



**Tesis en opción al título académico
de Máster en Automática**

**Título: *Plataforma de soporte a la detección de
Estados Estacionarios con fines de
monitorización en la CTE Felton***

Autor: David Díaz Martínez

Tutor: P. A. MSc. Ing. Luis Vázquez Seisdedos

***Santiago de Cuba
Diciembre 2014***



Pensamiento

Solo el ignorante cree en la suerte, el hombre fuerte cree en causa y efecto.

Emerson



Dedicatoria

Dedico este trabajo a mi padre, a quien además de mi existencia le debo todo lo que soy y lo que pueda ser mientras viva. Sólo espero que comprenda que mis ideales, esfuerzos y logros han sido también suyos e inspirados en él.



Agradecimientos

Al culminar una meta deseada, necesariamente se deben tener presente aquellas personas que por su constancia, dedicación y amor contribuyeron a que llegar hasta el final se convierta en una realidad. Por lo tanto, doy fe de gratitud:

- A mis padres por ser excelentes ejemplos de sacrificio y esfuerzo, por inspirarme a ser mejor y hacer de mí una persona útil, por dedicar todo su empeño a mi educación, por la mejor herencia que me han dado, pues con este triunfo compruebo que fue mejor enseñarme a pescar que dármele pescado.

- Al resto de la familia por la ayuda brindada de una forma u otra.

- A mi tutor Luis Vázquez Seisdedos por ser parte fundamental en mi formación desde el inicio, por nutrirme con su experiencia y sabios consejos, por confiar en mí y demostrarme que en él puedo confiar.

- A Álvaro Aguilera Castillo por considerarme y tratarme como su propio hijo.

- A mis excelentes profesores (compañeros de trabajo) por su apoyo, colaboración y sus valiosos aportes a mi formación profesional.

A los nombrados y demás personas que verdaderamente merecen este sincero agradecimiento y que imperdonablemente he olvidado, sin ellos llegar a la meta hubiera sido mucho más difícil.

A todos muchas gracias.



Resumen

Como consecuencia del continuo desarrollo en equipos de recogida de datos y sistemas informáticos de almacenaje, la cantidad de datos que se producen durante la operación continua de procesos industriales que son guardados se ha incrementado enormemente. Diversos autores han reconocido el potencial de conocimiento posible a descubrir a partir de estos para asistir en la operación y mejoras aplicables al proceso.

El siguiente trabajo constituye una herramienta para desarrollar, adaptar y/o extender técnicas de análisis de datos que ayuden a procesar estos grandes volúmenes de información que se denominan como minas de datos para extraer información útil y con ella soportar la aplicación de diversas tareas. La realización de este trabajo se centrará en estos problemas de obtención, manipulación y utilización de la información a partir de datos de operación.

El Explorador Industrial Multivariable (IME por sus siglas en inglés Industrial Multivariate Explorer) desarrollado sobre la plataforma MatLab™ permite utilizar una serie de herramientas las cuales están todas integradas en una interfaz de usuario para la extracción del conocimiento a partir de las minas de datos del proceso con el objetivo principal de realizar la detección automatizada de los Estados Estacionarios (EE) correspondientes a un proceso industrial (generadores de vapor de la CTE Felton) y de este modo determinar las desviaciones existentes entre estos valores (medidos) y los esperados para así monitorizar el efecto a largo plazo de las degradaciones de los componentes y equipos sobre el correcto funcionamiento del proceso.



Abstract

As a result of continuous development tools for data collection and storage systems, the amount of data produced during the continuous operation of industrial processes that they are stores has increased enormously. Several authors have recognized the potential of knowledge possible to reveal in these data to assist in the operation and improvements applicable to the process.

The following paper is a tool to develop, adapt and / or extend data analysis techniques to help process these big information volumes that they are denominated as data mining to extract useful information and with it support the implementation of various tasks. The realization of this work will focus on these problems of obtaining, handling and use of information from operational data.

The Industrial Multivariate Explorer (IME) developed on the MatLab™ platform enables a variety of tools which are all integrated in a graphic interface for extracting knowledge from process data mining with the aim of making the automated detection of Stationary States (SS) corresponding to a industrial process (steam generators of CTE Felton) and this way determining the existent deviations between these values (measured) and the expected for that way monitoring the long period effect of the degradations of the components and equipments on the correct operation of the process.



Índice General

| | |
|--|----|
| Introducción..... | 1 |
| Capítulo 1. Datos y su estructura en procesos industriales automatizados: La detección de estados estacionarios aplicada a la monitorización del desempeño. | 8 |
| Introducción..... | 8 |
| 1.1 Monitorización y análisis de los datos de procesos industriales..... | 8 |
| 1.1.1 Sistemas y métodos para la historización de los datos | 8 |
| Sistemas de Control Distribuido..... | 10 |
| Servidores OPC | 11 |
| 1.2 Análisis de procesos químicos: La quimiometría | 11 |
| 1.3 La calidad de los datos medidos..... | 13 |
| 1.3.1 Errores presentes en la medición..... | 13 |
| 1.3.1.1 Errores gruesos. Métodos para su detección..... | 15 |
| 1.3.1.2 Valores perdidos. Métodos para su tratamiento | 17 |
| 1.3.1.3 Datos atípicos..... | 20 |
| 1.3.2. Ruido en las mediciones | 22 |
| 1.3.2.1 Ruido blanco | 25 |
| 1.3.2.2 Ruido coloreado | 26 |
| 1.3.2.3 Técnicas de reducción de ruido | 27 |
| 1.3.2.4 Transformada Wavelet | 27 |
| 1.4 Técnicas de procesamiento de los datos..... | 29 |
| 1.4.1 Reconciliación de Datos..... | 31 |
| 1.4.2 Regresión no paramétrica | 34 |
| 1.4.3 Interpolación Polinómica | 35 |
| 1.4.3.1 Interpolación de Hermite..... | 36 |
| 1.5 Los estados estacionarios en la operación de los procesos industriales..... | 38 |
| 1.5.1 Métodos y técnicas estadísticas en la detección de estado estacionario..... | 39 |
| 1.6 Monitorización del desempeño con fines de diagnóstico. | 43 |
| 1.6.1 Necesidad de la monitorización de los efectos de las degradaciones en Centrales Termoeléctricas. | 45 |
| 1.6.2 Caso particular del generador de vapor de la unidad # 2 de la CTE Felton..... | 47 |



| | |
|--|----|
| Conclusiones del capítulo..... | 51 |
| Capítulo 2. El IME: una plataforma flexible para la experimentación industrial | 52 |
| Introducción..... | 52 |
| 2.1. Generalidades del Explorador Industrial Multivariable | 52 |
| 2.2. Análisis del módulo de DEE | 54 |
| 2.2.1. Estimación del ruido..... | 55 |
| 2.2.1.1 Estimación de ruido mediante Transformada Wavelet..... | 56 |
| 2.2.2. Eliminación del ruido (filtrado o suavizado) | 58 |
| 2.2.2.1. Filtro de Media Móvil..... | 59 |
| 2.2.3. Detección de Estados Estacionarios | 59 |
| <i>Fraccionamiento de la señal en intervalos de EE</i> | 60 |
| <i>Rectificación de coordenadas y verificación de tolerancia</i> | 63 |
| 2.2.4. Detección de coincidencias (DC) | 64 |
| 2.2.4.1. Detector de Coincidencias basado en lógica AND..... | 65 |
| 2.3. Análisis del módulo de Detección de Desviaciones | 68 |
| 2.3.1. Monitorización de las desviaciones a largo plazo..... | 69 |
| Construcción de polinomio interpolador mediante interpolación de Hermite | 70 |
| 2.3.1.2. Cálculo de las desviaciones por variable..... | 72 |
| 2.3.2. Organización y Agrupamiento por fechas..... | 73 |
| 2.4. Aplicación de los algoritmos al generador de vapor # 2 de la CTE Felton | 73 |
| Conclusiones del capítulo..... | 78 |
| Conclusiones Generales | 79 |



Listado de abreviaturas

| | |
|--|---|
| BAA: B omba de A gua de A limentación | PLS: P artial L east S quares |
| BDP: B ase de D atos P rimaria | RAFA: R ank A nnihilation F actor A nalysis |
| BDS: B ase de D atos S ecundaria | RB: R uido B lanco |
| CAP: C alentadores de A lta P resión | RC: R uido C oloreado |
| CAR: C alentador de A ire R egenerativo | PCR: P rincipal- C omponent R egression |
| COM: C omponent O bject M odel | RH: R e- H heater |
| CBP: C alentadores de B aja P resión | SCADA: S upervisory C ontrol A nd D ata A cquisition |
| CTE: C entral T ermo E létrica | SICC: S erial I dentification with C ollective C ompensation |
| DC: D etector de C oincidencias | SNR: S ignal- N oise R atio |
| DCS: D istributed C ontrol S ystems | SSDetector: S teady S tate D etector |
| DEE: D etección de E stado E stacionario | TFA: T arget F actor A nalysis |
| Deval: D eviations E valuator | VRG: V entilador de R ecirculación de G ases |
| DFT: D iscrete F ourier T ransform | VTI: V entilador de T iro I nducido |
| DWT: D iscrete W avelet T ransform | VTF: V entilador de T iro F orzado |
| EE: E stado E stacionario | UBET: U nbiased E stimated T echnique |
| EFA: E nvolving F actor A nalysis | UGE: U nidad de G eneración E létrica |
| ET: E stado T ransitorio | WFA: W indows F actor A nalysis |
| ECO: E COnomizador | WSS: W ide S ense S tationary |
| FFT: F ast F ourier T ransform | 1D: U ni- D imensional |
| IME: I ndustrial M ulti-variate E xplorer | |
| MAD: M edian A bsolute D eviation | |
| MAR: M issing a t R andom | |
| MCAR: M issing C ompletely a t R andom | |
| MNAR: M issing N ot a t R andom | |
| MNI: M issing N on- I gnorable | |
| PCA: P recalentador de A ire | |
| PDF: P robability D ensity F unction | |
| PLC: P rogramable L ogic D evice | |



Índice de figuras

Figura 1.1: Estructura simplificada de los Sistemas de Control Distribuido.....10

Figura 1.2: Errores típicos en la medición como ruido gaussiano.....14

Figura 1.3: Errores gruesos en la medición.....15

Figura 1.4: Gráfico de cajas múltiple de los residuos estandarizados frente al factor.....21

Figura 1.5: Representación de señales: a) Original b) Efecto del ruido.....23

Figura 1.6: Ilustración de a) ruido blanco b) su espectro de potencia.....25

Figura 1.7: a) Señal con ruido rosado b) Su espectro de amplitud.....26

Figura 1.8: a) Señal con ruido marrón b) Su espectro de amplitud.....26

Figura 1.9: Representación de un Sistema de Reconciliación de datos.....31

Figura 1.10: Aplicaciones de un Sistema DR.....32

Figura 1.11: Arreglo típico entre un Sistema de Control Distribuido y la Reconciliación de Datos, la Simulación y la Optimización.....33

Figura 1.12: Disposición de los subprocesos del Generador de Vapor de la unidad 2 de Felton.
.....48

Figura 2.1: Estructura general del IME.....53

Figura 2.2: Prestaciones del módulo de DEE.....55

Figura 2.3: Descomposición wavelet de nivel 3.....58

Figura 2.4: Metodología de trabajo del algoritmo de fraccionamiento.....61

Figura 2.5: Prolongación del segmento candidato para el cálculo de la media móvil.....63

Figura 2.6: Rectificación y eliminación de muestras no estacionarias.....64

Figura 2.7: Vectores de prueba para la determinación de la efectividad en la determinación de segmentos coincidentes.....66

Figura 2.8: Resultado en la determinación de segmentos coincidentes sobre las secuencias a:
a) Para $t_{min} \geq 0$ y b) $t_{min} \geq 200$ 67

Figura 2.9: Prestaciones del módulo de Detección de Desviaciones.....68

Figura 2.10. Resultados de la detección de EE para la Potencia.....75

Figura 2.11. Resultados de la detección de EE para el flujo de agua al generador de vapor...75

Figura 2.12. Resultados de la detección de EE para el flujo de combustible.....75

Figura 2.13: Desviaciones de: a) flujo combustible, b) Flujo de aire a los quemadores.....76

Figura 2.14: Desviaciones de: a) temperatura en gases de salida a la chimenea b) presión de vapor a la entrada del CAP.....76



Introducción

La evolución de los sistemas de automatización hacia los denominados Sistemas de Control Distribuido (DCS del inglés Distributed Control Systems) así como la facilidad de establecer compactos sistemas registradores de datos conducen a la generación de grandes volúmenes de datos que son almacenados en estructuras conocidas como bases de datos ya que, además de las mediciones instrumentales, se incluyen: el tiempo absoluto en que fueron registradas, etiquetas, unidad de medida, denominación, etc.

Es conocido que dentro de los registros históricos obtenidos a partir de mediciones instrumentales en procesos industriales existe potencial información que permite encontrar los factores que afectan la operación de la planta mediante la identificación de aquellas variables que están influyendo sobre la eficiencia y el comportamiento de la misma, sin embargo, este potencial está enmascarado dentro de estos grandes volúmenes de datos presentes en las bases de datos diarias, haciendo que sea difícil su análisis, usando métodos manuales. La supervisión del proceso implica el monitoreo continuo de variables de proceso en busca de anomalías que pueden representar problemas de operación o de afectación de los parámetros para los cuales fue diseñada o re-potenciada. El diagnóstico de anomalías es una sub-tarea de este análisis. Los datos transmitidos por los sensores de la planta al sistema de control pueden ser analizados para verificar la operación correcta y predecir el comportamiento futuro.

Las técnicas avanzadas de monitoreo pueden jugar un papel en el mejoramiento de la eficiencia energética. En particular, el diagnóstico termo-energético contribuye a la determinación de la variación del consumo de combustible, la identificación de las causas de estos incrementos a partir de condiciones de diseño y la cuantificación del efecto de cada una de estas causas.

La industria energética está trabajando en reducir los costos de generación por la adopción de nuevas estrategias de mantenimiento. Esto ha desarrollado gran interés en nuevos métodos de diagnóstico y pronóstico oportuno para anticiparse, identificar y resolver defectos, desgastes y anomalías de equipamiento y del proceso para garantizar los niveles de buena operación y seguridad de la planta, la eficiencia y de este modo hacerla menos vulnerable a la ocurrencia de accidentes. El diagnóstico es la disciplina



de la ingeniería que trata el análisis de las condiciones de operación de los sistemas energéticos con el objetivo de detectar, primero las desviaciones de las condiciones correctas de operación y luego la ocurrencia de posibles anomalías. Esas técnicas están basadas en la medición de variables físicas (temperatura, presión, potencia, etc.) en el punto de operación actual y en condiciones de referencia adecuadas. Cualquier desviación que ocurra entre las variables o sus características paramétricas en dos condiciones de operación puede ser atribuida a la presencia de anomalías.

En el área de control de procesos, la detección adecuada de los intervalos de tiempo en los cuales un proceso se encuentra en estado estacionario (EE), es sumamente importante para su monitorización y optimización. Si el proceso (Kelly y Hedengren, 2013) se encuentra en estado transitorio (ET), puede provocar estimaciones incorrectas de coeficientes paramétricos en el modelado, la toma de decisiones inapropiadas y la aplicación de los resultados de la optimización con dichos modelos puede ocasionar un operación inestable cuando el sistema está operando a lazo cerrado en tiempo real.

Los modelos de EE son muy utilizados en las funciones de evaluación del desempeño y rendimiento del proceso, identificación y ajuste del modelo, para seleccionar los segmentos de datos para modelado, en la optimización y control, detección de fallos, análisis de sensores, reconciliación de datos, análisis de procesos, formación de redes neuronales y para activar intervenciones en línea.

Es conocido que los procesos industriales nunca se encuentran en un estado estacionario (Bagajewicz, 2001). En la práctica, éste puede ser definido como un estado en que el proceso esté operando en la vecindad de algún punto estable o dentro de una región de estacionalidad, en la cual, se asume que la razón de cambio de la variable medida es estadísticamente no significativa o despreciable (Kelly y Hedengren, 2013).

En el campo específico de las centrales termoeléctricas, las condiciones de operación de estado estacionario están normadas por las máximas desviaciones y fluctuaciones de las lecturas de los valores medios estabilizados (Branislav y Radisa, 2011).

Muchos métodos se han utilizado para identificar los EE. Estos incluyen la regresión lineal, estadística F, estadística T, transformada wavelet, entre otros, sobre una ventana de datos, métodos estos que por lo general entran dentro de lo que se conoce como Control Estadístico de Procesos.



Particularmente en el proceso de generación de vapor de las unidades de generación interconectadas al sistema eléctrico, como es el caso de la termoeléctrica “Lidio Ramón Pérez” (Felton), con los valores de EE detectados a variables seleccionadas, se calculan las desviaciones. Los valores de buen desempeño de la operación de las variables son relativos (Can y Mazumder, 2013) porque con la explotación continua de la planta a través de los años se van alejando de las condiciones de diseño de la misma y necesitan ser actualizados luego de cada mantenimiento. La monitorización a diferentes plazos de tiempo consiste en determinar las desviaciones por variables entre cada valor esperado y su valor actual medido. La evolución de las variaciones en las desviaciones es reflejo de las acciones degradantes que afectan el buen desempeño de la operación. Las mismas ocurren en plazos de tiempo largos (Teruel y otros, 2005) por ejemplo, en el orden de horas, días o semanas. Entre los mecanismos físico-químicos que degradan la generación de vapor y los equipos tecnológicos relacionados y por tanto disminuyen la eficiencia (Kurz y otros, 2009) están: la corrosión y el ensuciamiento de superficies, las fugas de vapor de gases y agua, la pérdida de aislamientos en tuberías, el sobrecalentamiento de metales, el desgaste de impelentes, entre otros. La cuantificación del efecto de las diferentes causas es un punto importante para determinar la planificación del mantenimiento lo cual minimizaría el costo total de operación y un ingreso económico mayor que se revertirá en mejoras tecnológicas sucesivas.

Una vez que el inventario de posibles anomalías se define, entonces, el planteamiento científico estaría dado por la solución del problema inverso, a partir de la identificación de los efectos, determinar o inferir sus causas.

La operación de las unidades de generación, específicamente en la Central Termoeléctrica (CTE) de Felton se caracteriza por moverse en amplio rango en potencia, ininterrumpida y con paradas para mantenimiento de diferentes tipos. Hay que predecir los valores de buen desempeño relativo que deben esperarse para cualquier punto de operación. La solución tiene que incluir un procedimiento de modelado con sus medidas después de cada proceso de mejora por mantenimiento.

Sin embargo, la mayoría de los métodos existentes en la actualidad tanto para la detección de estados estacionarios (DEE) como para monitorización continúan



quedando como simples metodologías teóricas y muy pocas veces llegan a sintetizarse en aplicaciones computacionales que puedan ser usadas en la práctica.

En base a lo antes expuesto, para la realización de este trabajo se plantea como:

Problema de la investigación

La ausencia de una herramienta computacional capaz de monitorizar y calcular las desviaciones respecto a los valores deseados en las magnitudes a estado estacionario en las variables de los generadores de vapor de la CTE Felton a partir de las series temporales procedentes de observaciones instrumentales.

Objeto de la investigación

Las series temporales registradas a partir de mediciones instrumentales de variables correspondientes al área de generadores de vapor de la CTE Felton.

Objetivo

Diseñar e implementar una herramienta computacional en MatLab™ capaz de monitorear y detectar desviaciones respecto a los valores deseados en las magnitudes a estado estacionario de las variables en los generadores de vapor de la CTE Felton basándose en el procesamiento de las mediciones instrumentales.

Campo de acción

Algoritmos de cómputo capaces de detectar desviaciones en la operación de plantas energéticas bajo el desarrollo de una interfaz gráfica utilizando el Toolbox GUIDE de MatLab™.

Hipótesis

Si se diseña un conjunto de funciones capaces de detectar (a partir de las series temporales registradas en los generadores de vapor de la CTE Felton) la ocurrencia de desviaciones entre los valores a estado estacionario y los valores esperados de las variables y éstas logran ser integrados en una interfaz gráfica sobre MatLab™ entonces se dispondrá de una herramienta que contribuya a largo plazo al mejoramiento de la eficiencia energética, la seguridad, confiabilidad en esta entidad empresarial.



Para cumplir el objetivo planteado se propone el desarrollo de las siguientes **tareas de investigación**:

1. Caracterizar desde el punto de vista gnoseológico, histórico y en la actualidad, los aspectos referidos al pre-procesamiento, la reconciliación de datos y la detección de estados estacionarios sobre series temporales con mediciones industriales contaminadas con ruido y que portan perturbaciones de lento desarrollo y los algoritmos empleados para estos fines.
2. Diseño de un sistema de cómputo compuesto por bloques funcionales para la inserción de algoritmos de: estimación del nivel de ruido, filtrado, extracción de tendencias y detección de estados estacionarios en registros de proceso que cuente con la posibilidad de salvar de forma cronológica un sistema de reportes con los resultados de la detección.
3. Inclusión de un bloque funcional que implemente una solución al modelado del comportamiento de buen desempeño relativo por variable mediante una curva de buen desempeño relativo construida por interpolación con polinomios de segundo orden como base al cálculo de las desviaciones por variable para cualquier potencia (en el rango de operación previsto) entre los valores esperados bajo estas condiciones y los resultantes de la detección de EE.
4. Incorporación al sistema de cómputo de un bloque funcional que implemente una solución a la agrupación de estas desviaciones por intervalos de fecha y tiempo.
5. Integrar los bloques funcionales del sistema de cómputo mediante una interfaz gráfica de varias pantallas que garantice la flexibilidad en el uso interrelacionado y sincronizado de sus módulos componentes.
6. Comprobación de la efectividad de la solución computacional y los diferentes métodos seleccionados para cada una de las etapas mediante su aplicación al generador de vapor de la Central Termoeléctrica Felton.



Se han empleado los siguientes **métodos de investigación**:

- 1- Método de investigación documental y bibliográfica; para la sistematización del conjunto de conocimientos y teorías relacionadas con el objeto de estudio.
- 2- Método de investigación experimental para describir y caracterizar el proceso estudiado y sus principales regularidades.
- 3- Método inductivo-deductivo para el establecimiento de los parámetros y las expresiones empíricas desarrolladas.
- 4- Método de investigación correlacional y causal para establecer la relación existente entre las variables a través de la validación estadísticas de los resultados y la interpretación de los fundamentos y causa de dicha relación.
- 5- Técnicas computacionales para el procesamiento estadístico de los datos experimentales.

Significación Práctica

Se ha enfrentado una línea de investigación, la monitorización de sistemas de tiempo real con control de planta completa cuya ubicación está, de acuerdo a la jerarquía de la estructura de estos sistemas, por encima de la supervisión. Se ofrece al sector industrial una herramienta computacional abierta a nuevos códigos, que permite un pre-procesado de grandes volúmenes de series temporales procedentes de observaciones instrumentales conservadas en costosas computadoras de un proceso industrial pudiendo servirle de base a empresas de proyectos de automatización industrial para la implementación en tiempo real de acciones correctoras a largo plazo. El conocimiento para su manejo incluye tener presente las características dinámicas del proceso bajo estudio como los tiempos de respuestas preliminarmente conocidos por el experto industrial o el investigador por lo que se facilita la apertura de una nueva línea de investigación multidisciplinaria entre ingenieros químicos, mecánicos, informáticos industriales y automático.



Los resultados de este tipo de investigación introducen en el ámbito del Departamento de Control Automático la capacidad de enfrentar el estudio y caracterización de la solución de automatización existente en plantas complejas y como resultado fundamentarles vías de solución para superar deficiencias derivadas de una mala explotación de la planta ampliando así las perspectivas de cooperación entre la Universidad de Oriente y el potencial de industrias químico-tecnológicas en el Territorio.

La presente investigación se encuentra organizada de la siguiente forma: una introducción general en la que se exponen las principales motivaciones que llevaron a la realización de este trabajo y en la cual se encuentra además, la fundamentación del diseño metodológico del mismo.

Dos capítulos que constan de introducciones parciales para la mejor comprensión de los objetivos de los mismos, estos a su vez, se encuentran organizados por epígrafes, de manera que resulte más fácil su revisión por parte del lector.

Finalmente cada capítulo presenta sus conclusiones, además de las conclusiones generales, recomendaciones, bibliografía y anexos.

En el Capítulo I se presenta el estudio teórico del presente trabajo. Para ello se exponen las generalidades de la obtención de los registros históricos de procesos industriales y la aplicación de varias técnicas de procesamiento de estos registros para la extracción de información útil que pueda ser usada en la mejora del funcionamiento de la planta.

En el Capítulo II se exponen las características específicas de la aplicación desarrollada llamada IME, su interfaz de usuario, principales componentes, facilidades que ofrece y los algoritmos computacionales empleados.



Capítulo 1. Datos y su estructura en procesos industriales automatizados: La detección de estados estacionarios aplicada a la monitorización del desempeño.

Introducción

Las industrias en la actualidad cuentan con sistemas que registran y almacenan diariamente los valores de cada una de las variables que intervienen en el proceso. Estos datos, conocidos como registros históricos o series temporales conforman enormes bases de datos que a pesar de estar presentes no se aprovecha debidamente el potencial de conocimiento existente en ellas. Al procesar estos registros correctamente, se puede mejorar la operación de la planta mediante la identificación de aquellas variables que están influyendo sobre la eficiencia y el comportamiento de la misma y se puede realizar un diagnóstico preventivo de fallos y malos funcionamientos.

En el presente capítulo se realiza primeramente un estudio de los aspectos referidos a los sistemas encargados registrar y almacenar estos importantes registros y como a través de la minería de datos se puede extraer conocimiento útil y comprensible de ellos. Se exponen aspectos sobre la importancia que tiene la Detección de Estado Estacionario (DEE) del proceso, se revisan los conceptos que intervienen en la detección de estos estados así como la evolución histórica de los diferentes algoritmos, procedimientos y métodos dedicados a este fin. Se caracteriza el proceso industrial que sirvió como objeto de la investigación y se demuestra la necesidad de monitorizar las degradaciones de los equipos y componentes que intervienen en el proceso.

