Universidad de Cienfuegos Carlos Rafael Rodríguez

Facultad de Ingeniería

Departamento de Matemática



TESIS EN OPCIÓN AL GRADO ACADÉMICO DE MÁSTER EN CIENCIAS.

Modelación Matemática del Tope de la Columna Desbutanizadora de la Refinería de Cienfuegos.

Autor: Ing. Janier Delgado Puerto

Tutor: Prof. Nicolás González Suárez, Dr. C.

Cienfuegos 2014 "Año 56 de la Revolución"

DEDICATORIA

Dedicada al esfuerzo de toda una carrera, la carrera de la vida del autor.

AGRADECIMIENTOS

A todas aquellas personas bien intencionadas que han colaborado con el logro de estos resultados.

RESUMEN

La presente investigación se realizó en la Refinería de Petróleo de Cienfuegos, con el objetivo fundamental de obtener un modelo matemático que permita simular el proceso operacional del sistema del tope de la columna desbutanizadora del Bloque 400 de la Refinería de Petróleos de Cienfuegos. Entre las técnicas y herramientas empleadas se encuentran: corridas experimentales, entrevistas, observaciones directas, revisión de documentos. Desde el punto de vista estadístico matemático se identifican valores atípicos, análisis de tendencias lineales, se utilizan modelos paramétricos lineales, obtenidos por medio de la regresión lineal. El procesamiento de los datos se realiza a través del paquete de programas estadísticos IBM SPSS Statistics Versión 20.0 y el MATLABTM R2011b.

Como resultado se obtiene un modelo matemático de estructura lineal que permite simular el proceso operacional del sistema del tope de la columna desbutanizadora T-401/1 alrededor del punto de operación para el cual fue estimado, lo cual facilita el desarrollo e implementación de un Método de Control Inteligente.

Por último se exponen las conclusiones y recomendaciones que derivan del estudio y permiten definir una vía de seguimiento adecuada para dar continuidad a la temática desarrollada en la investigación.

ÍNDICE

RESUMEN		4
INTRODUC	CCIÓN	8
CAPÍTULO	1. Modelos y métodos aplicables a la modelación matemática	13
1.1. El j	proceso de modelación matemática	13
1.2. Tip	oos de modelos de sistemas	13
1.3. Mc	odelos de sistemas lineales e invariantes en el tiempo	14
1.3.1	Modelos de funciones de transferencia	15
1.3.2	Modelos en el espacio de estados.	19
1.3.3 estados	Correlación entre funciones transferenciales y ecuaciones en el espacio de	e 22
1.4. Mé	todos para obtener modelos de sistemas lineales e invariantes en el tiempo	23
1.4.1	Algoritmos de modelación no-paramétrica (Análisis de correlación):	23
1.4.2	Algoritmos no iterativos (mínimos cuadrados lineales):	25
1.4.3	Algoritmos no iterativos (variables intrumentales):	26
1.4.4	Algoritmos no iterativos (método de subespacios)	30
1.4.5	Algoritmos de estimación del error de predicción	34
1.5. Mc	odelos de sistemas no-lineales:	39
1.5.1	Modelos polinomiales no-lineales:	40
1.5.2	Modelos Hammerstein & Weiner:	41
1.5.3	Estimadores para modelos no-lineales:	42
1.6. Co	nclusiones del capítulo.	44
CAPÍTULO	2. Procedimiento para modelar sistemas en procesos industriales a	gran
escala		45
2.1 De	scripción del proceso	45

