.^2 x11.*x22.^2];
beta = regress(y11,X11);
figure(2)
scatter3(x11,x22,y11, 'filled')
hold on
x11fit = min(x11):1000:max(x11);
x22fit = min(x22):0.0001:max(x22);
[X11FIT,X22FIT] = meshgrid(x11fit,x22fit);
YFIT1 = beta(1) + beta(2)*X11FIT + beta(3)*X22FIT +
beta(4)*X11FIT.*X22FIT+beta(5)*X22FIT.^2+beta(6)*X11FIT.*X22FIT.^2;
mesh(X11FIT,X22FIT,YFIT1)
xlabel('RPM')
ylabel('MC')
zlabel('EC')
beta
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%REGRESION MT,N,PRC
x12=Ncorr';
x23=PRT';
y12=MTC';
X12 = [ones(size(x12)) x12 x23 x12.*x23];
gama = regress(y12,X12);
figure(3)
scatter3(x12,x23,y12, 'filled')
hold on
x12fit = min(x12):1000:max(x12);
x23fit = min(x23):0.01:max(x23);
[X12FIT,X23FIT] = meshgrid(x12fit,x23fit);
YFIT2 = gama(1) + gama(2)*X12FIT + gama(3)*X23FIT + gama(4)*X12FIT.*X23FIT;
mesh(X12FIT,X23FIT,YFIT2)
MTcal=gama(1) + gama(2)*x1 + gama(3)*x2 + gama(4)*x1.*x2;
xlabel('RPM')
ylabel('PRT')
zlabel('MT')
gama

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%REGRESION ET,N,PRC
x13=Ncorr';
x24=PRT';
y13=ET';
X13 = [ones(size(x13)) x13 x24 x13.*x24];
delta = regress(y13,X13);
figure(4)
scatter3(x13,x24,y13, 'filled')
hold on
x13fit = min(x13):1000:max(x13);
x24fit = min(x24):0.0001:max(x24);
[X13FIT,X24FIT] = meshgrid(x13fit,x24fit);
YFIT3 = delta(1) + delta(2)*X13FIT + delta(3)*X24FIT +
delta(4)*X13FIT.*X24FIT;
mesh(X13FIT,X24FIT,YFIT3)
xlabel('RPM')
ylabel('MT')
zlabel('ET')
delta

```