1.1 Monitorización y análisis de los datos de procesos industriales

1.1.1 Sistemas y métodos para la historización de los datos

La razón principal que debe estimular a la implementación de una solución de toma de datos históricos es ganar una comprensión más profunda de los datos de forma que sea posible mejorar la eficiencia y ahorrar dinero.

La utilización de herramientas y recursos correctos, junto con la captación continua de datos durante la puesta en marcha y parada, permite realizar análisis que darán una mejor perspectiva de las caídas de producción, permitiendo tomar medidas que hagan



aumentar el beneficio. Varios paquetes de software ejecutan estas tareas, equilibrando los datos históricos o implementando específicamente la historización de datos para análisis de los tiempos de parada.

Los datos históricos también ofrecen beneficios para el diagnóstico y el mantenimiento de equipos tales como bombas y válvulas. Esta información puede permitir seguir la degradación de una pieza a lo largo del tiempo de forma que pueda ocurrir el mantenimiento preventivo cuando sea necesario, pueden prevenir fallos inesperados debido a piezas rotas, desgaste prematuro u otros problemas mecánicos inesperados. Por ejemplo, puede rastrearse el nivel del par de un actuador de válvula a lo largo del tiempo para ver las varianzas de la norma.

En las últimas décadas han emergido compañías y productos especializados en la historización de datos de procesos. Sus productos soportan colecciones de datos para cientos de fuentes de datos como controladores lógicos programables (PLC del inglés Programmable Logic Controller), sistemas de control distribuido (DCS del inglés Distributed Control System), Sistemas de Supervisión y Adquisición de Datos (SCADA del inglés Supervisory Control and Data Acquisition) y Servidores OPC, proporcionan almacenaje eficiente y eficaz de datos y suministran datos a los clientes mediante una rica diversidad de herramientas que incluyen tendencias, displays e informes.

La colección de datos se realiza a través de un programa de interface de datos o de colección de datos para cada fuente de datos específicos. Los nuevos valores junto con el tiempo de colección se envían a un archivo de datos para su historización. Un creciente número de vendedores está haciendo que sus procesos estén disponibles a través de un OPC, así que la interface de colección de datos OPC está siendo el método más popular para obtener datos de procesos.

Resulta conveniente que la información pueda ser soportada y transferida a interfaces de programación más robustas. La mejor solución probablemente incluya soluciones que impliquen la combinación de aplicaciones de varios vendedores, por lo que el acceso abierto a datos históricos será crucial.



Sistemas de Control Distribuido

Un Sistema de Control Distribuido (DCS), es un sistema de control aplicado, por lo general, a un sistema de fabricación, proceso o cualquier tipo de sistema dinámico, en el que los elementos del tratamiento no son centrales en la localización (como el cerebro), sino que se distribuyen a lo largo de todo el sistema con cada componente o sub-sistema controlado por uno o más controladores. Todo el sistema de controladores está conectado mediante redes de comunicación y de monitorización.

Estos sistemas presentan un alto grado de integración entre sus componentes. Hoy en día cada vez son más numerosas las plantas de procesos químicos tales como refinerías de petróleo, termoeléctricas, centrales electronucleares y otras plantas de procesamiento, que son controladas por microcomputadoras basadas en DCS. Generalmente están compuestos por los siguientes componentes:

Controladores, módulos de control multifunción, *PLCs*, módulos de entrada salida, redes de comunicación, módulos de almacenamiento histórico, interface con computadoras, paquetes de software para el monitoreo en tiempo real, sistema de reportes, servidor de alarmas, optimización, control de procesos estadísticos y análisis basado en tendencias.

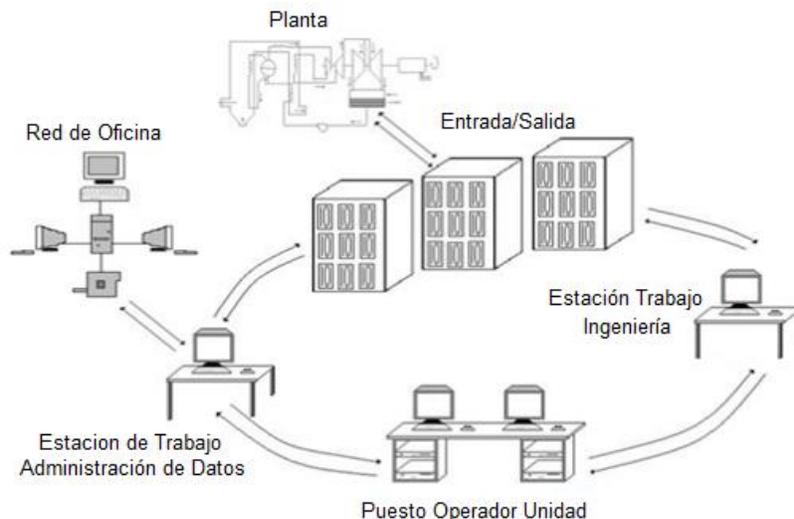


Figura 1.1. Estructura simplificada de los Sistemas de Control Distribuido

La estructura general de los DCS se muestra en la figura 1.1. Esta estructura permite gran flexibilidad en el control de proceso a gran escala. Gracias a ella, la operación en paralelo de varios paquetes de software es posible, tales como herramientas para el



control de procesos estadístico, optimización y sistemas de alerta temprana, los cuales necesitan acceso a las bases de datos del sistema para el procesamiento de la información sin interrumpir la ejecución normal del sistema de control y supervisión.

Servidores OPC

Los servidores OPC están basados en Microsoft Windows COM (del inglés Component Object Model). La arquitectura OPC fue desarrollada por la Fundación OPC y proporciona interface de software normalizada a hardware de bases de datos. Los tres tipos básicos de servidores que poseen son:

1. Servidor OPC de Acceso a Datos.
2. Servidor OPC de Alarma y Eventos.
3. Servidor OPC de datos históricos.

Estos servidores proveen una interfaz Cliente OPC de acceso a datos históricos, que facilita el uso de aplicaciones de acceso a datos. Permiten que las aplicaciones accedan a un banco de datos en tiempo real. Facilitan el desarrollo de aplicaciones sin sacrificar la funcionalidad de la Interfaz Cliente. Entre sus especificaciones cuenta con un juego estándar de objetos, interfaces y métodos que facilitan la interoperabilidad de los datos del sistema.

Una vez integrados estos sistemas los clientes pueden buscar y localizar los datos de interés, conectarse y recuperar datos en tiempo real. Para lograr la integración de toda la planta es necesario el establecimiento de una base de datos en tiempo real que garantice la integración de flujo de datos e información a través de varios sistemas de la planta.

1.2 Análisis de procesos químicos: La quimiometría

Diversos autores han reconocido el potencial de conocimiento inmerso en los datos registrados del proceso para asistir en la operación y mejoras aplicables al proceso, pero también han resaltado la poca capacidad actual para realizar con éxito la explotación de la información de dichos datos (Wang, 2001), (Stockill, 2002). Esto ha provocado la necesidad de desarrollar, adoptar y/o extender técnicas de análisis de datos que ayuden



a manejar estas minas de datos para extraer información útil y con ella soportar la aplicación de diversas tareas operativas.

Para caracterizar, evaluar, diagnosticar y pronosticar los procesos químicos-industriales, por ejemplo, la producción de energía, se requiere disponer de herramientas de cómputo para aplicar las técnicas de quimiometría. Bajo esta denominación se considera la ciencia de las mediciones realizadas sobre sistemas químicos para caracterizar su estado vía la aplicación de métodos matemáticos o estadísticos. Por ende es una ciencia basada en datos. La meta de la mayoría de las técnicas de la quimiometría es derivar un modelo empírico, a partir de los datos, que le permita al investigador estimar una o más propiedades del sistema desde sus mediciones instrumentales.

La quimiometría trata, específicamente, de todos aquellos procesos que transforman señales y datos más o menos complejos en información. Utiliza métodos de origen matemático, estadístico, y otros procedentes del campo de la lógica formal para dirigirse a problemas de química, bioquímica, medicina, biología e ingeniería química. Por todo ello, la quimiometría se sitúa en un campo interdisciplinar. Aunque sus métodos y herramientas provienen de otras disciplinas (como, de hecho, ocurre habitualmente en la química analítica), claramente los fines de la quimiometría están ligados a la química y su éxito depende de los problemas químicos que sea capaz de resolver.

En general, la calidad de los datos experimentales originales se evalúa con el uso de análisis estadístico. Se utilizan muy a menudo para este propósito parámetros tales como; la media, varianza, y la desviación estándar. La valoración puede ser llevada a cabo más allá mediante el empleo de algunos métodos tales como; la prueba T- Student y la prueba F, donde se asume que los datos tienen una distribución Gaussiana.

Con la disponibilidad de un conjunto de datos, las técnicas de procesamiento de señales tales como; métodos de filtrado y suavizado, métodos de transformación, y los métodos de tratamiento numérico podrán ser usados más adelante para analizar y mejorar los datos. Los objetivos son mejorar la relación señal-ruido (SNR del inglés Signal-Noise Ratio), para convertir los datos en una forma más comprensible físicamente, para extraerle la información útil y/o exacta, clasificar objetos, y así sucesivamente.

Para la interpretación de los datos los quimiometras han aplicado herramientas tales como; mínimos cuadrados parciales (PLS del inglés Partial Least Squares) y los métodos



basados en análisis factorial, incluyendo la regresión mediante componentes principales (PCR del inglés Principal-Component Regression), el análisis de factor designado (TFA del inglés Target Factor Analysis), el análisis de factor de evolución (EFA del inglés Evolving Factor Analysis), el análisis de factor de rango de aniquilación (RAFA del inglés Rank Annihilation Factor Analysis) y el análisis de factor de ventana (WFA del inglés Windows Factor Analysis).

Los avances en la computación y la ciencia de la información, las estadísticas, y las matemáticas aplicadas han introducido nuevos elementos dentro de la quimiometría. Las redes neuronales, los algoritmos genéticos y el análisis de Wavelet son ejemplos de estas metodologías.

1.3 La calidad de los datos medidos

En general, los datos de mediciones de procesos son de baja calidad en el sentido de que vienen afectados por errores aleatorios, errores gruesos (*outliers*) de diversos tipos (por ejemplo, valores picos o saltos en el valor de la media de una variable) y valores ausentes o perdidos. Además son altamente susceptibles al ruido, pérdidas, e inconsistencia. Su baja calidad conducirá a bajos resultados en la extracción de información.

Estos errores que por lo general se deben a fallos durante la medición del sensor (u otro componente en el enlace hasta la computadora), pueden reflejarse en forma de outliers (concepto que se establece cuando entre dos muestras el valor anterior y el posterior no se corresponde con la dinámica que representa la variable), no existe valor o aún peor; son incorrectos por lo que se requiere de una reconciliación de la variable bajo análisis respecto a su relación funcional con otras, consideradas como correctas. La comunidad internacional muestra resultados en la adecuación de métodos y algoritmos, para la etapa de pre-procesamiento y reconciliación de datos mediante soluciones de ingeniería computacional al pre-procesamiento.

1.3.1 Errores presentes en la medición

En cualquier planta química o de procesos moderna, cientos de variables como flujos, temperaturas, presiones, niveles, entre otras, son rutinariamente medidas y



automáticamente guardadas con el propósito de controlar el proceso, llevar a cabo una optimización en línea o la evaluación económica. Los valores de las variables medidas de un proceso son inevitablemente adulterados por un sin número de fuentes diferentes como fluctuaciones de suministro de energía, la transmisión de la red y ruido de conversión de la señal, cambios en ambiente, entre otras causas. El error en las variables de un proceso es la diferencia entre el valor medido o calculado y el valor verdadero. El error total en una medición, puede representarse convenientemente como la suma de las contribuciones de dos tipos de errores: los errores aleatorios y los errores gruesos. El error aleatorio se puede definir como el efecto acumulativo de cada una de las incertidumbres que ocasionan y dan lugar a que los datos de una serie de mediciones repetidas fluctúen al azar alrededor del valor medio de los mismos (media estadística) como se muestra en la figura 1.2. Los errores aleatorios son errores pequeños debido a la fluctuación normal del proceso (variación de la presión atmosférica, fluctuaciones de tensión para instrumentos eléctricos) o a la variación aleatoria inherente en la operación del instrumento. El término error aleatorio implica que ni la magnitud ni la señal del error pueden predecirse con certeza y la única manera posible para que estos errores puedan caracterizarse es mediante el uso de distribuciones de probabilidad.

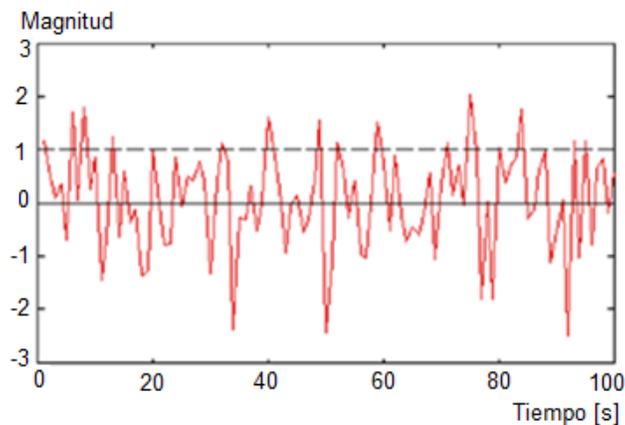


Figura 1.2. Errores típicos en la medición como ruido gaussiano

El ruido Gaussiano es normalmente distribuido con valor medio cero y varianza conocida. La función de densidad probabilística (PDF del inglés Probability Density Function) de la medición se describe por la fórmula:



$$P(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \quad (1.1)$$

Donde μ es el valor medio de las mediciones, y σ es la desviación estándar. La propiedad importante del error aleatorio es que le añade variabilidad a los datos, pero no afecta al rendimiento promedio al grupo.

1.3.1.1 Errores gruesos. Métodos para su detección.

El error grueso (del inglés Gross Error) o error sistemático puede definirse como aquel error que conduce un resultado discordante que difiere marcadamente del valor verdadero. Por lo general, el error grueso puede no aparecer por un corto tiempo y en otros casos se ve gradualmente a lo largo del tiempo. La contribución del error grueso a las medidas hechas con el mismo instrumento bajo condiciones iguales de proceso, siempre es la misma. El error grueso es menos frecuente pero la magnitud es más grande que el del error aleatorio (figura 1.3). Este error es originado por varios factores, entre los más destacados están: fugas en el proceso, corrosión, depósitos de sólidos, entre otros. No importa cuán cuidadosamente se lleva a cabo la medición, permanecerá sin ser detectado el error, incluso si la medición se repite.

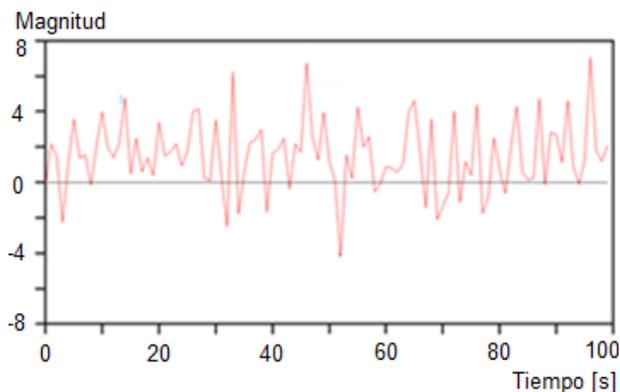


Figura 1.3. Errores gruesos en la medición

A diferencia de los errores aleatorios, los errores gruesos tienden a ser mantenidos ya sean positivos o negativos. Debido a esto, a veces se considera una desviación en la medición. En general, las mediciones con errores gruesos conducirán a información incorrecta sobre el proceso, mucho más que aquellas con errores aleatorios. Los errores en los datos medidos pueden conducir a un deterioro significativo en la operación de la



planta. Dado que el error grueso es más preocupante, se han creado una serie de pruebas basadas en la estadística para su detección.

La presencia de errores gruesos invalida la base estadística de métodos de conciliación de datos. Es también imposible, por ejemplo, preparar un modelo adecuado de proceso con base en mediciones erróneas. Para evitar estos defectos se debe revisar en busca de la presencia de errores sistemáticos en los datos de medición.

Hay diferentes formas para identificar un error grueso:

- i) Análisis teórico de todos los efectos que conducen a un error grueso.
- ii) Con mediciones de una variable dada de proceso por dos métodos con precisión diferente.
- iii) Comprobando la satisfacción de las ecuaciones de balance.

Esta tercera alternativa es en particular atractiva porque es relativamente simple y se basa en relaciones de validez absoluta, a saber, en la conservación de la masa y la energía. De las varias técnicas disponibles, las más usadas son la Prueba Global, la Prueba de Medición (Mah y Tamhane, (1982)), la Prueba de Medición Iterativa Modificada desarrollada por Serth y Heenan (1986), la Prueba Nodal (NT), Cociente de Probabilidad Generalizada presentado por Narasimhan y Mah (1987), la Prueba de Componentes Principales y la Prueba de Potencia Máxima-Mínima. Entre ellos, tres tipos de estrategias se han desarrollado para identificar y corregir múltiples errores gruesos de eliminación en serie, compensación serie, simultáneas o de compensación colectiva. El método Cociente de Probabilidad Generalizada permite identificar errores gruesos múltiples de cualquier tipo usando una estrategia serie de compensación. Para mejorar la eficiencia de la detección de errores gruesos en procesos industriales, una nueva estrategia son los métodos combinados con varias pruebas estadísticas. Estos métodos sólo detectan correctamente en presencia de un error grueso, con excepción de la Prueba Global. Cuando existen múltiples errores gruesos se requiere, además del uso de las pruebas anteriormente mencionadas, de una estrategia para la correcta identificación y localización de todos los errores gruesos. Estas estrategias se clasifican en dos grupos:



- *Estrategias Simultáneas*: Éstas identifican todos los errores gruesos presentes en los datos en una sola iteración. Entre las estrategia simultáneas, la de mayor eficiencia es la conocida como UBET (del inglés Unbiased Estimated Technique) (Bagajewicz (2010), Narasimhan y Jordache (2000)).

- *Estrategias en serie*: Estas estrategias se clasifican a su vez en:

- *Eliminación en serie*: Evalúa cuál de las medidas excede el criterio de Prueba, si existe más de una se toma la del mayor valor, y se elimina de los valores reconciliados, se realiza el mismo procedimiento en cada iteración, hasta que ninguna de las medidas exceda el criterio de Prueba. El método más conocido de esta estrategia es la Prueba de Medición Iterativa propuesta por Narasimhan y Jordache (2000).
- *Compensación en Serie*: Su principio de operación es igual al de la eliminación en serie, la diferencia radica en que en lugar de eliminar las medidas que fallan la prueba, estima su error y lo compensa. El método más potente de esta estrategia es la Identificación Serie con Compensación Colectiva (SICC del inglés Serial Identification with Collective Compensation) (Bagajewicz (2010), Narasimhan y Jordache (2000)).

El método más usado para detectar error grueso es la Prueba de Hipótesis Estadística: se declara un error grueso si el resultado de esta prueba excede un valor crítico el cual se selecciona siguiendo una función de distribución normal estándar. Si el valor de la prueba estadística no excede el valor crítico, entonces la hipótesis nula H_0 es aceptada, y esto significa que la medida no contiene error grueso. Si el valor de la prueba estadística excede el valor crítico, entonces la hipótesis alternativa H_1 es aceptada, lo cual significa que la medida contiene un error grueso.

1.3.1.2 Valores perdidos. Métodos para su tratamiento

En la mayoría de los estudios muestrales y/o censales, principalmente en la medición de unidades, existen múltiples obstáculos, tales como perder una medición, lo que genera



espacios vacíos que producen problemas en el análisis posterior. La presencia de valores perdidos (información ausente o faltante) es un problema común a cualquier investigación y no puede ser ignorado en el análisis de datos. Ignorar los datos ausentes puede tener repercusiones graves que van desde la pérdida de potencia del estudio hasta la aparición de desviaciones inaceptables (Little y Rubin, 1987). La eliminación de tramas con características especiales limita la representatividad o validez externa de los resultados del estudio. El análisis de valores perdidos ayuda a resolver varios problemas ocasionados por los datos incompletos. Además, los datos perdidos pueden reducir la precisión de los estadísticos calculados, porque no se dispone de tanta información como originalmente se pensaba. Otro problema radica en que los supuestos subyacentes a muchos procedimientos estadísticos se basan en casos completos y los valores perdidos pueden complicar la teoría exigida.

Desde hace ya varias décadas, se ha venido estudiando la forma de “llenar” estos espacios vacíos, con el fin de obtener un conjunto de datos completos para analizarse por la vía de los métodos estadísticos tradicionales. Sin embargo, esta situación se complica cuando se presentan en una matriz de datos formada por diversas variables sobre la cual se realizan estudios multivariantes, haciéndose necesario la aplicación de métodos que convenientemente imputen conjuntamente los datos (Geng y Li, 2003). Con la ayuda del avance de la computación, se han desarrollado nuevas formas de estudiar los datos faltantes multivariantes, obteniéndose una variedad de técnicas basadas en diferentes enfoques según las características de la data. Aun así, todavía son muchas las deficiencias que enfrentan las técnicas actuales y que son necesarias resolverlas, como las desviaciones en las estimaciones, alteración de la relación entre las variables, cambios en las varianzas, entre otros.

Se distinguen tres mecanismos de pérdida de datos:

- Datos perdidos completamente al azar (MCAR del inglés Missing Completely at Random). Cuando las características de las tramas con información son las mismas que las de las tramas sin información. Dicho de otra manera, la probabilidad de que una trama presente un valor ausente en una variable no depende ni de otras variables ni de los valores de la propia variable (Gleason y



Staelin, 1975). Las observaciones con datos perdidos son una muestra aleatoria del conjunto de observaciones.

- Datos perdidos al azar (MAR del inglés Missing at Random). Cuando las tramas con datos incompletos son diferentes significativamente de los que presentan datos completos en alguna variable, y el patrón de ausencia de datos puede ser predecible a partir de variables con datos observados en la base de datos del estudio que no muestran ausencia de datos. La probabilidad de que se produzca la ausencia de una observación depende de otras variables pero no de los valores de la variable con el valor ausente. Es imposible probar si la condición MAR es satisfecha y la razón es que dado que no se conoce la información faltante no se puede comparar los valores de aquellas tramas que tienen información con los que no la tienen.
- Datos perdidos no ignorables o no debidos al azar (MNI del inglés Missing Non-Ignorable, o MNAR del inglés Missing Not At Random). Cuando la probabilidad de los datos perdidos sobre una variable Y depende de los valores de dicha variable una vez que se han controlado el resto de las variables.

Métodos de imputación de datos

Cuando no se pueden ignorar los datos faltantes, la manera más adecuada de tratarlos es llenar esos espacios faltantes con valores admisibles; a este procedimiento es lo que denominamos imputación. La importancia de los procedimientos de imputación no radica sólo en reducir la desviación por la ausencia de datos, sino también en producir un conjunto de datos rectangulares y “limpios” sin datos faltantes (Geng y Li, 2003).

Los métodos de imputación consisten en estimar los valores ausentes en base a los valores válidos y/o casos de la muestra. La estimación se puede hacer a partir de la información del conjunto completo de variables o viene de algunas variables especialmente seleccionadas. Usualmente los métodos de imputación se utilizan con variables métricas (de intervalo o de razón) y deben aplicarse con precaución porque pueden introducir relaciones inexistentes entre los datos.

Son muchas las técnicas de imputación que han surgido, sobre todo desde la década de los setenta, que emplean enfoques univariantes y multivariantes. Se han empleado



enfoques basados en modelos como: funciones de verosimilitud, regresión y descomposición de matrices en valores singulares, entre otros. A pesar de estos avances, no se ha encontrado una metodología capaz de reproducir la data o que pueda resolver en forma totalmente satisfactoria el tratamiento de los datos faltantes, debido, generalmente, a problemas en cuanto a las alteraciones de la distribución de los datos, alteración en la relación de las variables, desviaciones en las estimaciones, inflación de la varianza, entre otros, razón por la cual, aún se sigue investigando en busca de mejorar las técnicas existentes.

Los métodos de imputación se pueden resumir en los siguientes casos:

- Sustitución por la media de las observaciones: Consiste en sustituir los valores ausentes por la media de los valores válidos.
- Sustitución por constante: Consiste en sustituir los valores ausentes por constantes cuyo valor viene por razones teóricas o relacionadas con investigación previa.
- Imputación por regresión: Consiste en estimar los valores ausentes en base a su relación con otras variables mediante el Análisis de Regresión.

1.3.1.3 Datos atípicos.

Un dato atípico (del inglés outlier) es un registro mayor o menor de lo esperado que se detecta por tener un residuo que es un valor “inusual”, muy grande o muy pequeño en relación con la distribución asociada a los residuos. También suelen llamarse valores aberrantes o inconsistentes.

Dado que los residuos estandarizados r_{it} son una muestra aleatoria de una distribución $N(0, 1)$, se verifica que aproximadamente un 68% de los r_{it} deben estar entre -1 y 1, y alrededor del 95% entre -2 y 2 y prácticamente todos entre -3 y 3. Por ello, un residuo estandarizado que diste más de 3 o 4 unidades del 0 ($|r_{it}| > 3$) se corresponde, potencialmente, con una observación atípica (Basseville y Nikiforov, 1993).

Para chequear la existencia de observaciones atípicas se pueden utilizar diferentes gráficos: el histograma de los r_{it} , el diagrama de cajas múltiple de los r_{it} por niveles del



factor tratamiento como se observa en la figura 1.4 o el diagrama de dispersión de los puntos.