2.2	Pro	cedimiento para la modelación matemática de sistemas en procesos		
industriales a gran escala47				
2.2.	.1	Fase I - Planificación del experimento.	.47	
2.2.	.2	Fase II: Pre-procesamiento de los datos.	.51	
2.2.	.3	Fase III: Selección de un set de modelos:	.56	
2.2.	.4	Fase IV: Obtención del modelo.	.58	
2.2.	.5	Fase V: Validación del modelo	.61	
2.3	Cor	clusiones del capítulo.	.69	
CAPÍTU	ЛO	3. Implementación del procedimiento para modelar el sistema del tope	e la	
columna	desl	butanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos	.70	
3.1	Fas	e I: Planificación del experimento	.70	
3.1.	.1	Selección de las señales a medir.	.71	
3.1.	.2	Selección del intervalo de muestreo.	.71	
3.1.	.3	Selección de las señales a manipular.	.71	
3.2	Fas	e II: Pre-procesamiento de los datos	.73	
3.2.	.1	Pre-procesamiento de los datos para modelos lineales de sistemas	.73	
3.2.	.2	Pre-procesamiento de los datos para modelos no-lineales de sistemas	.74	
3.3	Fas	e III: Selección de un set de modelos.	.74	
3.4	Fas	e IV: Obtención del modelo	.76	
3.4.	.1	Identificación del tiempo de retardo del sistema.	.76	
3.4.	.2	Obtención de modelos polinomiales lineales a partir del método de variable	es	
instrumentales (IV):7		.78		
3.4.3 Obtención de modelos polinomiales lineales a partir del		Obtención de modelos polinomiales lineales a partir del método de		
minimización del error de predicción79				

3.4	4.4 Obtención de modelos lineales en el espacio de estados a partir del métod	lo de	
min	nimización del error de predicción.	79	
3.4	4.5 Resumen de los modelos obtenidos.	79	
3.5	Fase V: Validación de los modelos	81	
3.5	5.1 Selección del mejor modelo	82	
3.6	Eficacia y eficiencia del modelo Box-Jenkins obtenido.	83	
3.7	Conclusiones del capítulo.	85	
CONCL	LUSIONES	86	
RECOM	MENDACIONES	87	
BIBLIOGRAFÍA			
ANEXOS			

INTRODUCCIÓN

El control automático surge para liberar al hombre de tareas repetitivas, donde la complejidad del sistema a controlar es elevada o la operación es riesgosa, en la práctica puede haber una gran cantidad de motivos por los cuales se opta por el control automático (Paraskevopoulos., 2002).

El presente trabajo pretende facilitar elementos matemáticos necesarios para implementar una técnica de control avanzado en el proceso de fraccionamiento de gases, específicamente en la columna desbutanizadora de la Sección 400, perteneciente a la Refinería de Petróleos de Cienfuegos. En dicha planta es de vital importancia el control de la composición con que sale el producto del tope, denominado Gas Licuado del Petróleo (GLP). Este producto está compuesto, aproximadamente, de un 70% de Propano (C3) y un 30% de butano (C4), con pequeñas concentraciones de Pentano (C5) que pueden fluctuar entre un rango de 0% a 2.5% de la mezcla, considerado como impureza. La concentración permitida de estas impurezas depende de las especificaciones de composición de los clientes; que pueden variar desde un requerimiento de 0% para las empresas niquelíferas de Moa, a un requerimiento que oscila entre un 1.4 y 1.8% para las empresas de Cuba Petróleos (CUPET). Actualmente el control de la composición de la mezcla de GLP que sale de la columna se realiza de forma manual, controlando fundamentalmente la temperatura del tope. La composición que debe tener el producto al salir por el tope es inferida según la experiencia de los propios operadores, en dependencia de los valores de las demás variables que intervienen en este proceso. Estos, además, conocen la composición de la mezcla por análisis que se le realizan en el laboratorio a muestras puntuales tomadas en el parque industrial, con frecuencia de una vez cada 12 horas; es decir, una vez por turno. Se cuenta con un potente cromatógrafo en línea (on-line) capaz de brindar la composición del GLP (o sea, de: C3, C4 y C5, además de otras variables presentes en la composición), pero ubicado aguas abajo de varios equipos que intervienen en el proceso después que el producto ha abandonado el tope de la columna.

El control de temperatura implementado en el tope es un lazo de control en cascada, lo cual se logra manipulando el caudal de la línea de reflujo hacia el tope. De esta forma se garantiza disminuir la temperatura del tope aumentando el flujo, y aumentar la temperatura

disminuyendo el reflujo. En dependencia del punto de ajuste deseado (set-point) de temperatura (inferido y establecido por el operador), el Controlador Indicador de Temperatura (TIC) genera set-point dinámicos al Controlador Indicador de Flujo (FIC) para controlar el caudal de la línea de reflujo al tope.