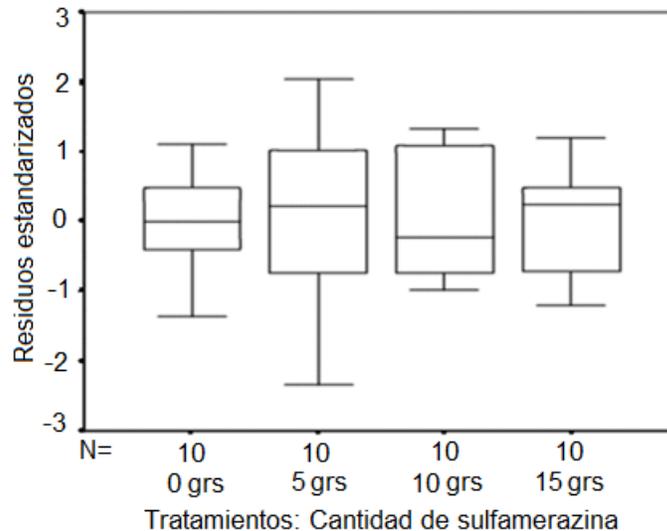


Figura 1.4. Gráfico de cajas múltiple de los residuos estandarizados frente al factor tratamiento.

Para la identificación de valores atípicos, dos de las pruebas utilizadas con mayor frecuencia en un grupo de datos único son la prueba de Nixon y la prueba de Grubbs (Chen y Romagnoli, 1998). La primera utiliza relaciones de los espacios entre datos de diferentes modos según la cantidad de valores en el grupo de datos. El valor crítico depende del tamaño de la muestra n , y de un nivel de representatividad elegido que es el riesgo de rechazar una observación válida.

La prueba de Grubbs utiliza una prueba estadística T , que es la diferencia absoluta entre el valor atípico X_0 y el promedio de la muestra \bar{x} dividida por la desviación estándar de la muestra σ .

Es importante señalar que las pruebas estadísticas se usan para identificar valores atípicos, no para retirarlos del grupo de datos. Técnicamente, una observación no debe retirarse a menos que una investigación halle una causa probable para justificar esta acción. Algunas compañías han definido procedimientos para estas investigaciones incluyendo la repetición del material asociado a la observación de valores atípicos, en caso de ser posible.



Si en la investigación no se encuentra una causa probable para retirar un valor atípico ¿Qué debe hacerse? Un enfoque sería realizar un análisis de datos con el valor atípico y sin él. Si las conclusiones son diferentes, entonces se considera que el valor atípico tiene influencia y esto debería indicarse en el informe (Basseville y Nikiforov, 1993). Otra opción es utilizar estimadores rigurosos para caracterizar los grupos de datos, tal como la mediana de la muestra en lugar de la media. Cuando hay múltiples valores atípicos en un grupo de datos, la investigación resulta más complicada pero existen procedimientos de prueba para estos casos. Un problema es que un valor atípico puede enmascarar otro valor en una prueba de valor atípico único. La prueba de Dixon supera esto redefiniendo los espacios a utilizar a medida que aumenta el tamaño de la muestra

1.3.2. Ruido en las mediciones

Es conocido por los ingenieros que trabajan con datos obtenidos de procesos industriales, que no existen señales libres de ruido. En condiciones ideales el nivel del ruido es insignificante con respecto al nivel de la señal, por lo que la reducción del mismo es innecesaria. En la práctica, sin embargo, es generalmente necesario remover el ruido que altera la señal para cualquier análisis posterior de los datos.

El ruido puede ser definido como una señal no deseada que interfiere con la medición de otra señal. El ruido en sí mismo es una señal que transmite información referente a la fuente que lo genera.

Si la señal del proceso está exenta de ruido entonces la identificación de "cerca del equilibrio" y "definitivamente no es estacionario" sería fácil, y visualmente distinguible. Sin embargo, los datos del proceso suelen ser ruidosos, y el nivel del ruido cambia con el tiempo y las condiciones del proceso. En consecuencia, en el procesamiento de las mediciones adquiridas y transmitidas con diferentes técnicas es necesario considerar el ruido que los contamina y la tasa de muestreo con que se leen. Por último, ya que la magnitud del ruido cambia con las condiciones de funcionamiento, el procedimiento debe ser independiente de la varianza del ruido.

Si un proceso está bien controlado, las variables controladas se mantienen constantes por largos períodos de tiempo. Estos son estados pseudo-estacionarios, períodos en los que la variable manipulada cambia en respuesta a las perturbaciones. En esos períodos,



el ruido puede dominar la dinámica de las variables relevantes del modelo. Cuando se tiene en cuenta el ruido en el proceso de ajuste de los parámetros del modelo, los valores de los parámetros se convierten en estadísticamente inciertos, el modelo se convierte en inválido y funcionalmente inútil o degradado.

Las señales industriales son afectadas por una amplia gama de ruidos. Típicamente, estos tienen naturaleza aditiva o multiplicativa. La figura 1.5.b muestra el efecto del ruido sobre una señal, se puede apreciar que la señal con ruido está distorsionada y es evidente que contiene otras componentes de frecuencias además de la original.

El ruido aditivo sigue la siguiente regla (Couch II, 2002):

$$x(t) = y(t) + v(t) \quad (1.2)$$

Donde: $x(t)$ señal contaminada con ruido

$y(t)$ señal original sin ruido

$v(t)$ ruido presente.

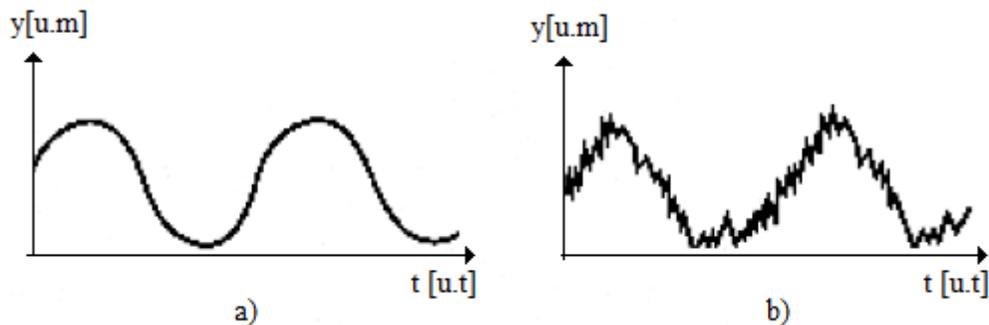


Figura 1.5. Representación de señales: a) Original b) Efecto del ruido.

Para medir la influencia del ruido sobre una señal se utiliza la SNR. Mientras mayor es la relación, menos efecto tiene la señal no deseada sobre la señal útil.

Se define como la proporción entre la potencia de la señal con la potencia del ruido:

$$SNR = \frac{P_S}{P_N} \quad (1.3)$$

Al operar con señales de tiempo discreto adquiridas de sistemas continuos es necesario tener en cuenta el teorema de muestreo de Nyquist-Shannon. La reconstrucción exacta de una señal periódica continua en banda base a partir de sus muestras, es matemáticamente posible si la señal está limitada en banda y la tasa de muestreo es



superior al doble de su ancho de banda.

El ruido gaussiano es una de las formas más comunes en las que éste aparece en las señales industriales, generalmente es causado por procesos aleatorios de la corriente eléctrica o por las agitaciones térmicas de los elementos conductivos. Debe su nombre a que sus amplitudes siguen una distribución de Gauss (Coursey, 2003), independientemente de que exista una correlación del ruido en el tiempo o no.

Dependiendo de su frecuencia o sus características de tiempo, el ruido de proceso puede clasificarse en una de varias categorías como sigue:

- 1- *Ruido de Banda Estrecha*: Un proceso de ruido con una banda estrecha tal como el zumbido de 50/60 Hz del suministro de electricidad.
- 2- *Ruido blanco*: Ruido puramente aleatorio que tiene un espectro de frecuencia plano. El ruido blanco teóricamente contiene todas las frecuencias con igual intensidad.
- 3- *Ruido blanco limitado en banda*: Un ruido con un espectro plano y un ancho de banda limitado que usualmente cubre el espectro del dispositivo o la señal de interés.
- 4- *Ruido coloreado*: Ruido de banda ancha cuyo espectro tiene una forma no plana; Los ejemplos son ruido de rosado, el marrón y el ruido autoregresivo.
- 5- *Ruido Impulsivo*: Consiste en pulsos de duración breve de amplitud y duración aleatorios.
- 6- *Ruido de Pulsos Transitorio*: Consiste en pulsos de relativamente larga duración

Dependiendo la forma concreta de la densidad espectral de potencia \hat{P}_{xx} se definen varios colores para el ruido, haciendo un paralelismo entre las longitudes de onda del espectro visible y las del espectro audible. Por ejemplo un ruido que contenga una distribución uniforme del espectro de frecuencias sería blanco al igual que el color blanco contiene todas las frecuencias del espectro visible, y un sonido con mayor preponderancia de graves estaría coloreado al rosa, ya que los colores con mayor cantidad de frecuencias bajas contienen mayor cantidad de rojo.

A su vez, la autocorrelación es otra de las propiedades en la caracterización de las



señales de ruido. Esta propiedad está relacionada con la forma de distribución del espectro de potencia.

1.3.2.1 Ruido blanco

El ruido blanco (RB) es un caso particular de proceso estocástico estacionario en sentido amplio (WSS, del inglés Wide-Sense Stationarity) en el cual las variables aleatorias que lo forman no están correlacionadas. La señal no está limitada en banda y su potencia es teóricamente infinita. Un ruido que tiene la misma potencia en todas las frecuencias en el rango de $\pm \infty$ debe tener necesariamente potencia infinita, y es por consiguiente sólo un concepto teórico. En la práctica, se considera que una señal es blanca si su correspondiente función de densidad espectral de potencia es constante en la banda de frecuencia de interés en la aplicación.

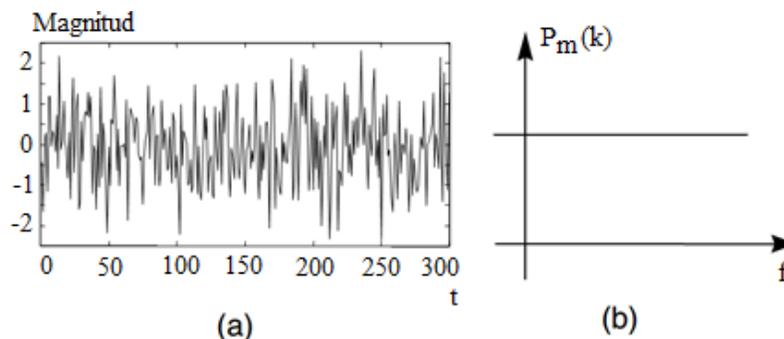


Figura 1.6. Ilustración de a) ruido blanco b) su espectro de potencia.

La función de auto-correlación de un ruido blanco de tiempo continuo con media cero con varianza σ^2 es la función delta dada por:

$$r_{NN}(\tau) = E[N(t)N(t + \tau)] = \sigma^2 \delta(\tau) \tag{1.4}$$

El espectro de potencia del ruido blanco, que se obtiene tomando la Transformada de Fourier de la ecuación 1.4. Está dada por:

$$P_{NN}(f) = \int_{-\infty}^{\infty} r_{NN}(t) e^{-j2\pi ft} dt = \sigma^2 \tag{1.5}$$

Esta ecuación muestra que el ruido tiene un espectro de potencia constante.



1.3.2.2 Ruido coloreado

Aunque el concepto de ruido blanco provee una aproximación razonable, matemáticamente conveniente y útil para una cierta cantidad de ruidos encontrados en sistemas de telecomunicación, en muchos otros procesos el ruido no es blanco. El término ruido coloreado se refiere a cualquier ruido de banda ancha con un espectro de potencia no blanco. Por ejemplo, la mayoría de los ruidos de audio-frecuencia, como el del movimiento de los autos, los ventiladores de computadora, taladros eléctricos y personas hablando en el trasfondo tienen un espectro no blanco predominantemente de baja frecuencia. Dos variedades clásicas de ruido coloreado son llamadas ruido rosado y ruido marrón.

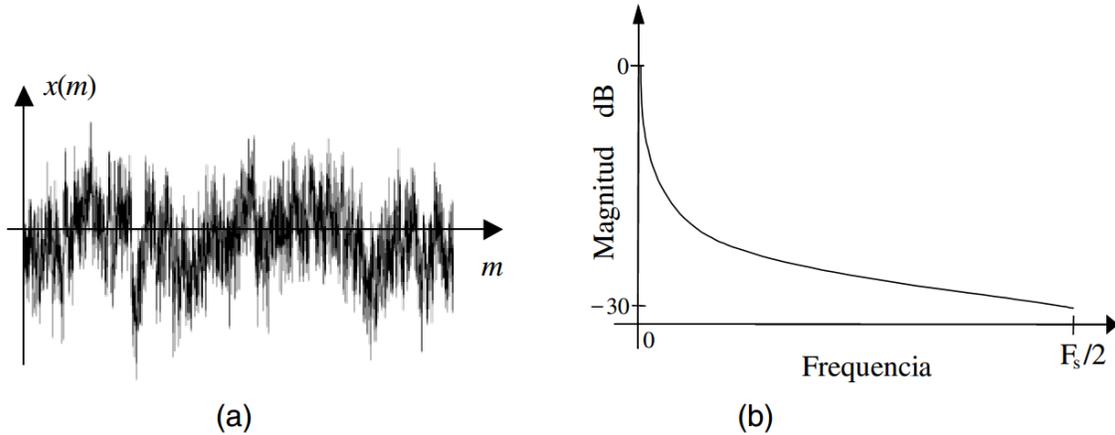


Figura 1.7. a) Señal con ruido rosado b) Su espectro de amplitud

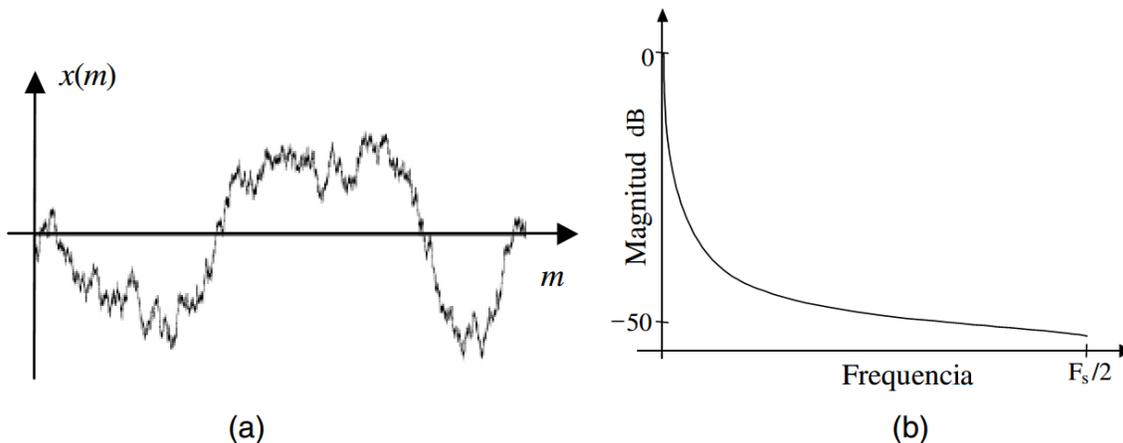


Figura 1.8. a) Señal con ruido marrón b) Su espectro de amplitud



1.3.2.3 Técnicas de reducción de ruido

Tradicionalmente para reducir el ruido se han utilizado diferentes técnicas de filtrado lineal, dentro de las que se destaca la utilización del filtro de Wiener, así como otros esquemas basados en filtros que trabajan directamente sobre la señal en cuestión. Otro conjunto de técnicas para remover el ruido de las señales son las basadas en el uso de transformadas, donde la más conocida y utilizada es la Transformada Discreta de Fourier (DFT del inglés Discrete Fourier Transform), ideal para señales estacionarias cuyos valores varían muy poco en el tiempo, pero no igual de efectiva para el análisis de señales con variaciones significativas en el tiempo (Rioul and Vetterli, 1991). Tanto las técnicas de filtrado lineal como las basadas en FT presentan importantes dificultades para el tratamiento de señales no estacionarias (Dolabdjian et al., 2002).

En este contexto aparece la transformada Wavelet (DWT del inglés Discrete Wavelet Transform).

1.3.2.4 Transformada Wavelet

El concepto clave de esta transformada es la adaptación del análisis en el tiempo y la frecuencia de forma simultánea. En este ámbito, ha emergido una amplia teoría acerca de la utilización de técnicas no lineales para reducir el ruido y se han diseñado diferentes esquemas para la supresión del ruido mediante la DWT y la aplicación de umbrales. La reducción de ruido basada en wavelet proporciona una base adecuada para separar el ruido de la señal.

Si se asume que los datos observados $x(t) = s(t) + \eta(t)$ contienen la señal pura $s(t)$ con un ruido auditivo $\eta(t)$ como funciones muestreadas en el tiempo y mediante $W(\cdot)$ y $W^{-1}(\cdot)$ se denotan la transformada Wavelet directa e inversa respectivamente, entonces la técnica de reducción de ruido a que se hace referencia involucra los siguientes pasos:

1. Obtención de los coeficientes wavelet de la señal ruidosa

$$y = W(x) = \zeta + v \quad \zeta = W(s); \quad v = W(\eta) \quad (1.6)$$

2. Modificación de los coeficientes obtenidos a partir de la aplicación de umbrales.



$$Z = D(y, \lambda) \tag{1.7}$$

donde $D(., \lambda)$ es un operador que para un valor de u actúa de una de las siguientes maneras:

$$D_s(u, \lambda) = \begin{cases} 0 & \text{si } |u| \leq \lambda \\ u - \lambda & \text{si } u > \lambda \\ u + \lambda & \text{si } u < -\lambda \end{cases} \tag{1.8}$$

$$D_s(u, \lambda) = \begin{cases} 0 & \text{si } |u| \leq \lambda \\ u & \text{si } |u| > \lambda \end{cases} \tag{1.9}$$

La fórmula (1.8) corresponde a la llamada umbralización suave o flexible mientras que la fórmula (1.9) se corresponde con la llamada umbralización fuerte o fija.

3. Reconstrucción de la señal $s(t)$ mediante la transformada Inversa $\hat{s}(t) = W^{-1}(Z)$

El objetivo de los métodos de supresión de ruido es minimizar la relación $\|s(t) - \hat{s}(t)\|$ en la reconstrucción, o sea suprimir el ruido manteniendo, en lo posible, las propiedades de la señal original. El valor de λ que regula cómo deben cambiar los coeficientes wavelet, se denomina umbral. La reducción de ruido mediante la aplicación de umbrales a los coeficientes wavelet de la señal comenzó con los trabajos de (Donoho et al., 1994). Aquí se introdujeron los métodos RiskShrink que es un umbral de tipo minimax y VisuShrink con un umbral universal y el método SureShrink basado en un umbral que minimiza el riesgo de Stein. Posterior a estos trabajos aparecieron algunos métodos de selección del umbral que son adaptativos con respecto a las características espaciales de la señal. Los métodos adaptativos han resultado ser más efectivos en las experiencias realizadas sobre todo en imágenes. Entre los métodos adaptativos un lugar especial ocupan los que están basados en las reglas de Bayes (Bayes Shrink, Modified Bayes Shrink, Normal Shrink y Eight Shrink).

Ha sido demostrado, desde el punto de vista teórico, que estas técnicas de umbralización wavelet son casi óptimas cuando se tiene poca información acerca de las características de las señales a analizar. Sin embargo, en muchos problemas prácticos se tiene algún conocimiento “a priori” acerca de la clase de señales que son de interés.



La optimalidad de un algoritmo para la reducción de ruido en una clase específica de señales no conlleva necesariamente a que este algoritmo sea óptimo para otro tipo de señales.

Dentro de los usos de esta poderosa herramienta podemos nombrar, además del análisis local de señales no estacionarias, el análisis de señales electrocardiográficas, sísmicas, de sonido, de radar, así como también es utilizada para la compresión y procesamiento de imágenes y reconocimiento de patrones.

1.4 Técnicas de procesamiento de los datos

Las técnicas de pre-procesamiento se agrupan en tres direcciones fundamentales; (i) limpieza para eliminarle el ruido y la inconsistencia, (ii) la integración para lograr un almacenamiento coherente ya que sus fuentes de procedencia son diversas, y (iii) la transformación de datos.

Uno de los primeros pasos de cualquier análisis, por tanto, debe tender a la llamada "rectificación de los datos", que consiste en estimar los valores verdaderos (libres de errores) de las variables del proceso y^* , a partir de las mediciones del proceso y tales que:

$$y = y^* + \varepsilon \quad (1.10)$$

Donde ε representa los errores que se añaden a la medición. Es una tarea clave, ya que si se logran minimizar los errores, los datos resultantes conducirán a mejores resultados en las aplicaciones que utilicen estos datos.

Aun cuando la rectificación representa un paso intermedio en el pre-procesamiento de los datos, en la industria química es un requisito de diversas tareas (control, monitorización, optimización de procesos, mantenimiento, etc.) por lo que en sí mismo representa un problema de gestión de información a resolver. Los desarrollos en esta área han ido surgiendo a lo largo de las cuatro últimas décadas y pueden dividirse en dos grandes bloques:

- Métodos de rectificación basados en modelos.
- Métodos de rectificación basados en filtros univariable.



Una vez que existe una solución para la visualización y se dispone del dato pre-procesado, corresponde aplicar las técnicas de procesamiento. El conocimiento puede representarse en forma de relaciones, patrones o reglas inferidos de los datos (previamente) desconocidos. De las técnicas para la identificación de patrones se tienen las modelizaciones. Éstas pueden ser paramétricas y no paramétricas. Si los valores de los atributos (variables) son series temporales, se investiga sobre sistemas estáticos si estos son cuasi-estacionarios en el tiempo y sobre sistemas dinámicos si muestran tendencias no constantes. Dentro de las no paramétricas se enmarcan aquellas que no se basan en una estructura predefinida y un conjunto de parámetros para ajustar.

En la práctica pueden ser de dos tipos: predictivos y descriptivos. Los predictivos pretenden estimar valores futuros o desconocidos de variables de interés, que denominamos variables objetivos, usando otras variables o campos de la base de datos, a las que nos referiremos como variables predictivas. Los modelos descriptivos, en cambio, identifican patrones que explican los datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos. En el manejo de bases de datos existen tareas que se desean implementar y métodos de cómputo para ejecutarlas. De acuerdo a como sea el problema, bien predictivo o descriptivo, así serán los métodos para darle solución.

Las tareas requieren métodos, técnicas o algoritmos para resolverlas. Una tarea puede ser resuelta con varios métodos como también un método puede solucionar varias tareas. Entre las tareas predictivas se tienen: la clasificación, estimación de probabilidad de clasificación, categorización, preferencias o priorización y regresión.

Entre las tareas descriptivas se tienen: agrupaciones (clustering), correlaciones, factorizaciones, reglas de asociación, dependencias funcionales y detección de valores e instancias anómalas. Las técnicas basadas en transformaciones al dominio frecuencia como la DFT, la Transformada Rápida de Fourier (FFT del inglés Fast Fourier Transform), y la DWT así como el análisis de correlación en el dominio tiempo constituyen opciones de cómputo muy útiles.



1.4.1 Reconciliación de Datos

Las inconsistencias de los datos se caracterizan mediante distribuciones de probabilidad. Las primeras se corrigen tempranamente mediante etapas de filtrado para maximizar la gestión de resultados durante su procesamiento y la segunda, atenuando los errores aleatorios. La técnica de reconciliación de datos es para mejorar la precisión, consistencia y confiabilidad de los datos y tiene el objetivo de reducir el impacto de los llamados errores aleatorios. Se fundamenta en un ajuste óptimo de los datos medidos tales que estos valores ajustados deben satisfacer condiciones basadas en leyes de conservación (por ejemplo, de masa, de energía, de elementos químicos, etc.) u otras restricciones del proceso. Esta vía de formalización, también conocida como reconciliación de errores de las mediciones o de materiales y energía fue propuesta alrededor de 45 años atrás y permite la estimación de variables no medidas.

La reconciliación de datos (DR) es la estimación de las variables del proceso sobre la base de la información contenida en las mediciones y modelos de procesos. Los modelos de procesos utilizados en la conciliación de datos son por lo general las ecuaciones de conservación de masa y energía.

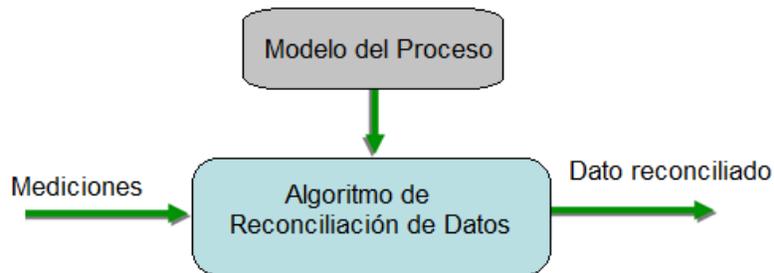


Figura 1.9. Representación de un Sistema de Reconciliación de datos

La técnica DR permite el ajuste de las mediciones de modo que las mediciones corregidas sean consistentes con los balances correspondientes. Esta información de los datos conciliados puede ser utilizada para diferentes propósitos tales como:

- Monitoreo
- Simulación
- Optimización
- Mantenimiento de instrumentos



- Modelado
- Control
- Administración
- Análisis de equipamiento

El desarrollo de la investigación durante los últimos 30 años ha dado lugar a dos grandes tipos de aplicaciones:

-Reconciliación del equilibrio de masa y energía. El ejemplo más simple es la reconciliación fuera de línea de las tasas de flujo alrededor de las unidades de proceso. Las velocidades de flujo reconciliados deben satisfacer el balance de masa global de las unidades.

-Estimación de parámetros del modelo. Se requieren estimaciones exactas y precisas de los parámetros del modelo con el fin de obtener modelos predictivos fiables para la simulación de procesos, diseño y optimización. Un enfoque para la estimación de parámetros es resolver el problema de estimación simultáneamente con el problema de conciliación de datos. Se espera que los parámetros del modelo conciliados sean más precisos y se puedan utilizar con mayor confianza.

Los modelos empleados en DR representan relaciones variables del sistema físico del proceso. Los datos reconciliados toman la información tanto de las mediciones y los modelos. En la conciliación de mediciones de estado estacionario, las restricciones de un modelo son las ecuaciones algebraicas. Por otro lado, cuando se trata de procesos dinámicos, se deben utilizar los modelos dinámicos que son ecuaciones diferenciales.

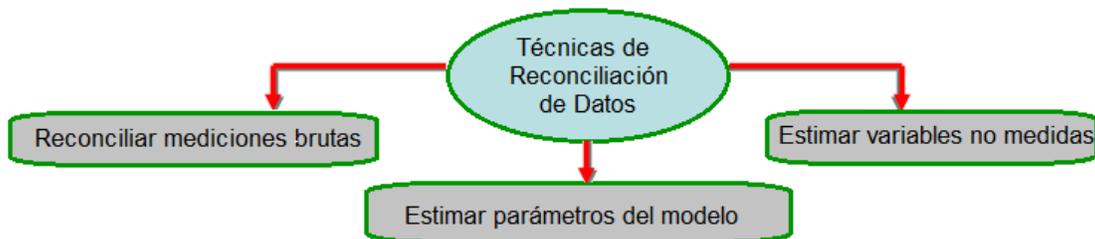


Figura 1.10. Aplicaciones de un Sistema DR

El enfoque integrado para tratamiento o conciliación de datos requiere un set de métodos matemáticos aplicados en la instrumentación de proceso y las mediciones. Esto provee



estimación completa y coherente de las variables de proceso. La metodología general puede estar dividida en tres pasos principales y ellos son:

- Clasificación de variables de proceso y la descomposición problemática
- Detección, identificación y estimación de errores gruesos
- El ajuste de medición y la estimación de las variables no medidas de proceso.

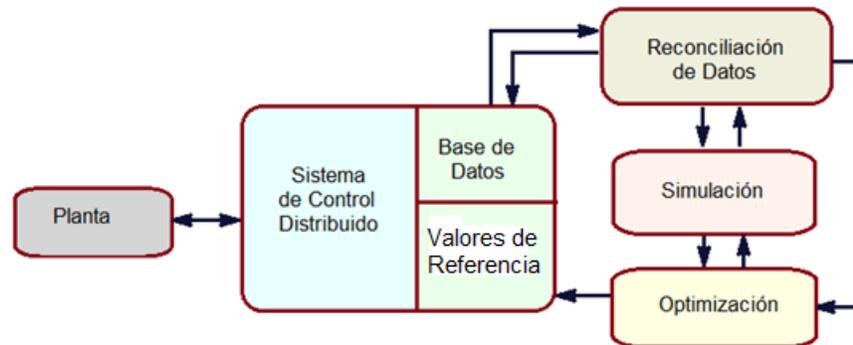


Figura 1.11. Arreglo típico entre un Sistema de Control Distribuido y la Reconciliación de Datos, la Simulación y la Optimización.

La idea original de reducir los sistemas de ecuaciones usados en el problema de conciliación se basa en un método de corrección basado sólo en un subconjunto reducido de ecuaciones y mediciones. La idea básicamente consiste en explotar la topología de proceso para clasificar las variables de proceso y eliminar del problema de las no medidas, dar como resultado un subconjunto de ecuaciones que involucre sólo las variables medibles. Varias estrategias han sido desarrolladas desde ese entonces para lograr la misma meta, o sea la descomposición de proceso para reducir la dimensionalidad del problema.

El ajuste de mediciones para compensar errores aleatorios involucra la solución del problema de minimización restringida, usualmente uno de mínimos cuadrados. Las ecuaciones de balance son incluidas en las restricciones; estos pueden ser lineales pero son generalmente no lineales. La función objetivo es usualmente cuadrática con relación al ajuste de mediciones y tiene la matriz de covarianzas de errores como pesos. Así, esta matriz es esencial en la obtención de conocimiento fiable de proceso. Algunos esfuerzos se han hecho para estimarla de las mediciones (Almasy y Mah (1984), Dorouach y otros



(1989), Keller y otros (1992)). La dificultad en la estimación de esta matriz está asociada con el análisis de la correlación serial y cruzada de los datos.

La estimación de parámetros es también una actividad importante en el diseño, evaluación y control de procesos. Ya que los datos tomados de los procesos químicos no satisfacen las restricciones del proceso, los métodos "error en variable" proveen estimaciones de parámetros y reconciliaciones de datos que son consistentes respecto al modelo. Estos problemas representan una clase especial de problema de optimización porque la estructura de mínimos cuadrados puede ser explotada en el desarrollo de métodos de optimización. Una revisión de este tema puede encontrarse en el trabajo de Biegler et al. (1986).

1.4.2 Regresión no paramétrica

La teoría clásica de la regresión se basa, en gran parte, en el supuesto que las observaciones son independientes y se encuentran idéntica y normalmente distribuidas. Si bien existen muchos fenómenos que pueden modelarse de esta manera, para el tratamiento de ciertos problemas, la normalidad de los datos es insostenible. En el intento de eliminar esa restricción se diseñaron métodos que hacen un número mínimo de supuestos sobre los modelos que describen las observaciones.

La teoría de los métodos no paramétricos trata, esencialmente, el desarrollo de procedimientos de inferencia estadística, que no realizan una suposición explícita con respecto a la forma funcional de la distribución de probabilidad de las observaciones de la muestra. Si bien en la Estadística no paramétrica también aparecen modelos y parámetros, ellos están definidos de una manera más general que en su contrapartida paramétrica.

La regresión no paramétrica es una colección de técnicas para el ajuste de funciones de regresión cuando existe poco conocimiento a priori acerca de su forma. Proporciona funciones suavizadas de la relación y el procedimiento se denomina suavizado. Los fundamentos de los métodos de suavizado son antiguos pero sólo lograron el estado actual de desarrollo gracias a los avances de la computación y los estudios por simulación que han permitido evaluar sus comportamientos. La técnica más simple de suavizado, los promedios móviles fue la primera en usarse (Härdle y otros, 2004), sin



embargo han surgido nuevas técnicas como la estimación mediante núcleos o la regresión local ponderada. Estos estimadores de regresión no paramétrica son herramientas poderosas para el análisis de datos, tanto como una técnica de estimación para resumir una relación compleja que no puede ser aprehendida por un modelo paramétrico, como para suplementar (o complementar) un análisis de regresión paramétrico.

En los análisis paramétricos se comienza haciendo supuestos rígidos sobre la estructura básica de los datos, luego se estiman de la forma más eficiente posible los parámetros que definen la estructura y por último se comprueba si los supuestos iniciales se cumplen. La regresión no paramétrica, en cambio, desarrolla un “modelo libre” para predecir la respuesta sobre el rango de valores de los datos. Básicamente está constituida por métodos que proporcionan una estimación suavizada de la relación para un conjunto de valores (denominado ventana) de la variable explicativa. Estos valores son ponderados de modo que, por ejemplo, los vecinos más cercanos tengan mayor peso que los más alejados dentro de una ventana de datos. Se pueden utilizar diversas funciones de ponderación, que son los pesos en que se basan los estimadores. La combinación de la función de ponderación y el ancho de la ventana inciden sobre la bondad de la estimación resultante. Dentro de los métodos de regresión no paramétrica más importantes se destacan.

- *Medias móviles*: calcular la media de y en una ventana alrededor de cada valor de x .
- *Medias móviles ponderadas* (suavización kernel): ponderar los datos en función de la distancia.
- *Regresión polinomial local*: ajustar un polinomio por mínimos cuadrados a los puntos en una ventana local, ponderados por la distancia.

1.4.3 Interpolación Polinómica

La interpolación es la estimación de muestras de una señal usando una medida ponderada de un número de muestras conocidas en los puntos de la vecindad. Los interpoladores son usados en formas diversas en la mayoría de los sistemas de procesamiento de señales y toma de decisiones. El objetivo de la interpolación es obtener una reconstrucción de alta fidelidad de lo desconocido o las muestras perdidas



de una señal. El enfoque clásico para la interpolación es construir un polinomio interpolador que atraviesa las muestras conocidas.

La interpolación polinómica es un método usado para conocer, de un modo aproximado, los valores que toma cierta función de la cual sólo se conoce su imagen en un número finito de abscisas. A menudo, ni siquiera se conocerá la expresión de la función y sólo se dispondrá de los valores que toma para dichas abscisas. Es decir, dado cierto número de puntos obtenidos por muestreo o a partir de un experimento, el objetivo será hallar un polinomio que pase por todos los puntos y que permita hallar aproximaciones de otros valores desconocidos para la función con una precisión deseable fijada (Stein, 1999). Por ello, para cada polinomio interpolador se dispondrá de una fórmula del error de interpolación que permitirá ajustar la precisión del polinomio.

Definición

Dada una función f de la cual se conocen sus valores en un número finito de abscisas $x_0, x_1, \dots, \dots, x_m$, se llama interpolación polinómica al proceso de hallar un polinomio $P_m(x)$ de grado menor o igual a m , cumpliendo $P_m(x_k) = f(x_k) \quad \forall k = 0, 1, \dots, m$. A este polinomio se le llama Polinomio interpolador de grado m de la función f .

Ajuste del polinomio interpolador

Se dispone de varios métodos generales de interpolación polinómica que permiten aproximar una función por un polinomio de grado m . El primero de estos polinomios es el método de las diferencias divididas de Newton. Otro de los métodos es la interpolación de Lagrange, y por último, la interpolación de Hermite. Estas formas diversas son matemáticamente equivalentes y pueden ser transformadas de una en otra.

1.4.3.1 Interpolación de Hermite

En determinadas aplicaciones se precisan métodos de interpolación que trabajen con datos prescritos y sus derivadas en una serie de puntos, con el objetivo de incrementar la aproximación en las proximidades de dichos puntos (Spitzbart, 1960). Dentro de esa clase de métodos esta la interpolación de Hermite.



La interpolación de Hermite, llamada así en honor a su inventor Charles Hermite, es similar a la de Newton pero con el añadido de que ahora también conocemos los valores que toma la derivada de la función f en las abscisas conocidas $x_0, x_1, \dots, \dots, \dots, x_m$.

El Polinomio Interpolador de Hermite de grado $2m + 1$ de la función f es un polinomio de la forma:

$$P_{2m+1}(x) = \sum_{i=0}^m f_i \Phi_i(x) + \sum_{i=0}^m f'_i \Psi_i(x) \quad (1.11)$$

donde:

$$\Phi_i(x) = \left(1 - 2l'_i(x_i)(x - x_i)\right) l_i^2(x) \quad (1.12)$$

$$\Psi_i(x) = (x - x_i) l_i^2(x) \quad i = 0, \dots, \dots, m \quad (1.13)$$

$l_i(x)$ son los llamados polinomios de Lagrange, que se calculan de este modo:

$$l_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j} = \frac{(x - x_0)(x - x_1) \dots (x - x_{j-1})(x - x_{j+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{j-1})(x_i - x_{j+1}) \dots (x_i - x_n)} \quad (1.14)$$

La interpolación de Hermite puede extenderse al conocimiento de las derivadas sucesivas de la función a interpolar en las abscisas tomadas, de modo que se puede obtener un polinomio cada vez más ajustado a la función real, ya que éste podrá cumplir otros requisitos como una determinada monotonía, concavidad, entre otros. En este caso, estaremos hablando de interpolación de Hermite generalizada y su cálculo se llevará a cabo de forma similar a la apuntada, pero obteniendo polinomios de grado cada vez mayor debido a las sucesivas derivadas de los coeficientes $l_i(x)$.

Para la interpolación de Hermite, al igual que en la de Lagrange, se tiene una fórmula del error de interpolación que, naturalmente, tiene en cuenta factores relacionados con las derivadas de f . Más concretamente, se dispone de una fórmula del error en el caso en que la función f sea $2m + 2$ veces diferenciable en un intervalo I mediante la siguiente expresión:



$$f(x) - P_{2m+1}(x) = \frac{f^{2m+2}(\xi(x))}{(2m+2)!} (x - x_0)^2 (x - x_1)^2 \dots (x - x_m)^2 \quad (1.15)$$

Para $x \in I$ y donde $\xi(x) \in \langle x_0, x_1, \dots, x_m \rangle$

La diferencia esencial entre la Interpolación de Hermite y la Interpolación de Lagrange reside en el cálculo a través de la construcción de los Polinomios de Lagrange. En este caso, su cálculo es arduo, largo y complicado; por lo que el uso de las llamadas diferencias divididas generalizadas simplifica mucho el cálculo del polinomio interpolador.

1.5 Los estados estacionarios en la operación de los procesos industriales

Las plantas de proceso continuo deben someterse a una serie de cambios en los estados de funcionamiento. Las plantas también se encuentran bajo una serie de cambios de un modo operativo a otro. Estos cambios se denominan transiciones. Los arranques y paradas son ejemplos comunes de las transiciones del proceso. Otras transiciones se producen debido a cambios en la calidad de las materias primas. Las transiciones a menudo resultan en productos fuera de las especificaciones. Por eso es importante acelerar el ajuste de la planta al nuevo estado de equilibrio.

La mayoría de las plantas de proceso se diseñan para operar en condiciones de estado estacionario (EE). En la práctica estas condiciones no se reúnen debido a que ellas están sometidas a cambios impredecibles. Sin embargo la suposición de estado estacionario es empleada de modo satisfactoriamente útil para el análisis de los datos recolectados (Bagajewicz, 2001).

La detección tanto de estado estacionario como de estado transitorio en señales ruidosas de proceso es muy importante. Los modelos de estado estacionario son ampliamente empleados en control de procesos, en la monitorización en línea, en su optimización, en las funciones de evaluación del desempeño y rendimiento del proceso, identificación y ajuste del modelo, para seleccionar los segmentos de datos para modelado, en la optimización y control, detección de fallos, análisis de sensores, reconciliación de datos, análisis de procesos, formación de redes neuronales y para activar las intervenciones en línea.



1.5.1 Métodos y técnicas estadísticas en la detección de estado estacionario

Se considera que los datos medidos en un proceso representan el verdadero valor de la medición con el ruido adicional y perturbaciones, y la condición de estado estacionario significa que el verdadero valor del proceso permanece sin cambios. Se debe tener en cuenta que aquí el concepto de estado estacionario es menos estricto que "estrictamente estacionario" en estadística. En esta rama, "estacionario" requiere no sólo que la media de la serie de datos temporales permanezca constante, sino también que la distribución y la autocorrelación, no cambien con el tiempo.

Abordar una investigación orientada a la DEE requiere de los métodos de la estadística matemática con su base en la teoría de las probabilidades (Ott, 2001) y se divide en dos partes: la descriptiva y la inferencial.

Las vías directas inmediatas en la implementación de un método para la DEE (Rhinehart, 2013) son: el empleo de ventana deslizante sobre L puntos, sea bien determinando la pendiente en su ajuste lineal regresivo o bien mediante la determinación sucesiva de medias. En ausencia de ruido, la pendiente sería idealmente cero y las medias exactamente iguales. Sin embargo la presencia del ruido produce fluctuaciones.

Los enfoques existentes en la identificación de estado estacionario van encaminados a realizar una regresión lineal en una ventana de datos y a continuación, realizar una prueba estadística tipo T sobre la pendiente de la regresión. Si la pendiente es significativamente diferente de cero, el proceso es casi seguro que no se encuentre en estado estacionario. Ésta suele ser una técnica fuera de línea. Las versiones en línea requieren un considerable almacenamiento de datos, el esfuerzo de cómputo asociado y la experiencia del usuario. Además, en medio de una oscilación definitiva, donde la pendiente de la regresión lineal fue temporalmente cero, el método daría una lectura falsa. En cada intervalo de tiempo, la ventana de datos debe ser actualizada y la regresión lineal debe ser recalculada. Si la ventana de datos es larga, habrá un considerable esfuerzo computacional, y el reconocimiento de los cambios se retrasaría. Si la ventana es corta, el ruido puede confundir el análisis. No hay reglas universales para elegir la longitud de la ventana de datos y la selección debe ser a juicio crítico.



Un método alternativo utiliza una prueba estadística de tipo F que no es más que una relación de las variancias que fueron medidas en el mismo set de datos por dos métodos diferentes. Los datos de la más reciente ventana son un promedio y la variancia es primero calculada convenientemente como la desviación media cuadrada de la media. La variancia también puede ser calculada a partir de la media de las diferencias cuadradas de los datos sucesivos. Si la serie temporal es estacionaria entonces, idealmente, la relación de las varianzas será la unidad. Sin embargo, debido al ruido aleatorio, la relación real de las varianzas no será exactamente la unidad, solo se acercará a ella. Alternativamente, si el proceso no está en estado estacionario la relación será inusualmente grande (con respecto a la distribución de la relación de valores en los cuales se esperan condiciones de estado estacionario).

Existe un método basado en la forma de onda en el cual las tendencias del proceso son extraídas de los datos medidos a través del procesamiento multi-escala basado en la forma de onda. El estado del proceso se mide utilizando un índice con un valor entre 0 y 1 según el módulo de la transformada de forma de onda de la señal extraída del proceso. Por último, un estado estacionario es identificado si el índice calculado es pequeño (cerca de cero). Este método tiene una mejor precisión para detectar cambios en el proceso debido a la buena localización de propiedad de la transformada de forma de onda, y es más conveniente para las aplicaciones en línea.

Cao & Rhinehart (1995) proponen otro método eficiente para la identificación en línea del estado estacionario del proceso. En este método, un filtro convencional de primer orden se utiliza para sustituir la media de la muestra. Con esta modificación, no se requiere períodos de tiempo (ventanas de datos), por tanto, los requisitos de almacenamiento y de cálculo se reducen de manera significativa.

Chen & Liao (2002) presentaron un método de pruebas en estado estacionario basado en técnicas estadísticas. El método asume que sólo los errores al azar existen con media cero y los períodos de tiempo sucesivos definidos de cada una de ellas se asumen constantes. En la primera etapa, una prueba se aplica para determinar si las matrices de covarianza de dos períodos consecutivos son iguales, entonces las medias de los dos períodos se analizarán usando la prueba T^2 de Hotelling.



Existe también un enfoque basado en la Transformada Wavelet para la identificación y localización de tendencias polinomiales en mediciones de ruido, para lo cual el estado estacionario es un caso especial para el polinomio de grado cero. Después de la aplicación de la transformada wavelet y la eliminación del ruido de las mediciones, el error de aproximación de mínimos cuadrados se calcula para cada sub-intervalo del coeficiente de aproximación wavelet. El estado estacionario se identifica mediante la fusión de sub-intervalos consecutivos constantes con aceptables errores comunes.

Los enfoques existentes (excepto el método de Cao y Rhinehart (1995)) se ejecutan en el período de tiempo especificado con evaluaciones globales, en lugar de probar el proceso en un punto arbitrario de tiempo, que es crítico para aplicaciones on-line y especialmente para aplicaciones en tiempo real.

Otro método reportado por la práctica es calcular la desviación estándar de las mediciones del proceso a través de una ventana móvil del historial de los últimos datos. Cada vez que el proceso no está en el estado estacionario, la desviación estándar medida es más grande que su valor de estado estacionario. Por lo tanto, cuando la desviación estándar medida es mayor que el valor de umbral, se establece la condición de estado no estacionario. Los autores señalan que el éxito con este método se basa en la capacidad de determinar correctamente las unidades de las variables clave, el período de tiempo usado para el cálculo y la desviación estándar (umbral). Además, cuando cambia la varianza del proceso, el valor umbral tiene que cambiar. Una vez más, el almacenamiento y la operación en la ventana de datos es una carga computacional.

En los métodos basados en ventanas deslizantes se requiere de la experticia del usuario para la elección del número de muestras a considerar. En estos casos se incluyen: (Alekman, 1994), (Jubien, y Bihary, 1994), (Kim et al., 2008) y (Kelly y Hedengren, 2013). Un análisis crítico de cada uno se presenta a continuación.

El método introducido por (Alekman, 1994) compara la media de un registro reciente con uno de referencia que está basado en el registro anterior y le aplica una prueba T de Student para determinar si la media cambió. El problema de esta técnica es que la condición de estado estacionario no es generalmente equivalente a la media.

El método de (Jubien, y Bihary, 1994) está basado en el cálculo de la desviación estándar sobre mediciones recientes dentro de una ventana deslizante. La desviación estándar



calculada tiene que estar dentro del límite que se establece por una condición declarada como de umbral. El éxito de este método depende de la habilidad que se logre para determinar el tiempo del período que debe emplearse para cada variable de proceso.

La solución (Kelly y Hedengren, 2013) es del tipo basada en ventana deslizante y utiliza la prueba de T de student para determinar si la diferencia entre los valores de la señal del proceso menos su media está por encima o por debajo un número de veces de la desviación estándar en su valor estadístico crítico.

Otros métodos tratan los datos secuencialmente sin necesidad de seleccionar un valor de ventana preliminar. En (Cao y Rhinehart, 1995) presentan una modificación al estadístico de la prueba de Fisher. Esto se logra incorporando un filtro de media móvil exponencialmente ponderada. Estos métodos tienen el inconveniente de que requieren que no exista autocorrelación en los datos de cada variable ni tampoco relación de correlación cruzada entre las variables. Esto último significa que en condición de estado estacionario, el ruido de una variable no debe estar correlacionado con el ruido de la otra (Brown y Rhinehart, 2000) pero salvando este inconveniente señalado es uno de los citados como efectivo y práctico.

Los trabajos previos en el área de detección de estado estacionario son resumidos en los siguientes casos:

a. Técnica basada en la regresión lineal con el objetivo de determinar el mejor ajuste con tendencia lineal para L valores anteriores de la variable. Heurísticamente, si el proceso está en EE, entonces la pendiente de la línea de tendencia será idénticamente igual a cero. Sin embargo debido a la presencia de ruido en el proceso, la pendiente puede fluctuar con valores cercanos a cero y consecuentemente, un valor de la pendiente desigual de cero, no es razón para rechazar la hipótesis de EE. Para aceptar o rechazar la hipótesis podría usarse una prueba basada en la estadística T de Student aplicado al cociente entre el valor de la pendiente y el error estándar de ese valor. Si el cociente mencionado excede el valor crítico puede considerarse que hay suficiente evidencia para rechazar la hipótesis de que el proceso está en EE.

b. Técnica basada en la evaluación de los valores promedios en ventanas sucesivas de datos. Se calculan estadísticos tales como; el promedio y la desviación estándar de los



datos en sucesivos conjuntos de muestras. Si el proceso está en estado estacionario los promedios deben ser iguales en cada ventana. Las diferencias entre dos promedios consecutivos provocadas por la presencia de ruido pueden ser analizadas con una prueba de Student. Los trabajos (Kelly y Hedengren, 2013) y (Kim y otros, 2008) son representantes de esta técnica.

c. Técnica basada en el análisis de la varianza, ver los trabajos de (Rhinehart, 2013), (Neha y otros, 2010). Se calculan las desviaciones estándar en sucesivos conjuntos de muestras y se aplica la prueba de Fisher para aceptar o rechazar la hipótesis de estado estacionario.

Los métodos agrupados en los casos *a* y *b* caen en la vertiente de soluciones en las que se procesan los datos en ventanas deslizantes. Las técnicas agrupadas en *c* tratan secuencialmente a los datos sin la necesidad de tales ventanas. En los métodos basados en ventanas deslizantes se requiere de la experiencia del usuario para la elección del número de muestras a considerar.

1.6 Monitorización del desempeño con fines de diagnóstico.

Monitorizar la operación es observar las medidas actuales a una o varias variables, y respecto a sus magnitudes esperadas de buen desempeño relativo determinar sus desviaciones respectivas. El propósito final de esta etapa, es crear las bases para detectar las causas que modifican la correcta operación en equipos e intercambiadores de calor y que en algunos casos pudiera conducir a sus fallos.

Diagnosticar los problemas que ocurren durante la operación de un sistema energético (Vázquez y otros, 2010), consiste en descubrir e interpretar los signos de un mal funcionamiento en los equipos y cuantificar sus efectos en términos de consumo adicional de recursos; es decir, saber dónde, cómo y qué parte del consumo de recursos puede ser ahorrado, manteniendo constantes la cantidad y especificaciones de los productos del sistema, así como las condiciones que afectan el comportamiento del sistema y que no pueden ser manipuladas por el operador.

La implantación de sistemas para la monitorización empleando los datos de estado estacionario de la planta puede tener dos criterios (Seborg y otros, 2004), el rendimiento global y la degradación del equipamiento.



La monitorización del rendimiento es para la evaluación continua a lo largo del tiempo de la capacidad de producción y eficiencia de una Unidad de Generación Eléctrica (UGE). Las evaluaciones son repetidas a intervalos regulares. Involucra dos cálculos: la producción actual y la esperada.

La determinación de los valores medios a los segmentos en estado estacionario aplicada a la monitorización a largo plazo tiene implícito un conocimiento interdisciplinario porque requiere del conocimiento de disciplinas tales como la termotecnia, la teoría de los sistemas de control incluyendo el modelado de sistemas y señales, la estadística, el procesamiento digital de señales entre otras.

La selección de variables se hace sobre la base de un criterio que permita evaluar los efectos de acciones degradantes. Éste puede estar orientado a observar los efectos de la desviación de las variables sobre el rendimiento actual de un subproceso o bien sobre un nivel global de la unidad de generación eléctrica. Para esto es preciso conocer, de la termotecnia, el rol de cada una de ellas en el ciclo térmico.