Investigaciones previas realizadas por Alonso Díaz, Herrera Fernández & Madrazo Valladares (2012); han evidenciado que el TIC actualmente implementado, el cual responde a una estrategia clásica de control Proporcional-Integral-Derivativo (PID), no muestra un buen desempeño ante determinadas variaciones requeridas del punto de operación, debido a que se evidencian tiempos de establecimientos altos que son alcanzados luego de tener oscilaciones relativamente grandes, lo cual no es beneficioso para el proceso en cuestión. Por estas razones, en repetida ocasiones, cuando se necesitan determinados cambios en el set-point de temperatura, el operador rompe la cascada de este lazo de control, dejando solamente al FIC (lazo interno) en automático, en aras de ganar en el tiempo de establecimiento y evitar las "dañinas oscilaciones" alrededor del valor deseado durante esta etapa. Todo esto es posible gracias a la experiencia de los operadores sobre el comportamiento de la planta, lo cual les permite trabajar directamente sobre el set-point del FIC, garantizando un flujo que dé como resultado el valor de temperatura deseado en el tope de la columna.

Tomando en consideración los elementos antes expuestos, y el criterio de autores como Preeti & Singh Beniwal (2012), sobre las ventajas demostradas por las técnicas de control avanzado con respecto a los controladores clásicos (PID) en aplicaciones como la actual, donde el uso de una estrategia de control basada en PID no tienen un buen desempeño, influenciado por las características del proceso, el cuál presenta altas no-linealidades y puede ser modelado matemáticamente en un sistema de orden superior, específicamente de tercer orden, con un tiempo muerto relativamente grande; en la organización objeto de análisis se trabaja en el desarrollo de una estrategia de control avanzado para mejorar el control de temperatura en el tope de la columna fraccionadora de gases. Dicha estrategia pertenece al campo de la inteligencia artificial, un controlador con lógica difusa, capaz de sintetizar las experiencias y habilidades operacionales adquiridas y desarrolladas por los operadores, para ser ejecutadas de forma automática. La implementación de este control difuso puede ser utilizada para ajustar la temperatura del tope de la columna de fraccionamiento de gases. El ajuste se logra mediante la generación de set-point dinámicos al Controlador Indicador de Flujo. En opinión de Medeiros & Meneghetti Ugulino (2007); en la industria este problema es usualmente resuelto por estrategias de control predictivo, sin embargo, una estrategia de control inteligente puede ser aplicada para este fin. La eficiencia de los controladores difusos ha sido corroborada en la generación de set-point y en la emulación de maniobras operacionales realizadas por operadores expertos en lazos de control automatizados. Investigaciones anteriores han dejado evidencia de esto, ver en (Al-Odienat. & Al-Lawama., 2008); (Rao. & Reddy., 2010); (Aslam. & Kaur., 2011); (Manjunath. & Raman., 2011); (Preeti. & Singh Beniwal., 2012).

A pesar de que para el método de control difuso que se pretende Fuyin & Weifeng (2009) plantean que no es indispensable contar con el modelo matemático del sistema, como lo es en el caso del diseño del control adaptativo o robusto según Paraskevopoulos (2002); o en un control avanzado como el Control Predictivo basado en Modelos, (MPC siglas en inglés) según Lawrynczuk & Tatjewski (2010); la idea de buscar un modelo matemático que represente y responda similar al sistema a controlar en aras de implementar esta nueva estrategia de control, es importante, debido a que posibilita el uso de herramientas de simulación, para conocer su desempeño y compararlo con los controladores que actualmente se utilizan en la planta. Además, dado el principio de la concepción del sistema de control que se propone, basado en la representación del conocimiento y maniobra de operadores expertos; dicha simulación sirve como una demostración que facilite, desde un punto de vista técnico-social, todo el proceso de aprobación de manejo al cambio, así como la aceptación en las plantas de producción y su posterior implementación; fundamentalmente en una industria petrolera como la de Cienfuegos, donde no existe experiencia de trabajo con este tipo de técnicas de control avanzado.