Las secuencias de medidas adquiridas de cada una de ellas son ininterrumpidas y además están contaminadas con ruido blanco y coloreado. Para detectar los segmentos de muestras libres de transiciones hay que emplear métodos del procesamiento de series temporales y de procesamiento digital de señales. Al emplear ventanas deslizantes para explorarlas hay que tener un conocimiento preliminar acerca de modelado de sistemas y señales para caracterizar las magnitudes de los coeficientes paramétricos a la dinámica de los subprocesos que las producen. Estos valores se necesitan para configurar los anchos de dicha ventanas deslizantes.

La observación de la forma de las series de tiempo adquiridas a la unidad de generación eléctrica como parte de un sistema de suministro eléctrico coloca la operación en niveles discretos de potencia de un amplio rango de operación de las variables. Para seleccionar las variables, además de lo señalado anteriormente, hay que distinguir entre cuales permiten reconocer el punto de operación de la UGE y cuáles no. Para encontrar una respuesta, es imprescindible conocer de los principios de la conversión electromecánica de energía y aquellos que rigen el ciclo térmico. Del primero su reacción sobre el sistema basado en el ciclo que le produce el flujo de vapor sobrecalentado. Del segundo que para producir ese flujo variable con la potencia, hay un conjunto de etapas cuyas variables



deben ser reguladas para que sus márgenes de variabilidad sean mínimos y lograr máxima eficiencia del ciclo térmico.

Por todas esas razones se precisa de una metodología que establezca armónicamente las etapas que se requieren para diseñar sistemas de monitorización de las desviaciones de magnitudes para cualquier valor de potencia.

1.6.1 Necesidad de la monitorización de los efectos de las degradaciones en Centrales Termoeléctricas.

La monitorización de las desviaciones es la vía de observar los efectos de la degradación de equipos e intercambiadores de calor pues éstas son reflejo de la reducción de la capacidad del rendimiento en el equipo que ha ocurrido en el tiempo. El grado de degradación, es un parámetro relativo pues se sustenta en el cálculo de las desviaciones a variables seleccionadas respecto a una condición relativa de referencia. Por ejemplo, el rendimiento actual de un calentador de agua de alimentación se mide por su Temperatura Terminal Diferencial (TTD). La diferencia entre la TTD medida y la TTD esperada es una evaluación del grado en la degradación en el calentador de agua de alimentación.

Las degradaciones (Toffolo y Lazzaretto, 2007) son debidas a modificaciones externas al subproceso (tales como variables ambientales, cambios de la calidad del combustible), inducidas sobre el subproceso (por ejemplo debidas al incorrecto funcionamiento del sistema de control en mantener las relaciones y calidad de la combustión) e intrínsecas del subproceso y provocadas por desgastes de partes, variación del espesor de paredes por ensuciamiento, etc. Los mecanismos esenciales que causan la degradación (Kurz y otros, 2009) son los siguientes: el ensuciamiento, la corrosión en caliente, la corrosión, la erosión, y la abrasión.

El ensuciamiento se debe a la adherencia de partículas a los planos aerodinámicos y superficies anulares.

La corrosión en caliente es la pérdida o el deterioro de material de los componentes expuestos a los gases de combustión calientes, causado por las reacciones químicas entre los componentes y ciertos contaminantes.



La corrosión es causada tanto por los contaminantes en la entrada de aire como en la entrada de combustible y los derivados de la combustión.

La erosión es el levantamiento abrasivo de material que está en el paso de la trayectoria de partículas duras e incompresibles que portan los flujos al chocar con la superficie.

La abrasión es causada cuando una superficie rotativa roza en una superficie estacionaria. El daño también puede deberse a los objetos extraños que golpean los componentes del camino de flujo.

Mientras algunos de estos efectos pueden ser eliminados o atenuados mediante limpieza o lavado, otros requieren del ajuste, la reparación, o reemplazo de componentes.

La monitorización a largo plazo de los procesos químicos y de unidades de generación eléctrica, así como el manejo de situaciones anormales tiene justificadas razones económicas. A escala internacional esta afirmación se ejemplifica con los siguientes casos:

- La deposición de partículas de cenizas sobre las superficies de transferencia de calor provoca pérdidas cuantiosas (Peña y otros, 2006) pues afectan la disponibilidad y la eficiencia en plantas termoeléctricas. Estas razones promueven investigaciones para la monitorización inteligente de procesos (Horowitz y otros, 2010), de la llama de los quemadores con la aplicación de técnicas espectroscópicas (Ballester y García, 2010), combinación de técnicas de análisis de respuestas transitorias (Finn y otros, 2010) y aplicaciones de software para asistir la evaluación a corto plazo de la operación (Branislav y Radisa, 2011), entre otras. Por ejemplo, el impacto económico en Estados Unidos ha sido alrededor de los 1.2 billones de dólares en la industria petroquímica.
- Además de los efectos de la corrosión y el ensuciamiento de superficies de intercambio (Seborg y otros, 2004) se deteriora la instrumentación, se producen perturbaciones inusuales por cambiar las normas del combustible y las características constructivas.
- En (Blanco y Peña, 2008) señalan que la corrosión y el ensuciamiento son las causas de alrededor del 90 % de los fallos en sobrecalentadores, recalentadores y economizadores. En la instrumentación las causas se localizan en el rozamiento o trabas en la válvula de control y la inexactitud de los sensores y como perturbaciones la mala atomización y los cambios de propiedades físico química de los carburantes.



En Cuba se han desarrollado múltiples y variadas investigaciones para la observancia de estas problemáticas. Por ejemplo se han abordado situaciones tales como las siguientes:

- El efecto en el diseño de la planta de los cambios en el combustible por otros de alto contenido de azufre (Cuza, 2011).
- El efecto de la corrosión y del ensuciamiento de superficies en sobrecalentadores en la zona de alta temperatura y en los Calentadores de Aire Regenerativos (CAR) así como de la composición y temperatura de los gases de salida sobre la eficiencia del ciclo y el medio ambiente (Prieto, 2008), (Fajardo, 2010) y (Correoso, 2012).
- La estimación de índices de sobre-consumo por desviaciones de los valores de consigna de variables seleccionadas, (Palmero, 2011).
- El efecto del deterioro del condensador de vacío con agua de mar sobre la eficiencia del ciclo.

En tanto no se introduzcan métodos y sistemas de cómputo, los datos que usualmente se emplean, resultan de la selección visual de tramas de secuencia de valores consideradas en estado estacionario.

A las problemáticas analizadas anteriormente se le adiciona las exigencias que le imponen la competencia y las rigurosas normas medio-ambientales de seguridad a la automatización de gran escala de la industria de proceso.

1.6.2 Caso particular del generador de vapor de la unidad # 2 de la CTE Felton

El objeto de estudio de esta investigación ha sido la unidad # 2 de la Central Termoeléctrica Felton. Esta fábrica cuenta con dos unidades de generación eléctrica de 250 MW cada una. Fue construida por la empresa checa SESTLMACE (SlovenskeEnergetickeStrojarnje) y puesta en operación para el sistema eléctrico cubano en Noviembre del 2001.

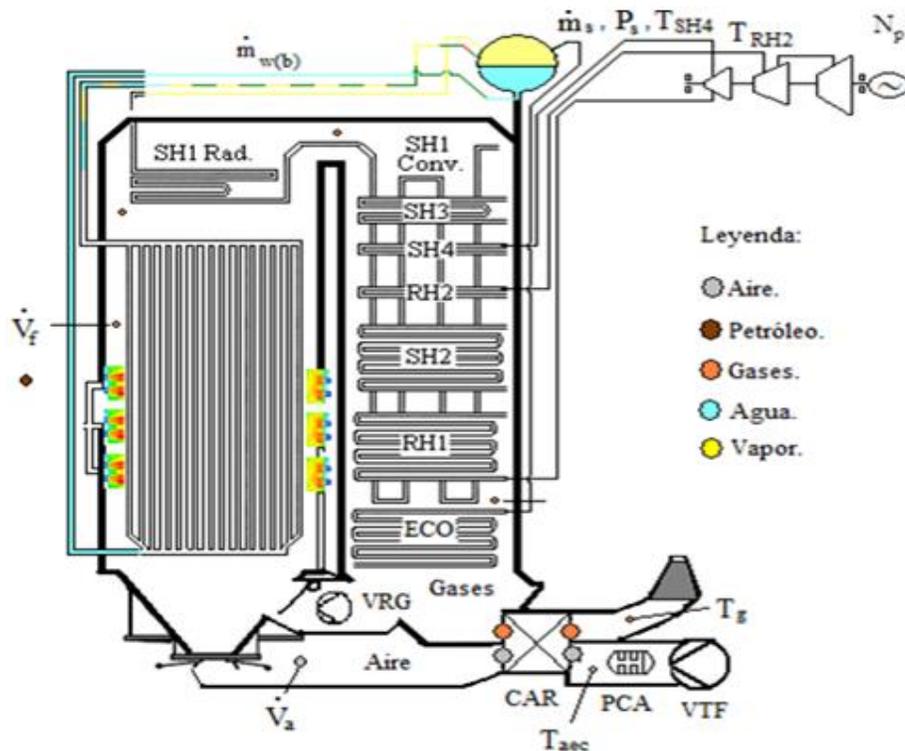


Figura 1.12. Disposición de los subprocesos del Generador de Vapor de la unidad 2 de Felton.

De las degradaciones de equipos y procesos, el ensuciamiento de superficies de intercambio es el de mayor incidencia sobre los factores de eficiencia y disponibilidad de unidad (por las razones que se exponen mas adelante) y por ende es necesario e importante monitorizar y analizar diariamente las variables que reflejan ensuciamiento en las superficies exteriores de transferencia de calor. La degradación por ensuciamiento en sobrecalentadores disminuye la temperatura del vapor vivo a la salida de caldera, en recalentadores la temperatura del vapor recalentado y la temperatura del agua de alimentar en el economizador. Para cada condición de carga, las acciones del sistema de control deben garantizar que todas esas variables alcancen sus valores normativos de presión y temperatura del ciclo térmico, y simultáneamente, se satisfaga el seguimiento en la entrega de potencia activa demandada. La degradación de intercambiadores por su ensuciamiento se refleja en una pérdida de energía y por ende debe incrementar el consumo de combustible lo cual conduce a una degradación del rendimiento al nivel de UGE.



El aumento de la suciedad en el CAR disminuye la transferencia de calor al aire afectando su temperatura al llegar a quemadores para la combustión. Afecta la transferencia en las superficies del eje convectivo porque el coeficiente de convección es sensiblemente inferior en una superficie metálica cubierta de cenizas, que en una superficie metálica limpia. La energía no transferida desde los gases se pierde en las altas temperaturas de los gases a la salida de la chimenea y afecta al medio ambiente. Un exceso de suciedad en los CAR, provoca una alta presión del hogar de caldera (horno), que en principio para mantener las condiciones de diseño de la caldera, obligaría a reducir carga de la central termoeléctrica y puede desencadenar combustiones secundarias. Éstas se facilitan ante altas temperaturas, turbulencia y tiempo de las partículas combustibles. Sus posibilidades se incrementan en los eventos de parada y arranque.

El ensuciamiento tiene su origen en el nivel de efectividad con que se logra la combustión. En la práctica industrial es bastante difícil de lograr una combustión completa. Existen muchos factores tales como: el buen desempeño del trabajo mecánico de los quemadores, la buena mezcla del aire y del petróleo con sus adecuadas temperaturas y con un mínimo exceso de aire. Las partículas de carbono sin quemarse, denominada hollín, arrastradas por las corrientes de gases impulsadas por los VTF y los VTI, se acumulan en todas las superficies de la caldera. Las incrustaciones producidas tienen el efecto de reducir el coeficiente total de transferencia de calor, usualmente denotado por U . Los sistemas de limpieza, sean por soplado de vapor y aire o sean por métodos ultrasónicos reducen las incrustaciones y por ende aumenta U . De todos modos el resultado de su acción sobre este coeficiente no supera el incremento que se consigue en las limpiezas del mantenimiento ya que siempre existe un efecto remanente acumulativo de soplado en soplado.



Sintetizando lo analizado en este capítulo se puede inferir que el diseño de métodos, algoritmos y procedimientos que localicen oportunamente el deterioro específico de componentes y equipos que intervienen en el proceso, por ejemplo, ensuciamiento de superficies, corrosión de tuberías, desgastes de partes o el incremento en el rozamiento de los cojinetes de motores, entre otros, es una etapa más compleja que requiere de resolver bien algunas etapas preliminares sin las cuales esto no sería posible. La primera fase es el pre-procesamiento de las series temporales con el objetivo de mejorar la calidad de los datos y evitar tener información incorrecta sobre el proceso. Luego se requieren procedimientos para la discriminación de la señal en segmentos de EE ya que considerar muestras que no se encuentren en este estado conducirá a errores en los resultados. Posteriormente es necesario conocer y establecer cuáles son las condiciones de buen desempeño de la planta para ser tomadas como referencia a la hora de realizar la comparación con las mediciones procedentes de la instrumentación del proceso y determinar cuánto se han desviado las variables de las condiciones deseadas. Esta primera parte de monitorización de desviaciones de variables es la base del diagnóstico operacional, de ahí su gran importancia.



Conclusiones del capítulo

1. Se describieron los sistemas empleados a nivel mundial para la obtención y almacenamiento de los registros de variables del proceso y los métodos para la extracción de información útil inmersa en estos registros.
2. Se muestra la importancia de considerar y eliminar los errores y el ruido en las mediciones a la hora de procesar los datos y las diferentes técnicas existentes para lograr este objetivo llegando a la consideración de que el procesamiento mediante Transformada Wavelet es el más adecuado para su implementación en este trabajo por su buen comportamiento ante señales correlacionadas.
3. Se expone la importancia de la identificación de los estados estacionarios en procesos industriales y los diferentes métodos que existen en la actualidad para lograr este fin. El método elegido se basa en el cálculo de la desviación estándar en una ventana deslizante de datos que presenta buen desempeño en señales procedentes de plantas térmicas.
4. Se realiza un estudio de los diferentes métodos de regresión no paramétrica y de ellos se elige la Interpolación de Hermite para la construcción de polinomios interpoladores por preservar la forma y monotonía de los datos.
5. Se fundamenta por qué la monitorización del efecto de las degradaciones y ensuciamientos en el equipamiento del proceso es fundamental para identificar las causas que afectan la correcta operación de la planta y trabajar en función de la planificación del mantenimiento en aras de minimizar el costo total de operación y lograr un mayor ingreso económico.



Capítulo 2. El IME: una plataforma flexible para la experimentación industrial

Introducción

La Industria Termo-energética representa un sector importante en el desarrollo de nuestra economía. Debido a esto, se ha hecho un considerable esfuerzo para mejorar el diseño y la operación de las plantas tal que los procesos operen en forma segura y eficiente, se asegure la mejor calidad de la producción, se mantenga la viabilidad económica y se reduzca el impacto ambiental de sus actividades. De ahí la necesidad de diseñar herramientas que permitan determinar las principales causas que afectan el buen funcionamiento de tales plantas. En este capítulo se fundamentan los aspectos necesarios para diseñar un sistema que calcule la secuencia de desviaciones de variables seleccionadas a partir de las series de tiempo que portan las mediciones que pueden estar contaminadas con ruido blanco y coloreado. La aplicación aquí desarrollada denominada IME está encaminada a lograr este objetivo. A continuación se exponen sus generalidades, su interfaz de usuario, principales componentes, facilidades que ofrece y los algoritmos computacionales empleados.

2.1. Generalidades del Explorador Industrial Multivariable

El IME es un software construido usando las facilidades del MatLab™ versión 7.7 del 2008. El mismo es un laboratorio virtual para procesamiento de las series temporales extraídas de generadores de vapor con fines de monitorización. El software utiliza algunas herramientas que han sido programadas a partir de la necesidad de resolver determinados aspectos indispensables para cada una de las variables del proceso bajo estudio. El diseño del IME posibilita la descomposición del procesamiento en dos partes o módulos fundamentales. El dato de salida del primero es entrada al segundo. Esta concepción flexibiliza la introducción de mejoras.

Ambos módulos cuentan con una interface de usuario amigable que facilita la interacción con la gran cantidad de funcionalidades con que cuenta el sistema. Para ello se usó el Toolbox GUIDE (del inglés Graphical User Interface Development Environment) del



MatLab™, es un ambiente de desarrollo integrado para la confección de interfaces gráficas de usuario que permite crear de modo interactivo esta interface y ejecutar programas que necesiten ingreso continuo de datos. Tiene las características básicas de todos los programas visuales como Visual Basic o Visual C++, aunque todavía con unas posibilidades más limitadas. Desde el punto de vista de su aspecto GUIDE se parece a una moderna herramienta para el diseño de interfaces de usuario rápidas, que no son más que programas que facilitan la producción y puesta a punto de software en relativo poco tiempo. Hay que decir que antiguamente MatLab™ no proporcionaba esta herramienta y la tarea de realizar un programa en dicho lenguaje con ventanas y botones era bastante tediosa, pero con GUIDE esto deja de ser así, por tanto, si bien no es uno de los avances más importantes de MatLab™, si ha sido uno de los más retribuidos por los usuarios.

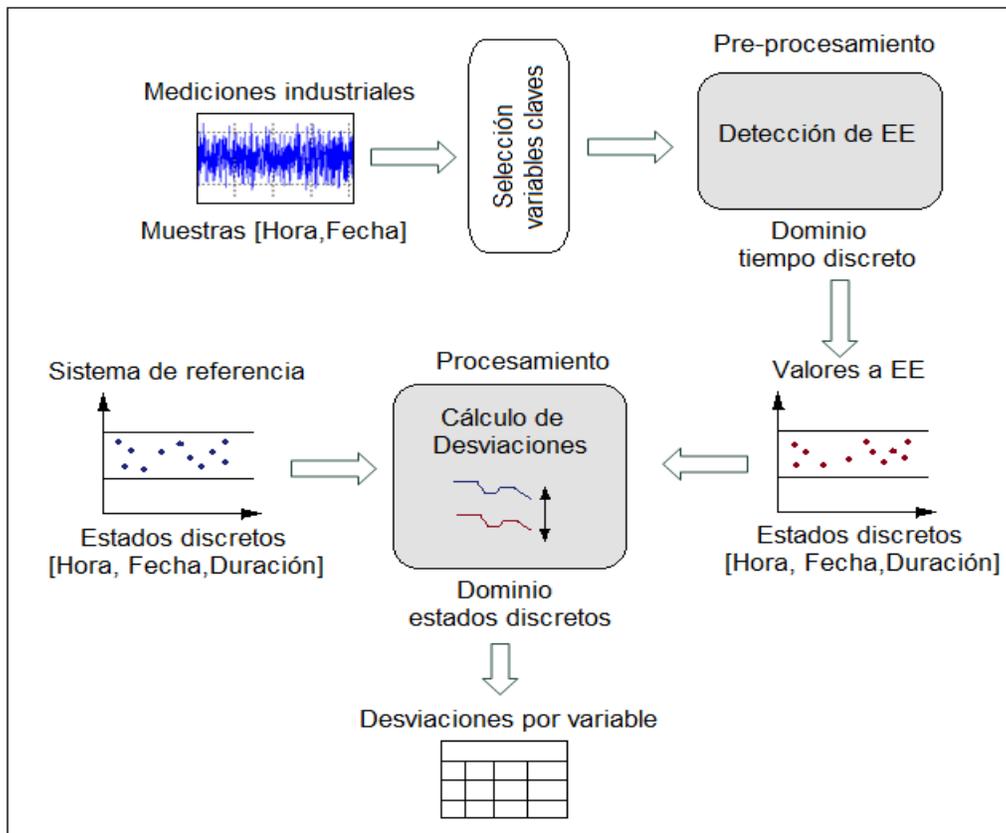


Figura 2.1. Estructura general del IME



Es importante señalar que los métodos y algoritmos explicados en este trabajo son solo los seleccionados para su implementación en cada una de las etapas del procesamiento pero no son los únicos que pueden ser incorporados. El IME es una herramienta escalable y flexible a la introducción de nuevos métodos en dependencia de las ventajas o desventajas de cada uno de ellos en cuanto a cantidad de parámetros en entrada requeridos, carga computacional, exactitud en la estimación, etc. El IME está concebido de manera que se mantiene la estructura de entrada-salida de cada bloque funcional independientemente del método utilizado por lo que se garantiza la escalabilidad y operatividad del sistema.

2.2. Análisis del módulo de DEE

El primer módulo se nombró SSDetector (del inglés Steady State Detector). Aquí es donde están programadas las funciones de Estimación de Ruido en las mediciones, Detección de Estados Estacionarios, Detector de Coincidencias, Filtro por Umbral de Tiempo Mínimo y otras.

Primeramente el usuario debe cargar el juego de datos a procesar procedente de mediciones instrumentales del proceso con su acuñado de fecha y hora. Este dato de entrada es un fichero MatLab™ (con extensión .mat) que ha sido previamente convertido desde un fichero Excel y se carga en forma de una estructura que contiene las mediciones de todas las variables. Una vez cargado, se muestran en un panel en la parte izquierda todas las variables que contiene la estructura. Teniendo en cuenta que el procedimiento se debe realizar una variable a la vez por las características intrínsecas de cada una de ellas, en este panel se selecciona la que se va a procesar y a partir de aquí se cuentan con numerosas posibilidades de operaciones sobre dicha variable.

Dentro de las facilidades podemos citar el graficado dinámico, es decir, la posibilidad de poder graficar los resultados de cada etapa o bloque de manera muy particular en dependencia de la estructura de salida de los datos de cada uno de ellos, de esta manera el usuario puede ir valorando y decidiendo si los cálculos que va realizando son correctos y si cumplen con los requisitos esperados y en caso de no ser así, rectificar en cada paso sin tener que llegar al final de todas las operaciones.

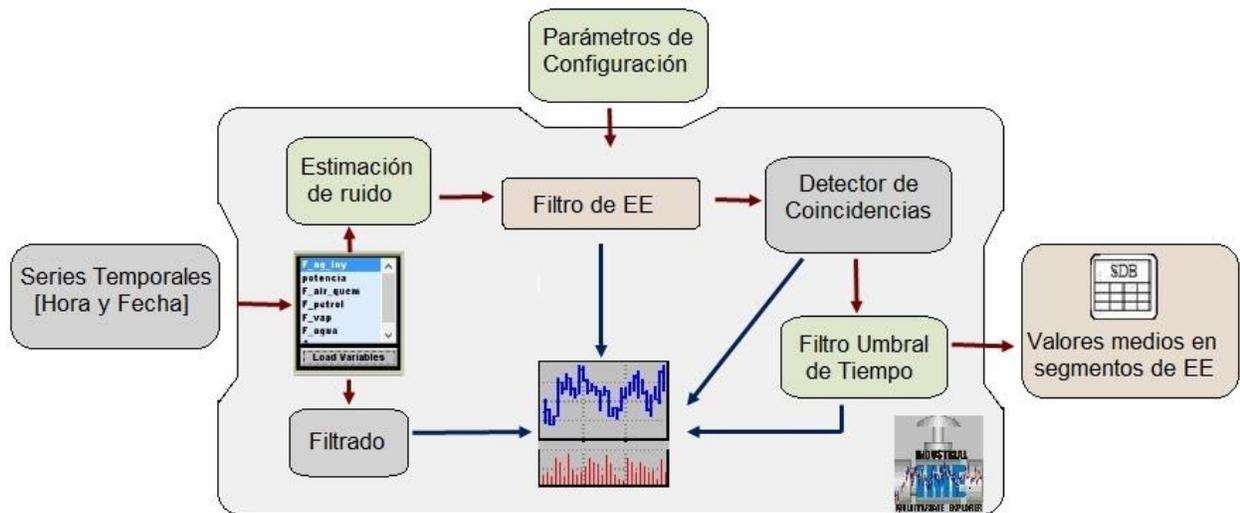


Figura 2.2. Prestaciones del módulo de DEE

También se debe destacar la posibilidad de manejar diferentes estructuras de datos tanto de entrada como salida de cada uno de los módulos como son: ficheros Excel y archivos con extensión (.mat) los cuales son los que usa internamente el Matlab para almacenamiento de datos.

2.2.1. Estimación del ruido.

En el procesamiento de las mediciones adquiridas y transmitidas con diferentes técnicas es necesario considerar el ruido que los contamina y la tasa de muestreo con que se leen.

Teniendo en cuenta el carácter abierto de la aplicación, se pueden introducir técnicas de filtrado lineal, como el filtro de Wiener que trabaja directamente sobre la señal en cuestión o técnicas basadas en el uso de transformadas, como la DFT, ideal para señales estacionarias cuyos valores varían muy poco en el tiempo, pero no igual de efectiva para el análisis de señales con variaciones significativas en el tiempo, aspecto este en el cual la DWT es más robusta, es por eso que fue el método seleccionado para su implementación en el IME.



2.2.1.1 Estimación de ruido mediante Transformada Wavelet

En muchos casos la varianza σ^2 del ruido que contamina las mediciones no es conocida de antemano y es necesario su estimación. Una clase de estimadores hace uso de las características de la DWT.

Las Wavelets, funciones bases de la Transformada Wavelet, son generadas a partir de una función Wavelet básica, mediante traslaciones y dilataciones. Estas funciones permiten reconstruir la señal original a través de la Transformada Wavelet inversa. La Transformada Wavelet no es solamente local en tiempo, sino también en frecuencia.

De manera muy general, la Transformada Wavelet de una función $f(t)$ es la descomposición de $f(t)$ en un conjunto de funciones $\Psi_{s,t}(t)$, que forman una base y son llamadas las “Wavelets”. La Transformada Wavelet se define como:

$$W_s(s, t) = \int f(t)\Psi_{s,t}^*(t)dt \quad (2.1)$$

Por otra parte, la desviación estándar del ruido σ que contamina una señal puede ser estimada a partir de la mediana de la desviación absoluta (MAD) de los coeficientes wavelets, lo cual se basa en la hipótesis (Donoho et al., 1994) de que la MAD de la banda de mayor frecuencia es proporcional a la desviación estándar del ruido. El estimador que se obtiene es:

$$\hat{\sigma}_{MAD} = \frac{1}{0,6745} MAD(y_i^H) \quad (2.2)$$

Donde y_i^H son los coeficientes wavelet de la banda de mayor frecuencia en la DWT. La operación MAD para un conjunto de datos X_i es definida como la mediana de las desviaciones absolutas de la mediana del conjunto, o sea:

$$MAD(X_i) = Median_i(|X_i - Median_j(X_j)|) \quad (2.3)$$

Este estimador es muy usado en los algoritmos para reducir el ruido blanco gaussiano en mediciones industriales (Cedeño, 2013). Sin embargo, si el ruido no es blanco la densidad espectral del mismo no es constante y por tanto este estimador no nos proporciona un valor correcto del nivel del ruido ya que el mismo sólo contiene



información de la potencia del ruido en la frecuencia más alta. Si el ruido es estacionario y correlacionado se ha establecido (Johnstone, 1997) que la varianza del ruido que afecta los coeficientes wavelets depende del nivel de descomposición y la varianza de las primeras bandas puede ser estimada por la fórmula 2.2, reemplazando los coeficientes de la banda de mayor frecuencia por los coeficientes de la banda a analizar. De esta forma podemos obtener estimaciones del ruido correspondientes a diferentes frecuencias. Si se desconoce la densidad espectral del ruido un estimador global razonable es, sin duda, la media aritmética de los estimadores obtenidos para cada banda. Si se incluyen, por ejemplo, las 4 primeras bandas se obtiene el estimador $\hat{\sigma}_{MADc}$ que se define como:

$$\hat{\sigma}_{MADc} = \frac{1}{4(0,6745)} \left(MAD(y_i^{H1}) + MAD(y_i^{H2}) + MAD(y_i^{H3}) + MAD(y_i^{H4}) \right) \quad (2.4)$$

donde y_i^{Hk} son los coeficientes wavelet de la banda k de la DWT ordenadas en orden decreciente de la frecuencia. Obsérvese que la fórmula 2.4 representa la media aritmética de las estimaciones del ruido en las 4 primeras bandas obtenidas por la fórmula 2.2.

Para llevar a cabo la estimación del ruido mediante Transformada Wavelet se utiliza la función madre Daubechies 4 y los siguientes comandos en MatLab™:

`[C,L] = wavedec(X,N,'wname')` -- Devuelve la descomposición wavelet de la señal X al nivel N , utilizando bien la wavelet específica ' $wname$ ' o un filtro de descomposición wavelet específico (Lo_D and Hi_D). La estructura de descomposición de la salida contiene el vector de descomposición wavelet C y el vector de contabilidad L . La estructura es organizada como se muestra en la figura 2.3.

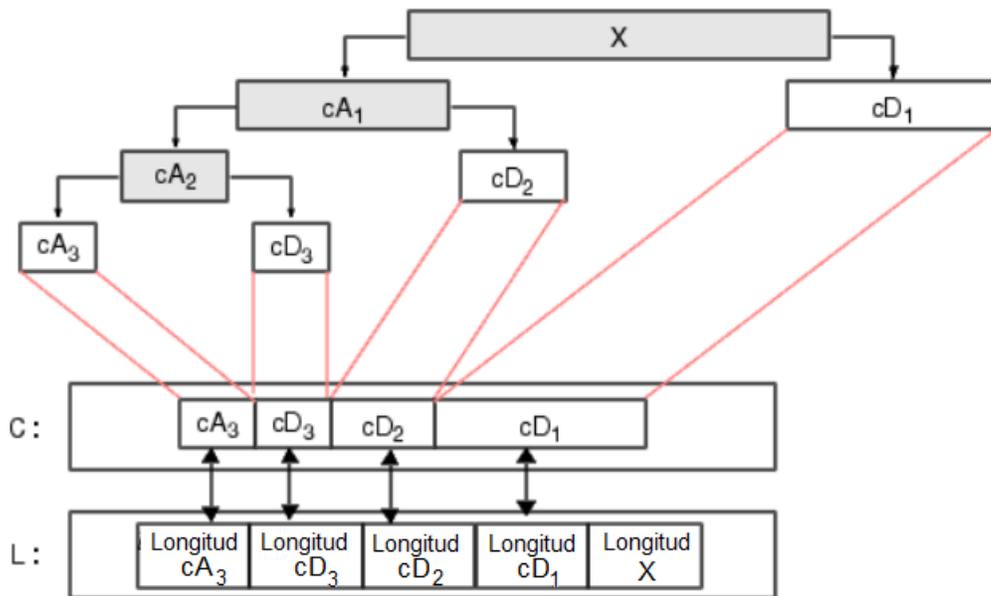


Figura 2.3. Descomposición wavelet de nivel 3

$D = \text{detcoef}(C, L, N)$ --Función de análisis wavelet uni-dimensional. Extrae los coeficientes detallados a nivel N de la estructura $[C, L]$ obtenida con `wavedec`.

$M = \text{median}(A)$ --Retorna la media de los elementos del arreglo A .

$B = \text{abs}(X)$ --Retorna el valor absoluto de X o la magnitud compleja en caso de ser un número imaginario.

Este algoritmo y otros que puedan ser empleados dan como resultado valores cuantitativos de la magnitud del ruido en la señal medida, en este caso la varianza (y por ende la desviación estándar) del ruido, pero hasta este momento todavía no se ha aplicado un proceso de filtrado o suavizado para removerlo. Este valor (desviación estándar del ruido) es usado como punto de partida para la configuración del método de detección de estados estacionarios y para tener una idea bastante clara del grado de suavizado que se le debe aplicar a la señal en caso de ser necesario.

2.2.2. Eliminación del ruido (filtrado o suavizado)

En condiciones ideales el nivel del ruido es insignificante con respecto al nivel de la señal, por lo que la reducción del mismo es innecesaria. En la práctica, sin embargo, es



generalmente necesario remover el ruido que altera la señal para cualquier análisis posterior de los datos. Si el nivel de ruido en las mediciones es muy alto, esto puede conllevar a errores en los resultados, por lo tanto, en ocasiones es necesario aplicar un filtrado a estos datos para eliminar ese ruido.

Este es un paso no obligatorio; el usuario, en base a la estimación proporcionada por el método anterior, decide si es necesario o no filtrar la señal.

2.2.2.1. Filtro de Media Móvil

Para esto, actualmente se cuenta con un procedimiento de filtrado llamado “Filtro de media móvil”. Básicamente consiste en fijar un ancho de ventana e ir calculando el valor medio de la variable dentro de esa ventana hasta llegar al final de la muestra y guardar los valores calculados dentro de un arreglo. Aunque presenta su equivalencia, este método tiene la ventaja de que no se necesitan configurar los parámetros de los coeficientes del numerador y denominador del polinomio de un filtro tradicional (como cuando se usa el comando *filter* del MatLab™). El ancho (en número de muestras) de la ventana de exploración es una magnitud mucho más comprensible para un usuario no experto en cuestiones de filtrado de series temporales. Una vez filtrada la señal, se puede elegir si de ahí en adelante se trabajará con esa nueva data o si se mantiene la señal original. El comando usado es:

$M = \text{mean}(A)$ --Retorna el valor medio de los elementos del arreglo A .

2.2.3. Detección de Estados Estacionarios

El punto de partida de la investigación es un registro finito X , el cual se considera una señal aleatoria. El propósito es detectar los segmentos de sus muestras que satisfacen una condición de EE y a cada uno calcularle su valor medio y de esta manera el espacio muestral de registros industriales X se transforma en otro, el de los valores medios de sus segmentos \bar{X} .

Como se mostró en el capítulo 1, se han desarrollado gran cantidad de métodos para la DEE entre los cuales lógicamente, existen diferencias: algunos presentan ventajas sobre otros pero también inconvenientes, o bien aumenta la precisión de la estimación pero



crece considerablemente la carga computacional o bien pueden ser inefectivos para señales contaminadas con un tipo específico de ruido pero muy eficientes con otro tipo de señales, etc.

Todas estas metodologías son aptas para ser implementadas en el IME ya que cuenta con una estructura modular preparada para el proceso de DEE. En este caso se elige un método que consiste en calcular la desviación estándar de las mediciones del proceso a través de una ventana móvil del historial de los últimos datos. Cada vez que el proceso no está en el estado estacionario, la desviación estándar medida es más grande que su valor de estado estacionario. Por lo tanto, cuando la desviación estándar medida es mayor que el valor de umbral prefijado, se establece la condición de estado no estacionario.

2.2.3.1. Detección de Estados Estacionarios mediante fraccionamiento basado en cálculo de desviación estándar (σ).

El método que se propone se basa en la detección de segmentos de muestras de una variable que satisfacen una condición de estacionalidad expresada por un nivel de tolerancia en la variabilidad de la media y la desviación estándar. Su fundamento se resume en los siguientes aspectos:

- a) Fraccionamiento de la señal en intervalos de EE basada en cálculo de desviación estándar (σ).
- b) Rectificación de coordenadas y verificación de tolerancia para cada segmento candidato.

Fraccionamiento de la señal en intervalos de EE

El algoritmo de fraccionamiento está basado en el cálculo de parámetros estadísticos sobre ventanas deslizantes de ancho L tomadas de la serie temporal:

$$X = \{x_1, x_2, \dots, x_N\}.$$

En cualquier instante k la media de las L muestras más recientes viene dada por:



$$\bar{x}_k = \frac{1}{L} \sum_{i=k-L+1}^k x_i \tag{2.5}$$

Este valor se conoce como la media de ventana móvil porque el promedio en cada instante k -ésimo se calcula con el conjunto de L valores más reciente. En otras palabras, en cualquier instante, la ventana móvil de L valores se utiliza para calcular la media de los siguientes datos de la secuencia. Esta magnitud es equivalente a la producida por un filtro pasa bajo. Las medias móviles pueden obtenerse de manera computacionalmente eficiente a partir de la siguiente expresión que vincula la media de ventana móvil en el instante k con la media de ventana móvil en el instante precedente.

$$\bar{x}_k - \bar{x}_{k-1} = \frac{1}{L} \left[\sum_{i=k-L+1}^k x_i - \sum_{i=k-L}^{k-1} x_i \right] = \frac{1}{L} [x_k - x_{k-L}] \tag{2.6}$$

$$\bar{x}_k = \bar{x}_{k-1} + \frac{1}{L} [x_k - x_{k-L}] \tag{2.7}$$

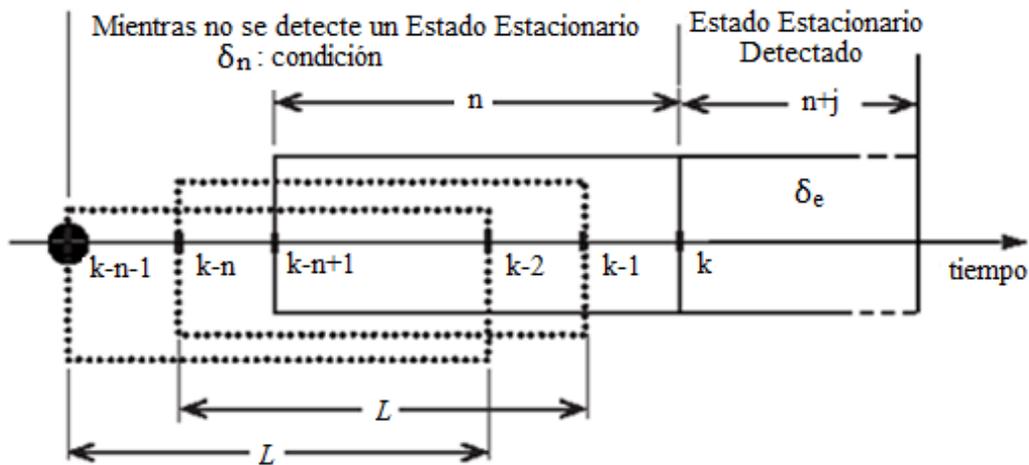


Figura 2.4. Metodología de trabajo del algoritmo de fraccionamiento



La varianza de la ventana móvil en el instante k es:

$$v_k = \frac{1}{L-1} \sum_{i=k-L+1}^k (x_i - \bar{x}_k)^2 \quad (2.8)$$

Por lo que la desviación estándar puede calcularse:

$$\sigma_k = \left(\frac{1}{L-1} \sum_{i=k-L+1}^k (x_i - \bar{x}_k)^2 \right)^{\frac{1}{2}} \quad (2.9)$$

Básicamente, en este algoritmo se va calculando la desviación estándar σ de manera recursiva en una ventana móvil y se va comparando con la resultante de la estimación de ruido (modificable por usuario si lo desea a través de la interface gráfica) y si es menor, entonces se considera que el segmento comprendido por la ventana se encuentra en EE, se avanza un paso y se repite el procedimiento hasta que deja de cumplirse la condición, momento en el cual se posiciona el inicio de la nueva ventana al final de la anterior y se inicia el ciclo de nuevo. Los autores señalan que el éxito con este método se basa en la capacidad de determinar correctamente las unidades de las variables clave, el período de tiempo usado para el cálculo y la desviación estándar (umbral).

De este primer paso del método se obtiene una lista con las coordenadas que identifican la secuencia de segmentos α_j (EE) y complementariamente (por exclusión) los segmentos β_i (ET). En otras palabras, se obtiene una matriz de dos columnas, de la forma $SS_i = [k_{i1} \ k_{i2}]$. El número de filas corresponde a la cantidad de estados estacionarios detectados, los valores de la 1ra columna corresponden a los puntos de inicio de dichos segmentos y los de la 2da columna a los puntos finales. Las consideraciones anteriores pueden resumirse en siguiente algoritmo que se muestra en el anexo 1.

La principal función usada es:

$s = std(X)$ --Devuelve la desviación estándar del vector X

Como ocurre con todos los métodos, existe un por ciento de probabilidad de que se incluya en un segmento β_i una muestra que realmente pertenece a un estado



estacionario. Este tipo de errores en los que una muestra contigua que pudiera incluirse en el segmento de EE erróneamente se clasifica como perteneciente a un segmento transitorio se conoce como error de tipo I (Rhinehart, 2013). Por el contrario, los casos en los cuales una muestra que realmente pertenece a un estado transitorio se clasifica como perteneciente a un segmento estacionario se conocen como errores de tipo II. Para disminuir los errores de tipo II a la salida del paso anterior se implementa el siguiente paso del método. Ya que el método de ventana móvil tiene una estructura simple, su aplicación a sistemas reales impone una carga pequeña de cálculo.

Rectificación de coordenadas y verificación de tolerancia.

En la primera etapa de la rectificación para cada segmento candidato $X(k)$ se calcula su media móvil Y_M con el objetivo de filtrar y suavizar los datos. Para que el vector de medias móviles resultante tenga la misma dimensión que el vector original se realiza una prolongación simétrica, en forma de espejo, del segmento candidato tal y como se muestra en la figura 2.5 con lo cual se garantiza que no se afecten los valores de la media de las primeras y últimas muestras. La media móvil se determina utilizando un ancho de ventana L_f y para el cálculo de la media en el instante k se consideran $L_f/2$ valores en instantes anteriores y posteriores. Obviamente, el vector Y_M tendrá una menor variabilidad que el vector $X(k)$.

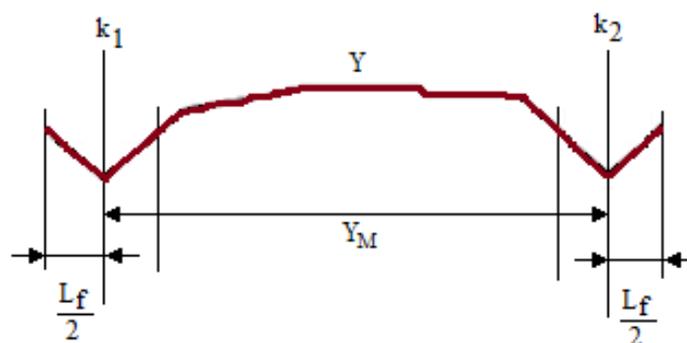


Figura 2.5. Prolongación del segmento candidato para el cálculo de la media móvil.

Una vez determinada la tendencia $Y_M(k)$ al segmento $X(k)$ empleando la técnica de la ventana móvil, el algoritmo de rectificación de coordenadas aplicado al segmento



candidato $X(k)$ de media μ , calcula la desviación absoluta que existe entre Y_M y μ , y evalúa el cumplimiento de la siguiente condición:

$$|Y_M(k_1) - \mu| < e_R \quad (2.10)$$

Aquellas muestras que no satisfagan la desigualdad son discriminadas. El segmento resultante se dice que tiene sus coordenadas rectificadas (figura 2.6).

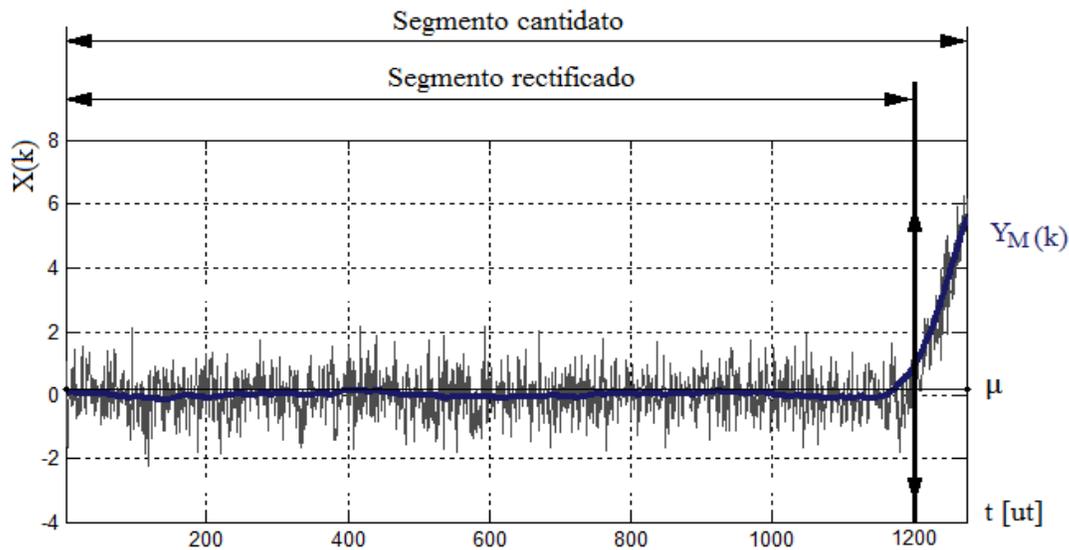


Figura 2.6. Rectificación y eliminación de muestras no estacionarias

2.2.4. Detección de coincidencias (DC)

En el IME, la DEE se realiza para cada una de las variables seleccionadas. Sin embargo, los procesos térmicos son, por lo general, multivariables. Esto demanda del diseño de un algoritmo que determine las coordenadas que corresponden a secuencias de muestras en EE coincidentemente para todas las variables (Ver anexo 2). En la determinación de cada segmento de coincidencia multivariable hay que imponer la restricción de validar solo aquellos segmentos que satisfagan una duración mínima t_{min} . La razón es de índole física pues un EE existe en un sistema físico cuando el conjunto de variables que caracterizan sus entradas y salidas se mantienen en condición de estacionalidad en un intervalo de tiempo más largo que la constante de tiempo mayor. Además de esto, como existe la posibilidad de que se creen segmentos de dimensión



muy pequeña que no tienen mucho sentido desde el punto de vista práctico, se cuenta con la posibilidad de establecer un Umbral de Tiempo Mínimo de manera que los segmentos que sean menor que este valor serán eliminados. Esto se conoce como *Filtro de Umbral de Tiempo*.

2.2.4.1. Detector de Coincidencias basado en lógica AND

El cómputo ofrece a través de su interfaz de usuario, la opción de tomar en cuenta hasta un máximo de 8 matrices de coordenadas, una por cada variable procesada. Su función es determinar, basándose en la función lógica AND, la matriz de coordenadas multi-dimensional a partir del conjunto de matrices de coordenadas por variable existente. El algoritmo se basa en crear vectores binarios (ceros y unos) unidimensionales para cada variable procesada de la misma longitud que la señal original. Se llenarán con (1) los espacios comprendidos entre las coordenadas de inicio y fin de EE y con (0) los espacios restantes (estados transitorios). De esta manera, al aplicar la operación lógica AND entre todas las variables, solo dará como resultado (1) en aquellas posiciones coincidentes para todos los vectores. Por último, se hace el procedimiento inverso; se convierte el vector unidimensional binario resultante a una matriz de coordenadas de 2 columnas llamada *matriz de coincidencias*, en la cual el número de filas podría ser mayor que el número de filas de la matriz de coordenadas por variable de mayor dimensión. El detector de coincidencia determina las coordenadas de aquellos segmentos en los cuales se cumple la condición de estacionalidad de las series temporales seleccionadas a tal efecto. Estos son los *Estados Estacionarios Multivariantes*.

Para verificar que el algoritmo opera correctamente se generan una serie de vectores de prueba cuyas coordenadas de inicio y fin de segmentos son conocidas para de esta manera determinar si las coordenadas del vector resultante coinciden con las esperadas (figura 2.7). En éstas se alternan los segmentos que representan posiciones valores cuyas muestras fueron detectadas en EE y complementariamente en ET. Las coordenadas delimitan los EE. El sistema puede aceptar hasta un máximo de 8 variables pero en este caso para la experimentación se consideraron 3 variables que son suficientes para validar el método. La característica de estas 3 distribuciones de segmentos es su desplazamiento entre ellas de modo irregular.

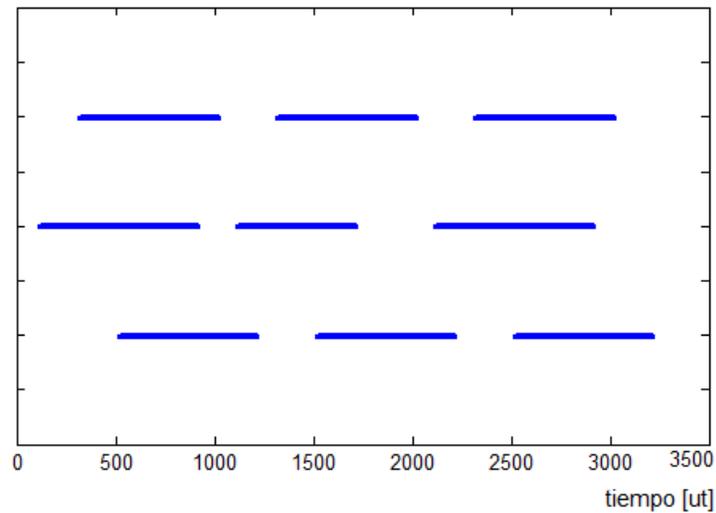


Figura 2.7. Vectores de prueba para la determinación de la efectividad en la determinación de segmentos coincidentes.

La lista de pares de coordenadas de entrada de las secuencias de la figura 2.7 son las siguientes:

$$S_1 = [(500,1200), (1500,2200), (2500,3200)]$$

$$S_2 = [(100,900), (1100,1700), (2100,2900)]$$

$$S_3 = [(300,1000), (1300,2000), (2300,3000)]$$

En la figura 2.8.a se muestra el resultado del diseño de una estructura formada por tres segmentos en EE por variable. La efectividad del filtro por umbral de tiempo se verifica configurando a t_{min} con dos valores. Por defecto $t_{min} = 0$ y segundo es $t_{min} > 0$. Al disponer de la lista de coordenadas que resulta para $t_{min} = 0$ se prueba evaluar su desempeño ante la segunda.

La lista de pares de coordenadas coincidentes de salida mostrados para $t_{min} = 0$ es:

$$S_C = [(500,900), (1500,1700), (2500,2900)]$$

En $t_{min} \geq 300$ se discrimina un segmento coincidente que no tendría sentido físico. El resultado (figura 2.8.b) es:

$$S_C = [(500,900), (2500,2900)]$$

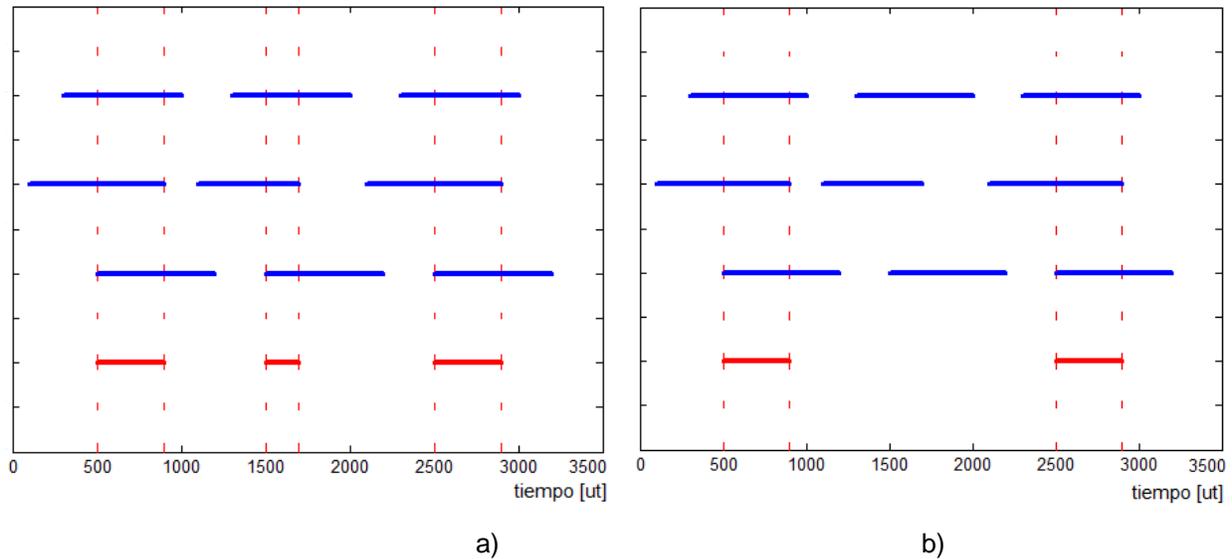


Figura 2.8. Resultado en la determinación de segmentos coincidentes sobre las secuencias a:
a) Para $t_{min} \geq 0$ y b) $t_{min} \geq 200$.