Basado en los aspectos abordados se plantea el siguiente problema científico:

Necesidad de un modelo matemático que permita simular el proceso operacional del sistema del tope de la columna desbutanizadora T-401/1 del Bloque 400 de la Refinería de Petróleos de Cienfuegos.

Hipótesis:

Si se obtiene un modelo matemático del sistema del tope de la columna desbutanizadora T-401/1 del Bloque 400 de la Refinería de Petróleos de Cienfuegos, es posible simular su proceso operacional.

Objeto de investigación: El sistema del tope de la columna desbutanizadora del Bloque 400 de la Refinería de Petróleos de Cienfuegos.

Campo de acción: Modelación matemática del sistema del tope de la columna desbutanizadora del Bloque 400 de la Refinería de Petróleos de Cienfuegos.

El Objetivo General de la investigación es:

Obtener un modelo matemático que permita simular el proceso operacional del sistema del tope de la columna desbutanizadora del Bloque 400 de la Refinería de Petróleos de Cienfuegos.

Para el cumplimiento de este objetivo es necesario llevar a cabo los siguientes **Objetivos Específicos:**

1. Analizar modelos y métodos aplicables a la modelación matemática de sistemas en la industria de procesos.

2. Diseñar un procedimiento que posibilite modelar matemáticamente el proceso operacional del sistema del tope la columna desbutanizadora del Bloque 400 de la Refinería de Petróleos de Cienfuegos.

3. Obtener un modelo matemático para el sistema del tope de la columna desbutanizadora del Bloque 400, que sea capaz de describir con eficiencia y eficacia el proceso operacional de dicho sistema.

Tipo de investigación:

La investigación que se pretende es de tipo descriptiva - correlacionar, ya que además de medir y recolectar información sobre las variables del proceso, pretende saber cómo se puede comportar la variable de salida (temperatura del tope de la columna) conociendo el comportamiento de otras variables relacionadas medidas. La utilidad principal de este proyecto es proporcionar una herramienta matemática (modelo) para simular, con la mayor precisión posible, el comportamiento del sistema.

Las tareas de la investigación se pueden resumir como sigue:

- Valoración crítica sobre las herramientas matemáticas aplicables a la modelación de sistemas tecnológicos de la industria de procesos.
- Excitación de las variables del proceso en condiciones de operación.
- Creación de la base de datos de acuerdo al propósito de la investigación.
- Selección de los métodos de obtención de modelos.
- Obtención de modelos matemáticos del proceso tecnológico del sistema del tope de la columna desbutanizadora T-400/1 del Bloque 400.
- Validación de los modelos obtenidos.

Metodología aplicada en la investigación

• Analítico-sintético

Para procesar la información relacionada con los aspectos abordados en la investigación, así como arribar a conclusiones parciales y generales.

• Análisis hipotético-deductivo

A partir de la hipótesis inicial y muestra utilizada, se estableció un modelo para simular el proceso operacional del tope de la columna desbutanizadora T-401/1 en la Refinería de Petróleos de Cienfuegos.

Modelación matemática

Aplicación de modelos paramétricos, obtenidos por medio de la regresión lineal, así como se identifican valores atípicos y tendencias. El empleo de paquetes de programas como IBM SPSS Statistics Versión 20.0 y el MATLABTM R2011b.

Organización del informe

El trabajo quedó estructurado de la siguiente forma:

En el Capítulo I se abordan diferentes modelos matemáticos, así como los métodos utilizados en su obtención, teniendo como soporte la literatura científica que trata la problemática desde el punto de vista teórico-práctico.

En el Capítulo II se describe el proceso operacional objeto de estudio y se diseña un procedimiento para la modelación matemática de sistemas de procesos industriales a gran escala que operan en configuración de lazo cerrado.

En el Capítulo III se presentan los resultados obtenidos en el proceso de modelación matemática del sistema en estudio, trabajando sobre la base de un conjunto de variables de entradas y salidas, y aplicando el procedimiento desarrollado en el Capítulo II.

CAPÍTULO 1. Modelos y métodos aplicables a la modelación matemática. Introducción.