Tanto a simple vista como numéricamente queda demostrada la efectividad del algoritmo de coincidencia y el Filtro por Umbral de Tiempo.

Los comandos utilizados para la coincidencia son:

`B = zeros(m,n)` --Crear un arreglo de dimensión (m*n) con todos los valores cero.

`B = ones(m,n)` --Crear un arreglo de dimensión (m*n) con todos los valores uno.

Operación &: --La expresión `expr1 & expr2` representa la operación lógica AND entre los valores, arreglos o expresiones `expr1` y `expr2`. Si `expr1` es verdadero y `expr2` es verdadero, entonces la operación AND entre esas entradas resulta verdadero. Si cualquiera de esas entradas es falso, el resultado será falso. En MatLab™, si `xpr1` y `expr2` son arreglos, se realiza la operación lógica AND en cada elemento del arreglo y la salida es un arreglo.

Con la aplicación del método IME y del algoritmo de coincidencia se tienen los segmentos de muestras coincidentes y a partir de estos se calculan los valores medios de cada variable. De cada segmento se conoce su correspondiente fecha y tiempo inicial. Estos son los datos de salida del 1er módulo y de entrada del 2do que opera sobre los valores



medios de la secuencia de EE por cada variable y podrá predecir el vector de datos con los valores esperados de buen desempeño para cada variable seleccionada. Está dotado de un procedimiento para la construcción de una función vectorial \vec{H} mediante interpolación 1D de las variables seleccionadas con la potencia para la predicción de sus valores esperados. Bajo estas premisas calculan los vectores δY_{ITj} con las correspondientes desviaciones para cada j-estado estacionario.

2.3. Análisis del módulo de Detección de Desviaciones

El 2do módulo se nombró Deval (del inglés de **D**eviations **E**valuator). Aquí están programadas las funciones de:

- Cargar datos en condiciones de buen desempeño y mediciones actuales
- Constructor de funciones interpolantes para predicción de valores esperados de las variables
- Cálculo de desviaciones respecto a valor deseado de cada variable
- Organización y Agrupamiento por fechas

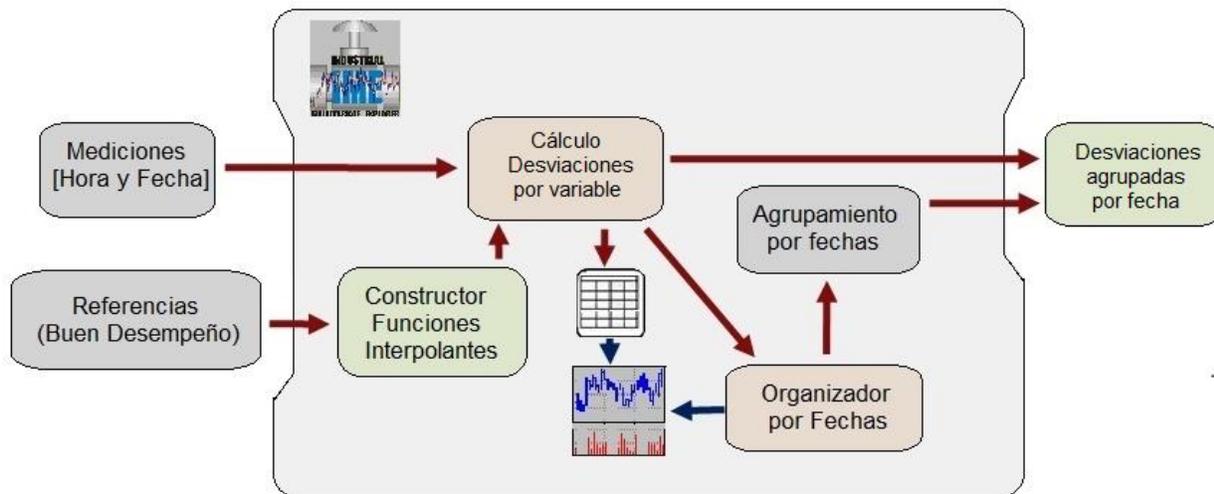


Figura 2.9. Prestaciones del módulo de Detección de Desviaciones

Cargar Datos: Primeramente se debe cargar el juego de datos correspondiente a las referencias (tomado en condiciones de buen desempeño de la planta) y luego los datos de las mediciones que serán procesadas. En todas las etapas del procesamiento que se necesite cargar o exportar un juego de datos se puede hacer desde un fichero MatLab



(con extensión `.mat`) o desde un fichero Excel. Para el manejo de ficheros Excel se emplearon las siguientes funciones:

`[num, txt, raw] = xlsread(filename)` --Devuelve el dato contenido en la 1ra hoja del fichero Microsoft Excel llamado `filename`. Dicho dato queda dividido de la siguiente manera: en `num` quedan los datos numéricos, en el arreglo de celdas `txt` quedan los datos tipo texto y en `raw` quedan tanto los datos numéricos con textuales en la misma disposición de celdas en que se encontraban originalmente en el fichero Excel.

2.3.1. Monitorización de las desviaciones a largo plazo

Como se abordó en el epígrafe 1.6.1, las degradaciones de la UGE y sus efectos sobre las mediciones ocurren en intervalos de tiempo de varias horas y en ocasiones, por sus efectos acumulativos, de varios días. La monitorización de las desviaciones entre el valor actual y el valor esperado requiere de un modelo que prediga este último.

La implementación de este procedimiento para la monitorización requiere el desarrollo de los siguientes bloques funcionales:

- i. Constructor de las funciones interpolantes para la predicción de los valores de referencia para cada valor actual de potencia N_p . Recibe como entrada los valores medios que recibe de la DEE cuando los registros son adquiridos bajo condición de buen desempeño de la operación.
- ii. Tabulador de valores medios por filas atendiendo a la fecha inicial y tiempo (primera columna). En la segunda columna la potencia N_p y a continuación cada variable seleccionada.
- iii. Calculador de las desviaciones por variables para cada potencia N_p .
- iv. Organizador de las desviaciones por fecha a partir de la menos reciente.
- v. Agrupador de las desviaciones a diferentes escalas de tiempo, por ejemplo; por intervalo de días consecutivos, días, semanas y de meses con delimitación fecha y tiempo inferior y superior.



2.3.1.1. Constructor de funciones interpolantes

Para llevar a cabo el 1er punto (construir funciones interpolantes para la predicción de los valores de referencia para cada valor actual de potencia) se pueden usar los numerosos métodos de regresión no paramétrica conocidos.

Cualquiera de los métodos generales de interpolación polinómica existentes que permiten aproximar una función por un polinomio de grado m como son el método de las diferencias divididas de Newton, la interpolación de Lagrange o la interpolación de Hermite y otras pueden ser probadas en el IME y comprobar ventajas y desventajas de cada una de ellas.

El procedimiento deber partir de una muestra de series temporales de un conjunto de p variables Y_i seleccionadas y cada uno de estos valores tiene asociando un elemento D_T de su fecha y tiempo correspondiente.

Se tienen una cantidad q de valores medios que resultaron del procesamiento del 1er módulo en la detección de segmentos de estados estacionarios para cada una de las p variables sobre las series de tiempo. Las sucesiones de las p variables, son adquiridas cercanas al arranque de la unidad, después de un mantenimiento, por lo que estas medidas son consideradas de buen desempeño relativo. Se identifican con el subíndice 'o'. Cada uno de estos valores, en condición de buen desempeño, se obtuvo para cada valor medio en potencia $N_{p_{oi}}$:

El procedimiento se realiza de forma independiente para cada variable:

La variable Y_{oi} es función de la potencia $N_{p_{oi}}$ $i = 1 \dots \dots \dots p$

Consiste en construir por interpolación entre los q -nodos una serie de funciones por tramos $H_i(x)$ que permita predecir, en lo sucesivo, a Y_i para cualquier valor de N_p (acotado dentro del rango para el cual fue previsto en la construcción de las funciones).

Construcción de polinomio interpolador mediante interpolación de Hermite

Se conoce que la secuencia de valores de las variables esenciales en los generadores de vapor (flujos de agua de alimentación al domo y de vapor sobrecalentado y recalentado a turbina, presiones y temperaturas de gases y vapor sobrecalentado y



recalentado), exhiben monotonía con las mediciones de N_p en el tiempo, es decir, no existe punto crítico intermedio. De las características antes mencionadas y por diseño del sistema de control en una EGU, sus variables controladas (Temperatura, Presión y Flujo) pueden ser consideradas auto-reguladas pero existe otro tipo que no tienen valores máximos en todo el rango de operación, es decir, sus mediciones son monótonas con la potencia.

Se elige la interpolación cúbica polinomial por tramos de Hermite porque es un método en el cual las pendientes entre dos puntos consecutivos están determinadas de tal modo que se conserva la forma de los datos y la monotonía.

Se construye un polinomio Interpolador de Hermite de grado $2m+1$ de la forma:

$$H_{2m+1}(x) = \sum_{i=0}^m f_i \Phi_i(x) + \sum_{i=0}^m f'_i \Psi_i(x) \quad (2.11)$$

donde:

$$\Phi_i(x) = \left(1 - 2l'_i(x_i)(x - x_i)\right) l_i^2(x) \quad (2.12)$$

$$\Psi_i(x) = (x - x_i) l_i^2(x) \quad i = 0, \dots, m \quad (2.13)$$

$l_i(x)$ son los llamados polinomios de Lagrange, que se calculan de este modo:

$$l_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j} = \frac{(x - x_0)(x - x_1) \dots (x - x_{j-1})(x - x_{j+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{j-1})(x_i - x_{j+1}) \dots (x_i - x_n)} \quad (2.14)$$

Donde siempre $l_i(x) \neq 0$

Se obtiene una serie de funciones $H_i(x)$ correspondientes a cada uno de los i tramos de los q valores medios detectados.

Se integran todas las H_i en la función vectorial

$$\vec{H} = (H_1, H_2, \dots, H_q) \quad (2.15)$$



para posteriormente, poder predecir un vector de q -valores de buen desempeño de la variable Y_i para cada elemento N_{p_i} .

Para lograr este objetivo, en la plataforma MatLab™ se utiliza el comando:

`v = ppval(pp, xx)` --Devuelve el valor de la estructura polinomial f contenida en `pp` a la entrada de `xx`. Se obtiene `v` reemplazando cada entrada de `xx` por el valor de f en esa posición.

2.3.1.2. Cálculo de las desviaciones por variable

Primeramente se procede a cargar el nuevo juego de datos a analizar correspondiente a mediciones de cualquier otro momento de la operación de la planta que contiene los valores medios en estado estacionario de potencia N_{p_i} y la variable a procesar Y_i : $[Y_i, N_{p_i}]$.

Evaluando en la función \vec{H} , cada uno de los valores de N_{p_i} según corresponda, se predicen cuáles deben ser los valores esperados Y_{oi} :

$$Y_{oi} = \vec{H}(N_{p_i}) \quad (2.16)$$

Luego se define:

$$\delta Y_i = Y_i - Y_{io} \quad (2.17)$$

δY_i : Sucesión de las desviaciones de la variable Y_i

Y_i : i -ésimo valor medio medido de la variable Y

Y_{io} : i -ésimo valor medio esperado de buen desempeño de la variable Y .

Este procedimiento se repite para cada una de las variables $[Y_1, Y_2, \dots, Y_p]$.

Las funciones MatLab™ utilizadas son las siguientes:

Existe una solución de computo mediante el comando **pchip** del paquete MatLab™ que preserva la monotonía, C^1 (Abbas y otros, 2012). El empleo de ese comando viene asociado al también comando **ppval**. El segundo calcula el valor de la variable



dependiente para cualquier valor de la variable independiente utilizando el polinomio obtenido en **pchip**.

`pp = pchip(x,y)` --Devuelve e la estructura polinomial por partes para su uso por `ppval` que puede ser un vector fila o columna. El vector `y` debe ser de la misma longitud que `x`.

De este modo se van determinando, para cualquier punto de operación en potencia, N_{pi} , las desviaciones relativas que ocurren en ventanas de corto plazo y evaluar su evolución en ventanas de largo período mediante agrupación por fechas y tiempo.

2.3.2. Organización y Agrupamiento por fechas.

Como cada variable Y_i (y por ende cada δY_i) tiene asociado su atributo de fecha y tiempo D_i , haciendo pequeños grupos de los elementos de δY_i por estos atributos se podrá examinar los efectos de las degradaciones entre estas ventanas de corto plazo. Los comandos usados para lograr este objetivo son:

`B=sort(A)` --Ordena los elementos a lo largo de la dimensión del arreglo A y los organiza de manera ascendente.

Los resultados de este módulo se dan tanto en forma gráfica como en formato numérico a través de una tabla que relaciona los valores que debería tener la variable (en condiciones de buen desempeño), los valores reales, la desviación entre ellos y su atributo correspondientes de fecha y tiempo.

Con este esquema se logra también que la aplicación se convierta en un laboratorio para el ensayo de algoritmos y métodos. Los registros de entrada se agrupan bajo el término denominado base de datos primaria (BDP). A las desviaciones que se determinan como resultado final se agrupan bajo el término denominado base de batos secundaria (BDS).

2.4. Aplicación de los algoritmos al generador de vapor # 2 de la CTE Felton

Sobre la base del criterio de monitorización indicado anteriormente se ilustra la comprobación de la efectividad de los métodos programados en el IME en el generador de vapor de la unidad # 2 de la Termoeléctrica de Felton. Por razones de seguridad



industrial que afectan negativamente la disponibilidad de momentos de toma de registros para la experimentación, los resultados se presentan de manera gráfica y tabulada a partir de un conjunto específico de ellos exportados en varios períodos de oportunidad. Para el establecimiento de magnitudes de referencia relativa, los días cercanos a un arranque después de un mantenimiento es buen momento. Durante los períodos de paradas de la unidad para mantenimiento se producen correcciones (por ejemplo, limpieza de superficies, eliminación de salideros, ajustes de holguras, entre otras) y por ende se actualiza el comportamiento de buen desempeño relativo.

Se eligen un conjunto de variables de las más representativas de la dinámica del sistema como son la potencia activa de generación, el flujo de combustible, la temperatura de los gases de salida de la chimenea, el flujo de agua al generador de vapor, el flujo del vapor sobrecalentado que sale del SH4, la temperatura a la entrada del economizador, la presión de vapor principal a la entrada del Calentador de Alta Presión (CAP) y el flujo de aire a los quemadores como algunas de las más importantes.

Para el establecimiento del sistema de referencia en este ejemplo se aprovecha una parada de mantenimiento parcial realizada a principios de enero del 2013. Las mediciones correspondientes a los días sucesivos inmediatos a esta parada se procesan en el 1er módulo del IME y se obtiene (luego de realizar la coincidencia multivariable) un total de 11 valores a Estado Estacionario de estos registros que serán suficientes para cubrir los puntos del rango referencial relativo de operación de potencia. Estos datos quedan tabulados en tablas que asocian cada valor de potencia con el correspondiente de las restantes variables en ese momento de tiempo y son entradas a la implementación de la función vectorial \vec{H} .

Dado que ésta es una metodología de trabajo concebida para ser aplicada a largo plazo, para notar el deterioro progresivo de las condiciones de buen desempeño, se procesan datos correspondientes a los meses sucesivos de trabajo de la planta que comprenden desde el 20 de enero hasta el 12 de marzo del 2013.

Las gráficas mostradas a continuación son ejemplos de los resultados del cálculo de los EE para tres de las variables procesadas en dos de las fechas mencionados anteriormente (20 de enero y 22 de febrero de 2013).

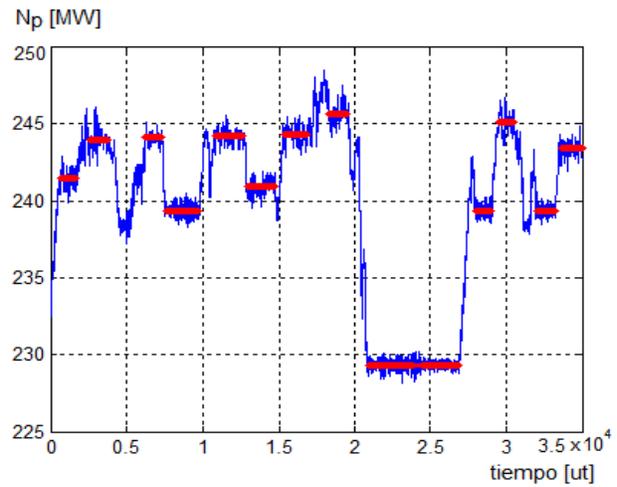
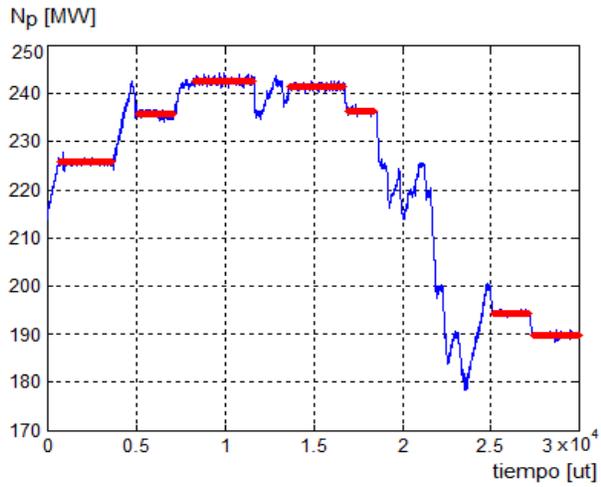


Figura 2.10. Resultados de la detección de EE para la Potencia

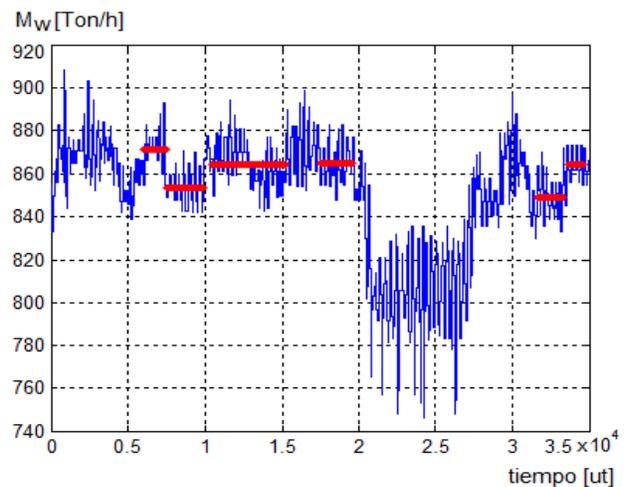
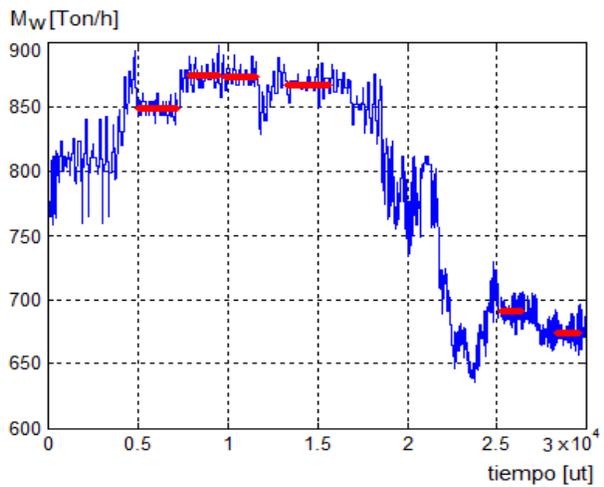


Figura 2.11. Resultados de la detección de EE para el flujo de agua al generador de vapor

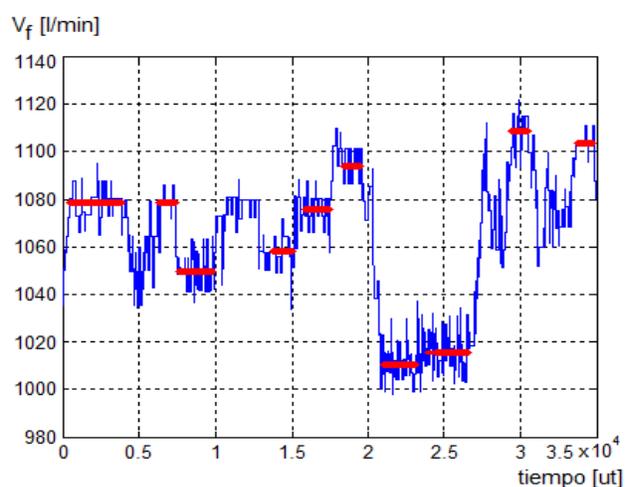
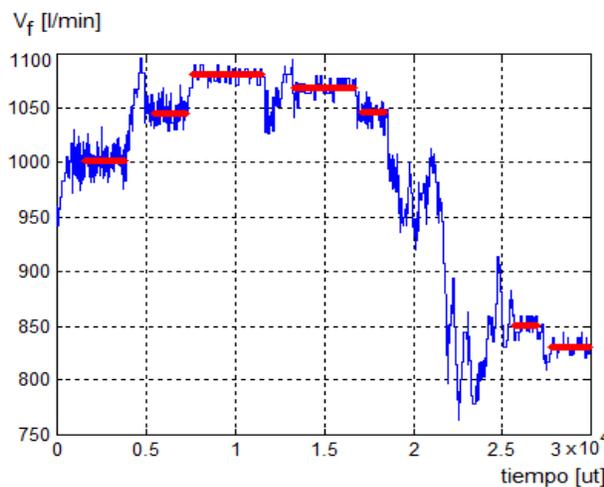


Figura 2.12. Resultados de la detección de EE para el flujo de combustible



A la hora de pasar al cálculo de desviaciones (2do módulo), para mayor claridad y entendimiento de los resultados gráficos se tomaron no más de 12 valores resultantes de ésta estimación de EE en todo este rango de tiempo. Su agrupación en ventanas de 10 días (como ejemplo) permite que en ventanas de largo plazo se obtengan los resultados que se grafican en la figura 2.13 y 2.14. Esto hace que los resultados de los días 16 y 22 de febrero queden en un solo grupo. La leyenda de colores es como sigue:

—●— Referencia —●— 20 enero 2013 —●— 16 y 22 febrero 2013 —●— 12 marzo 2013

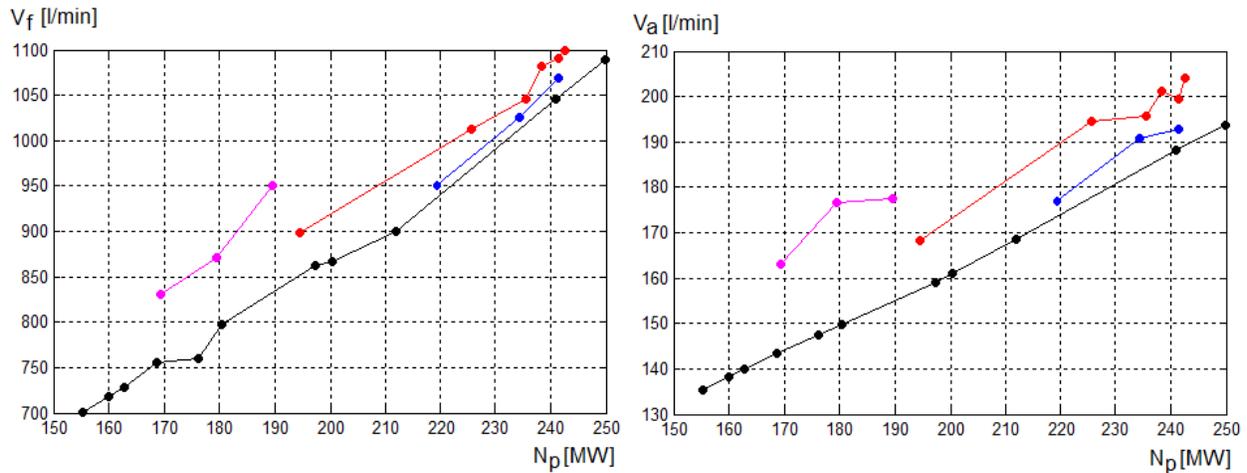


Figura 2.13. Desviaciones de: a) flujo combustible, b) Flujo de aire a los quemadores

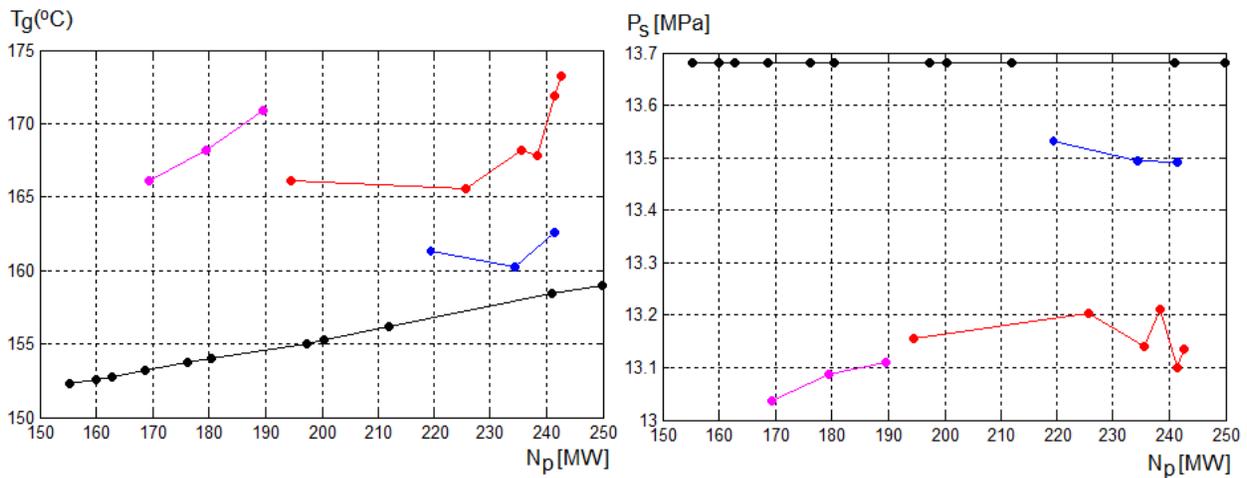


Figura 2.14. Desviaciones de: a) temperatura en gases de salida a la chimenea b) presión de vapor a la entrada del CAP



En las figuras, las sucesiones de puntos negros son los valores de referencia de las variables bajo análisis para cada uno de los valores de potencia. Cuenta con 11 puntos específicos para cada variable resultantes del proceso de interpolación polinómica por tramos y representan los valores que debería presentar dicha variable en condiciones de buen desempeño relativo. Los puntos (coloreados) de los restantes grupos muestran los valores reales que presentaban esas variables en cada momento. En cada una de ellas se pueden notar palpables desviaciones respecto a la referencia y sobre todo como éstas desviaciones se van acrecentando con el transcurso de los días derivándose el siguiente análisis termo-energético:

La degradación por ensuciamiento se manifiesta, entre otras, en las mediciones de temperatura. Se observa una incidencia en el incremento tanto de la temperatura de los gases a la salida debido a ensuciamiento de las superficies del eje convectivo como a la entrada e interior del CAR (no mostradas en figura). Toda esta energía adicional perdida en forma de calor en los gases a la atmósfera debe ser contrarrestada por el sistema de control mediante un incremento del flujo de combustible por lo que existe un sobreconsumo. Estas afectaciones provocan también una caída en la presión a la entrada del cilindro de alta presión, así como la disminución de las temperaturas a la salida del sobrecalentador y del recalentador y otras reacciones en cadena debido a la naturaleza intrínsecamente multivariable del proceso.