En opinión de algunos autores Paraskevopoulos (2002); Wigren & J. Schoukens (2013); un concepto fundamental en la ciencia y la tecnología es el de la modelación matemática. Un modelo matemático describe el conocimiento que se tiene sobre los procesos o sistemas, dichos sistemas pueden ser físicos, sociológicos o hasta económicos. Según Zill (1997) a la descripción matemática de un sistema o fenómeno se llama modelo matemático. La determinación de un modelo matemático de un proceso o sistema es conocido como identificación de sistemas (R. Pintelon. & J. Schoukens., 2001); (Paraskevopoulos., 2002). Tomando como base tales conceptos, durante el desarrollo de esta investigación se emplea el término de modelación matemática en lugar de identificación de sistemas para un mismo fin.

1.1. El proceso de modelación matemática.

De acuerdo al criterio de Ljung (1999); la modelación matemática por lo general es una tarea compleja y requiere de muchos tipos de habilidades; para ello no es suficiente conocer de un campo de aplicación específico como el automovilismo, la electrónica y la química, por mencionar algunos; sino que se requiere además de dominar técnicas de medición, teoría estadística y métodos numéricos; evidenciando ser un campo multidisciplinario. Para muchos autores, por ejemplo, Zill (1997); Ljung (1999) y R. Pintelon & J. Schoukens (2001) y Liu & Gao (2012), la modelación matemática es un proceso cíclico lógico donde se destacan etapas claves como la planificación del experimento, la recolección y procesamiento de los datos, la selección de un set de modelos, su obtención mediante el método seleccionado y la validación del mismo, la cual puede ser una última etapa una vez que haya sido probado que el modelo realmente es bueno para el fin que ha sido previsto. De lo contrario, se deben retomar algunas de las etapas previas en busca de mejorar el proceso de modelación.

1.2. Tipos de modelos de sistemas.

Ciertamente los modelos de sistemas pueden ser de varias formas y con determinado grado de formalismo matemático. La intención del uso determina el grado de sofisticación que es requerido para hacer el modelo útil Ljung (1999). En la vida diaria muchos sistemas son tratados con el uso de modelos mentales, los cuales no envuelven ningún formalismo

matemático, un ejemplo de ellos puede ser conducir un auto, en el cual el conductor como un reflejo condicionado conoce que girar el volante hacia la izquierda induce un giro del auto hacia la izquierda. Para otros sistemas es apropiado describirlos usando tablas numéricas y gráficos, estas descripciones son conocidas como modelos gráficos y son aplicables a sistemas lineales. Tales sistemas lineales pueden ser únicamente descritos por sus respuestas al impulso o al paso o por sus funciones de frecuencia (Ljung., 1999). En aplicaciones más avanzadas se hace necesario el uso de modelos que describan las relaciones entre los sistemas de variables en términos de expresiones matemáticas, como por ejemplo, ecuaciones diferenciales. En este caso son llamados modelos matemáticos; los que pueden ser adicionalmente caracterizados por un número de adjetivos tales como: (de tiempo continuo o discreto, determinísticos o estocásticos, lineales o no-lineales, etc.); en dependencia del tipo de ecuación diferencial usado. El uso de modelos matemáticos es inherente en el campo de la ingeniería y la física, de hecho, la mayor parte del campo de la ingeniería trata cómo hacer buenos diseños basados en modelos matemáticos, ya que los mismos son instrumentos para la simulación y la predicción (Ljung., 1999). Por tal motivo, en las siguientes secciones del capítulo el objetivo principal está centrado en describir, de una breve manera, las características básicas de modelos matemáticos considerados importantes para el desarrollo del la presente investigación.

1.3. Modelos de sistemas lineales e invariantes en el tiempo.

Un modelo completo lineal e invariante en el tiempo, de tiempo discreto, como tal es dado por:

$$y(t) = G(q)u(t) + H(q)e(t)$$
 (1.1)

 $f_e(x)$ la función de densidad probabilística (PDF siglas en inglés) de e con:

$$G(q) = \sum_{k=1}^{\infty} g(k)q^{-k}, \qquad \qquad H(q) = 1 + \sum_{k=1}^{\infty} h(k)q^{-k}, \qquad (1.2)$$

Donde:

- y(t) variable de salida en el tiempo t.
- u(t) variable de entrada en el tiempo t.
- e(t) disturbios o ruidos en el tiempo t (secuencia de variables aleatorias).
- G(q) función de transferencia de la entrada (u) a la salida (y).
- H(q) análoga a G pero para función de transferencia de (e) a (y).