Este es un ejemplo indiscutible de como el IME con sus diferentes métodos y algoritmos (abiertos a cambios y modificaciones) puede ser usado de forma efectiva para monitorear a largo plazo el comportamiento de las variables en los generadores de vapor de la CTE Felton (extensible a otras) constituyendo la base para el análisis en retrospectiva de cuáles son las causas de estos malos comportamientos y para el diseño de métodos y procedimiento para alcanzar metas orientadas al desarrollo de sistemas de cómputo que actúen como asistente automatizado de la operación, sistema de alerta temprana entre otras posibilidades.



Conclusiones del capítulo

Se diseñó un bloque funcional que permite implementar flexiblemente la inserción de métodos que determinen el nivel de ruido en las mediciones y su filtrado en caso de ser necesario. En particular, se desarrolla la programación de un método basado en Transformada Wavelet Discreta que resulta efectivo tanto para muestras contaminadas con ruido blanco como con ruido coloreado.

Se creó una estructura que posibilita la discriminación de la señal bajo análisis en segmentos de EE y ET de manera independiente por variable teniendo en cuenta las características dinámicas de cada una de ellas empleando el cálculo recursivo de la desviación estándar y cuya salida son las coordenadas de inicio y fin de cada segmento de EE.

Se implementó el método de interpolación de Hermite que facilita el ajuste de los polinomios interpoladores necesarios para la predicción de valores de referencia de las variables para cualquier potencia y se crearon las condiciones para darle solución al cálculo de las desviaciones entre los valores esperados y los medidos de estas variables.

El empleo de interfaces graficas introduce un ambiente adecuado y de fácil manejo para el ensayo y evaluación rápida y efectiva de los resultados de las diferentes propuestas de solución que necesiten ser consideradas.



Conclusiones Generales

Al término de la presente investigación se arriba a las siguientes conclusiones:

Con el estudio de las técnicas de detección de ruido y de los métodos de identificación de estados estacionarios en procesos industriales, se ha enfrentado una línea de investigación, la monitorización de sistemas de tiempo real cuyo rol es fundamental para crear las bases en la identificación de las causas que modifican la correcta operación de la planta.

Se diseñó un módulo computacional basado en bloques funcionales abiertos a cambios para la inserción de algoritmos de estimación del nivel de ruido, filtrado, extracción de tendencias y detección de estados estacionarios en registros de proceso que constituye la primera fase de la monitorización y los datos de entrada al siguiente módulo.

Se creó un segundo módulo también compuesto por bloques funcionales que integra procedimientos de interpolación con polinomios de segundo orden para la predicción de valores de referencia de las variables para cualquier potencia y el cálculo de las desviaciones entre los valores esperados y los resultantes de la detección de EE.

Se logró la integración de los bloques funcionales del sistema de cómputo en una interfaz gráfica de varias pantallas que garantiza la flexibilidad en el uso interrelacionado y sincronizado de sus módulos componentes quedando así constituida la plataforma computacional IME que sirve de base a la determinación de estados estacionarios con fines de monitorización.

La solución instalada en la estructura propuesta, ha revelado su efectividad tanto ante experimentos con series temporales sintéticas como con registros industriales del generador de vapor de la Central Termoeléctrica Felton.



Recomendaciones

Al culminar este trabajo, y desarrolladas todas las tareas propuestas en el mismo sugerimos:

Introducir métodos para el tratamiento de posibles errores en las mediciones como errores gruesos, valores atípicos y valores perdidos.

Proponer y ensayar con otros métodos para la extracción de tendencias que requiere el método IME en el paso de la rectificación de coordenadas.

Establecer un catálogo portador del conocimiento de expertos de las unidades de generación eléctrica que relacione a las desviaciones de variables con posibles anomalías para pasar a la fase de diagnóstico. También podrá ser enriquecido con los sucesos que se desarrollen en la industria.



Bibliografia

Alekman, S.L., 1994. *Control for the process industries*. Putman Publications, Chicago, IL, 7(11):62, November.

Almasy, G. A., and Mah, R. S. H., 1984. *Estimation of measurement error variances from process data*. Ind. Eng. Chem. Process Des. Dev. 23, pp 779-784.

Bagajewicz, M. J., 2001. *Process Plant Instrumentation: Design and Upgrade*, Technomic Publishing Company, Inc. ISBN No. 1-56676-998-1

Bagajewicz, M. J., 2010. *Smart Process Plants. Software and Hardware Solutions for Accurate Data and Profitable Operations*. McGraw-Hill.

Ballester, J, García, AT., 2010. *Diagnostic techniques for the monitoring and control of practical flames*. Prog Energy Combust Sci; 36/4:375–411.

Basseville, M., Nikiforov, I. V., 1993. *Detection of abrupt changes—theory and applications*. Englewood Cliffs, NJ, USA: Prentice-Hall.

Biegler, L. T., Damiano, J. J., and Blau, G. E., 1986. *Non-linear parameter estimation: A case study comparison*. AIChE 1. 32, pp 29-43.

Blanco, J.M., Peña, F., 2008. *Analytical study of the effects of the clogging of a mechanical precipitator unit in air preheaters in a high-performance thermoelectric power plant based on available data*. ASME J. Eng. for Gas Turbines and Power, 130/2.

Branislav, S., Radisa, J., 2011. *Software system “DIORES” for the operation diagnosis of a steam power plant unit*. Energy; 36(2):1187-95.

Brown, P.R., Rhinehart, R.R., 2000. *Development and demonstration of a method for automated steady-state identification in multivariable systems*, Hydrocarbon Processing 79, September 9, pp 79–83.

Cao, S., Rhinehart, R., 1995. *An efficient method for on-line identification of steady state*, Journal of Process Control 5, pp 363–374.



- Cedeño, A., Trujillo, R., 2013. *Estudio comparativo de técnicas de reducción de ruido en señales industriales mediante Transformada Wavelet Discreta y selección adaptativa del umbral*. Revista Iberoamericana de Automática e Informática industrial, vol 10.
- Chen, J., Romagnoli, J. A., 1998. *A strategy for simultaneous dynamic data reconciliation and outlier detection*. Computers and Chemical Engineering, 22(4/5), 559–562.
- Chen, J. H., Liao, C. M., 2002. *Dynamic process fault monitoring based on neural network and PCA*. Journal of Process Control, 12(2), pp.277-289.
- Couch, II, L.W., 2002. *Digital and analog communication systems*. 6th ed. Prentice Hall.
- Coursey, W.J., 2003. *Statistics and Probability for Engineering Applications with Microsoft Excel*. Newnespress. Elsevier Science.
- Cuza, O., 2011. *Influencia del cambio de combustible en el generador de vapor de la termoeléctrica de Felton*. Tesis en opción al grado académico de Máster en Eficiencia Energética. Instituto Superior MineroMetalúrgico de Moa.
- Dolabdjian, C., Fadili, J., Leyva, E. H., 2002. *Classical low-pass filter and real-time wavelet-based denoising technique implemented on a DSP: a comparison study*. The European Physical Journal Applied Physics 20, pp 135–140.
- Donoho, D. L., Johnstone, I. M., 1994. *“Ideal spatial adaptation by wavelet shrinkage”* Biometrika, vol. 81, p 425–455.
- Fajardo, M. 2010. *Evaluación de la efectividad del sistema de limpieza acústico en la Unidad # 4 de la CTE “Antonio Maceo”*. Tesis en opción al grado académico de Máster en Eficiencia Energética. Universidad de Oriente.
- Finn, J., Wagner J, Bassily, H., 2010. *Monitoring strategies for a combined cycle electric power generator*. Appl Energy, 87(8):2621–7.
- Geng, Z., Li, K., 2003. *Factorization of posteriors and partial imputation algorithm for graphical models with missing data*. Statistics and probability letters. 64, 369-379.



Gleason, T., Staelin, R., 1975. *A proposal for handling missing data. Psychometrika*. Vol 40, 2. 229-252.

Härdle, W., Müller, M., Sperlich, S. y Werwatz, A., 2004. *Nonparametric and Semiparametric Models*. Springer.

Horowitz, G, et. al., 2010. *Monitoreo Inteligente de Procesos*, TAMI – YPF- Reporte final.

Johnstone, I. M., Silverman B. W., 1997. “*Wavelet Threshold Estimators for Data with Correlated Noise*”. *Journal of the Royal Statistical Society. Series B*, vol. 59.

Jubien, G. y Bihary, G. 1994. *Control for the process industries*. Putman Publications, Chicago, IL, 7(11):64, November.

Keller, J. Y., Zasadzinski, M., and Darouaeh, M., 1992. *Analytical estimator of measurement error variances in data reconciliation*. *Comput. Chem. Eng.* 16.

Kelly, J. D., Hedengren J. D., 2013. “*A steady-state detection (SSD) algorithm to detect non-stationary drifts in processes*”, *Journal of Process Control*, vol 23, p 326– 331.

Kim, M., Yoon, S.H., Domanski, P., 2008. Payne W. *Design of a steady-state detector for fault detection and diagnosis of a residential air conditioner*, *International Journal of Refrigeration*, vol 31, p 790–799.

Kurz, R., Brun, K., Wollie, M., 2009. *Degradation Effects on Industrial Gas Turbines*. *Journal of Engineering for Gas Turbines and Power*, Vol. 131, pp 062401- 062407.

Little, R., Rubin, D., 1987. *Statistical Analysis with Missing Data*. Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc. New York.

Mah, R. S. H. and Tamhane, A. C., 1982. *Detection of Gross Error in Process Data*. En: *AIChE J.*, Vol. 28, No. 5, pp. 828-830.

Narasimhan, S., Kao, C., Mah, R., 1987. *Detecting changes of steady states using mathematical theory of evidence*, *AIChE Journal*. 33, 1930–1932.



- Narasimhan, S, Jordache C., 2000. *Data reconciliation & gross error detection: an intelligent use of process data*. Houston, USA: Gulf Professional Publishing.
- Neha, S., Vilankar, K. P., Rhinehart, R.R., 2010. *Type-II critical values for a steady-state identifier*, *Journal of Process Control*, vol 20, p 885-90.
- Ott, L., 2001. *An Introduction to Statistical Methods and Data Analysis*. 5th ed. Wadsworth Group. Division of Thomson Learning, Inc.
- Palmero, M., 2011. *Herramienta para el Diagnóstico del Ciclo Térmico de la Empresa Termoeléctrica "Lidio Ramón Pérez"*. Tesis en opción al grado académico de Máster en Eficiencia Energética. Instituto Superior Minero Metalúrgico de Moa.
- Peña, F., Blanco, J.M., 2006. *Evaluation of the physical dew point in the economizer of a combined cycle burning natural gas*. *Appl. Therm. Eng*, 27, pp. 2153-2158.
- Prieto, L.F., 2008. *Análisis del comportamiento de la caída de presión en el eje convectivo del generador de vapor en la Unidad # 6 de la CTE "10 de octubre" con la quema del crudo cubano*. Tesis en opción al grado académico de Máster en Eficiencia Energética. Universidad de Oriente.
- Rhinehart, R.R. 2013. *"Tutorial: Automated Steady and Transient State Identification in Noisy Processes"*, Proceedings of the 2013 American Control Conference, Washington DC, 17-19 June, pp. 4477-4493.
- Rioul, O., Vetterli, M., 1991. Oct. *Wavelet and signal processing*. IEEE Signal processing Magazine 8 (4), pp 14–38.
- Seborg, D.E, Edgar T.F, Mellichamp D.A., 2004. *Process Dynamics and Control*. Capítulo 21 Process Monitoring. Pp 567 – 590. Editora John Wiley & Sons, Segunda Edición.
- Serth, R. W., & Heenan, W. A., 1986. *Gross Error Detection and Data Reconciliation in Steam-Metering Systems*. En: AIChE J., Vol. 32, No. 5, p. 737.

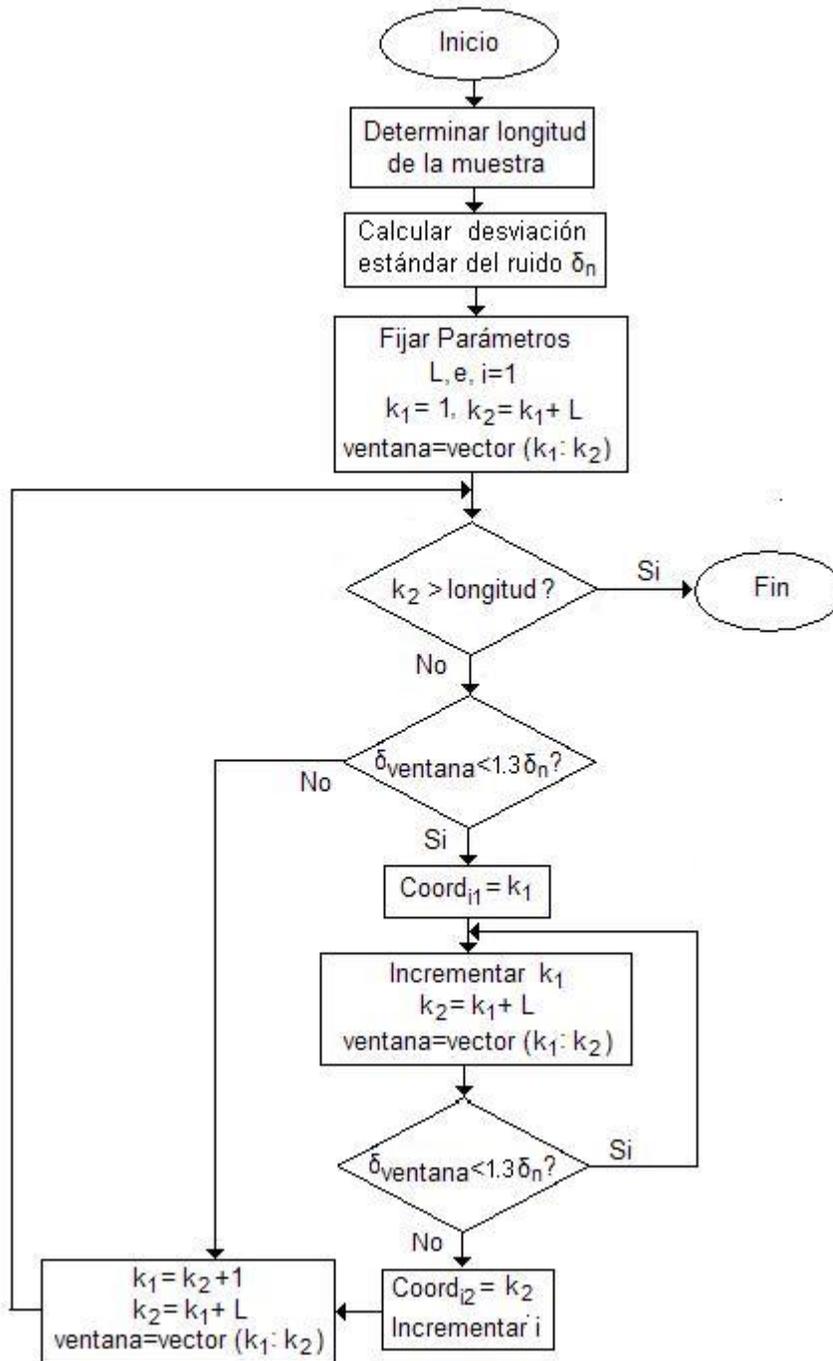


- Spitzbart, A., 1960. *A Generalization of Hermite's Interpolation Formula*. American Mathematical Monthly. 67(1): pp. 42–46.
- Stein, M.L., 1999. *Interpolation of Spatial Data. Some Theory for Kriging*, Springer, New York.
- Stockill, D., 2002. *Handling Uncertainty through the Plant Life Cycle using Statistics and Data Mining*. Amsterdam: Elsevier. Amsterdam, Netherlands, pp.70.
- Teruel, E., Cortés, C., Díaz, L. I., Arauzo, I., 2005. *Monitoring and prediction of fouling in coal-fired utility boilers using neural networks*, Chemical Engineering Science 60, pp. 5035 – 5048.
- Toffolo, A., Lazzaretto, A. 2007. *A New Thermo-economic Method for the Location of Causes of Malfunctions in Energy Systems*. Journal of Energy Resources Technology, Vol. 129.
- Vázquez, L., Llosas, Y., et. al., 2010. “*El diagnóstico energético de la operación en centrales térmicas con el monitoreo de los índices de sobre-consumo.*” Revista electrónica Ciencia en su PC, vol3, URL: 169.158.189.18/cienciapc.
- Wang, X. Z., 2001. *Knowledge Discovery through Mining Process Operational Data. In Application of Neural Networks and other Learning Technologies in Process Engineering*. (Mujtaba, I. M. y M. A. Hussain eds.). Imperial College Press, London, pp. 287-327.
- Xia, R., Meng, K., Qian, F., 2007. Wang, Z.-L. *Online wavelet denoising via a moving window*. Acta Automatica Sinica 33 (9), pp 897 – 901.



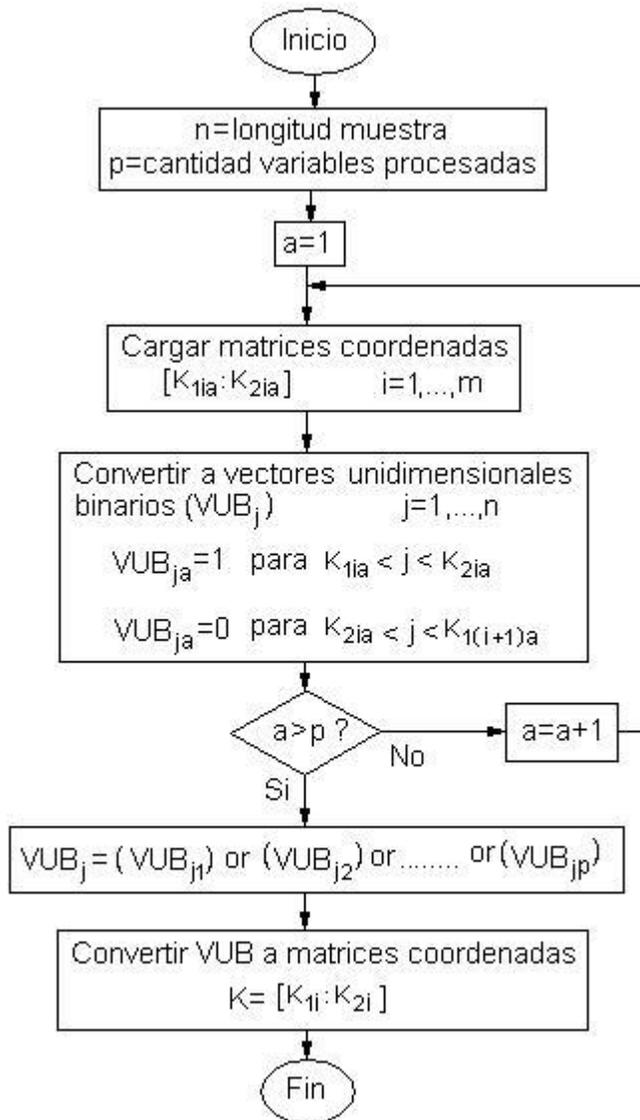
Anexos

Anexo 1. Diagrama de flujo del algoritmo para fraccionamiento basada en cálculo de desviación estándar (σ)





Anexo 2. Diagrama de Flujo del algoritmo de coincidencias multivariable



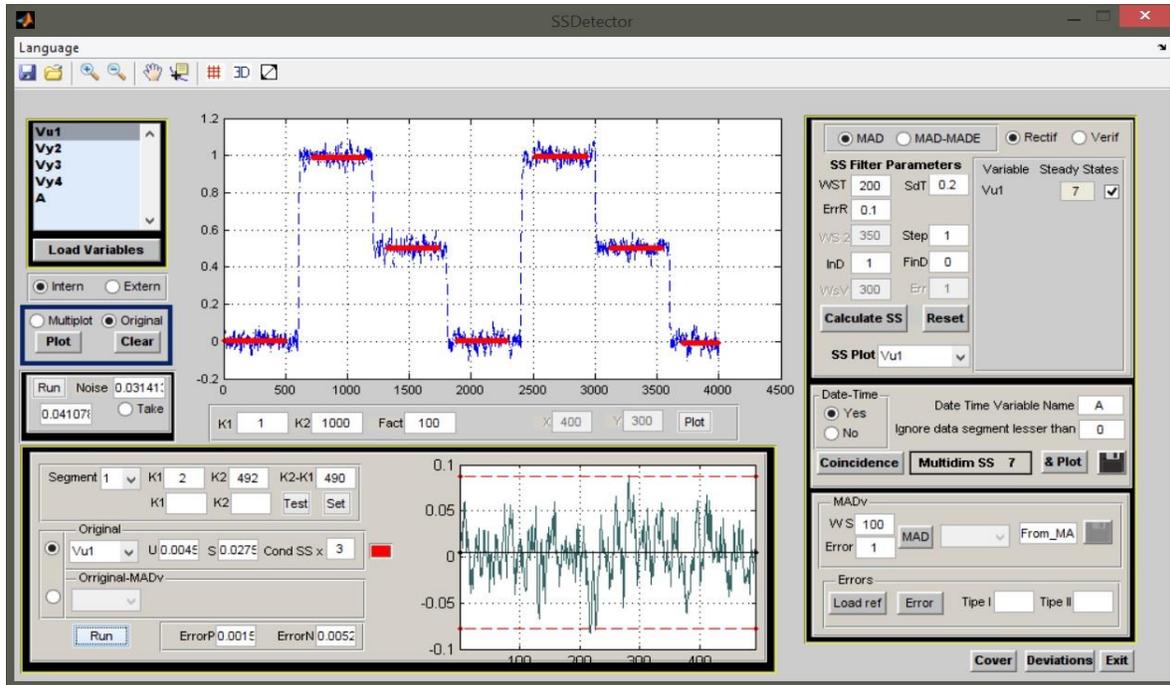


Anexo 3. Listado de variables usadas del generador de vapor CTE Felton.

| Nomenclatura | Etiqueta. | U/M | Significado |
|------------------|-----------|----------------------|---|
| N_P | LCM51TCA | [MW] | Potencia activa de la unidad. |
| \dot{V}_f | PE10F201 | [l/min] | Flujo volumétrico de combustible al generador de vapor. |
| \dot{V}_a | NG00F801 | [Nm ³ /s] | Flujo volumétrico de aire a los quemadores. |
| $\dot{m}_{w(b)}$ | RL50F801 | [Ton/h] | Flujo másico de agua al generador de vapor. |
| X_{VTF} | NG10G201 | % | Apertura de compuerta entrada de aire |
| X_{VRG} | NS10G201 | % | Apertura de compuerta entrada de recirculación de gases |
| P_e | NG10P201 | [kPa] | Presión de aire a la salida del FDF |
| P_h | NR00P203 | [kPa] | Presión de gases en el horno. |
| P_{ec} | NR10P205 | [kPa] | Presión de gases a la entrada del RRH |
| P_{sc} | NR10P206 | [Pa] | Presión de gases a la salida del RRH |
| $T_{w(eco)}$ | RL50T202 | [°C] | Temperatura a la entrada del economizador. |
| T_{sSH4} | NA41T801 | [°C] | Temperatura del vapor que deja SH4 |
| T_{sRH2} | NE21T801 | [°C] | Temperatura del vapor que deja RH2 |
| T_g | NR10T207 | [°C] | Temperatura en la chimenea de los gases de salida. |
| T_{gee} | NR10T210 | [°C] | Temperatura de los gases ante del ECO |
| T_{gec} | NR20T208 | [°C] | Temperatura de los gases a la entrada del RRH |
| T_{aec} | NG10T211 | [°C] | Temperatura del aire a la entrada del RRH. |
| T_{asc} | NG10T213 | [°C] | Temperatura del aire a la salida del RRH |



Anexo 4. Interfaz gráfica del módulo SSDtector



Anexo 5. Interfaz gráfica del módulo Deval