Muy frecuentemente la PDF de $f_e(x)$ no está especificada como una función, pero si descrita en términos de unas pocas características numéricas, típicamente los momentos de primer y segundo orden (μ_1 y μ_2), que representan: μ_1 la media (μ_e) y μ_2 la varianza (σ_e^2) del error respectivamente, los mismos son:

$$\mu_1 = E(e(t)) = \int_{-\infty}^{\infty} x f_e(x) dx \qquad \qquad \mu_2 = E(e^2(t)) = \int_{-\infty}^{\infty} x^2 f_e(x) dx \qquad (1.3)$$

Es además común asumir que e(t) es Gaussiano, resultando ser completamente especificado por la PDF (Ljung., 1999) (R. Pintelon. & J. Schoukens., 2001). Luego E(e(t)) = 0 y $E(e^2(t)) = \lambda$, o sea, e(t) está normalmente distribuida con media cero y varianza λ ; $e(t) \sim N(0, \lambda)$.

En muchos de los casos no es práctico hacer las especificaciones expuestas en (1.2) por enumeración de las secuencias infinitas $\{g(k)\}, \{h(k)\}$ junto con la función $f_e(x)$. Según Ljung (1999), en su lugar se opta por trabajar con estructuras que permitan las especificaciones de G y H en un número finito de valores numéricos, como lo son los casos típicos de las funciones de transferencia rotacional y las descripciones en espacio de estados dimensionales finitas; dándose lugar a los modelos de funciones de transferencia y a los modelos en espacio de estados.

Los modelos más usados de estos tipos son obtenidos de datos de entadas y salidas de los sistemas que describen, aplicando diseños cuidadosos e implementación de señales de prueba, resultando entonces modelos obtenidos en base a datos de pruebas. Como consecuencia los modelos de entradas y salidas tienen algunas desventajas, así lo evidencian autores tales como Ljung (1999); R. Pintelon & J. Schoukens (2001); Weiyong & Dexian (2008); Pitta & Odloak (2012). Primero, es difícil de llevar a cabo la aplicación de señales de prueba a procesos con altos requerimientos de seguridad, como puede ser la industria de procesos petroquímicos. Segundo, para procesos no-lineales, no se puede esperar que el modelo estimado sea capaz de predecir con exactitud la conducta del proceso más allá del rango de los datos de prueba.

1.3.1 Modelos de funciones de transferencia.

Para Ljung (1999), quizás la forma más inmediata de parametrización para G y H sea representarlos como funciones rotacionales, obteniéndose así los modelos de funciones de transferencia también conocidos como modelos polinomiales. La estructura general de dichos modelos de funciones de transferencia para sistemas que se analicen en tiempo

discreto con una entrada y una salida, se convierte en una ecuación polinomial de la siguiente forma (Ljung., 2007b):

$$A(q)y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t)$$
(1.4)

Con:

$$A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + \dots + a_{na} q^{-na}$$

$$B(q) = b_1 q^{-1} + b_2 q^{-2} + \dots + b_{nb} q^{-nb}$$

$$F(q) = 1 + f_1 q^{-1} + f_2 q^{-2} + \dots + f_{nf} q^{-nf}$$

$$C(q) = 1 + c_1 q^{-1} + c_2 q^{-2} + \dots + c_{nc} q^{-nd}$$

$$D(q) = 1 + d_1 q^{-1} + d_2 q^{-2} + \dots + d_{nd} q^{-nd}$$

(1.5)

$$\theta = \begin{bmatrix} a_1 \dots a_{na} \ b_1 \dots b_{nb} \ f_1 \dots f_{nf} \ c_1 \dots c_{nc} \ d_1 \dots d_{nd} \end{bmatrix}^T$$
(1.6)
Donde:

A, B, C, D y F: son polinomios que contienen el operador de tiempo q^{-1} .

 θ : vector de parámetros del modelo.¹

 $a_1, \ldots, a_{na}, b_1, \ldots, b_{nb}, f_1, \ldots, f_{nf}, c_1, \ldots, c_{nc}, d_1, \ldots, d_{nd}$: parámetros del modelo.

 q^{-1} : operador de tiempo de retardo, compactamente representa una ecuación en diferencias, definido por la siguiente operación:

$$q^{-1}u(t) = u(t-1).$$
(1.7)

La **Tabla 1.1** resume las estructuras comunes de modelos polinomiales lineales, representando cada uno de ellos casos particulares de la ecuación polinomial general (1.4). Estos modelos se diferencian en cómo algunos de esos polinomios son incluidos en la estructura y, producto de sus diferencias, proporcionan niveles de variación en la flexibilidad para la modelación de las dinámicas y las características del ruido de los sistemas.

¹ Es retomado y explicado en próximas secciones.

Nombre de la	Polinomios usados en forma de	Modelo de ruido
estructura del	tiempo discreto	
modelo.		
Auto Regresivo		El modelo de ruido es $\frac{1}{A}$ y el ruido está
con entrada		acoplado a la dinámica del modelo.
exógena (ARX).	A(q)y(t) = B(q)u(t) + e(t)	ARX no hace un modelo de ruido y de
		la dinámica independientemente. Es
		usado para obtener un modelo simple
		en verdad.
Media movida		Proporcionado más flexibilidad para la
auto regresiva		modelación del ruido al usar los
con entrada		parámetros de C (una Media Movida
exógena		de ruido blanco). Usado cuando las
(ARMAX)	$A(q)y(\iota) = B(q)u(\iota) + U(q)e(\iota)$	perturbaciones dominantes tienen
		lugar en las entradas del sistema. Tales
		perturbaciones son llamadas
		perturbaciones de carga.
Box – Jenkins		Provee una parametrización
(BJ)		completamente independiente de la
		dinámica del proceso y el ruido, con
	B(a) = C(a)	funciones polinomiales rotacionales.
	$y(t) = \frac{D(q)}{F(q)}u(t) + \frac{U(q)}{D(q)}e(t)$	Es usado cuando el ruido no tiene
	- (4)	lugar en las entradas del sistema, pero
		es primario a las perturbaciones
		medibles. Provee flexibilidad
		adicional para la modelación del ruido.
Error de salida	B(a)	Usado cuando se desea parametrizar la
(OE)	$y(t) = \frac{F(q)}{F(q)}u(t) + e(t)$	dinámica del proceso pero no se desea
		obtener el modelo de ruido.

Tabla 1.1: Estructuras de modelos polinomiales lineales. Fuente: (Ljung., 2007b).

Caso multivariable.

Hasta ahora, se ha hecho alusión a sistemas caracterizados solo con una entrada y una salida (SISO: siglas en inglés). Los sistemas de la vida real generalmente son multivariables, pudiendo contener múltiples variables de entrada o múltiples variables de salida, o ambas (MIMO: siglas en inglés).

Considerando el caso donde la entrada u(t) es un vector *m*-dimensional y la salida y(t) sea un vector *p*-dimensional, la mayoría de las ideas que han sido descritas en esta sección para los sistemas SISO, según Ljung (1999), tienen un camino directo hacia el homólogo multivariable.

Para los casos multivariable análogo a (1.5), por ejemplo, se introducen los polinomios

$$A(q) = I + A_1 q^{-1} + A_2 q^{-2} + \dots + A_{na} q^{-na}$$

$$B(q) = B_1 q^{-1} + B_2 q^{-2} + \dots + B_{nb} q^{-nb}$$

$$F(q) = I + F_1 q^{-1} + F_2 q^{-2} + \dots + F_{nf} q^{-nf}$$

$$C(q) = I + C_1 q^{-1} + C_2 q^{-2} + \dots + C_{nc} q^{-nc}$$

$$D(q) = I + D_1 q^{-1} + D_2 q^{-2} + \dots + D_{nd} q^{-nd}$$

(1.8)

Siendo estos ahora matrices polinomiales que contiene el operador de retardo q^{-1} , significando que A(q), B(q), F(q) C(q), D(q) son matrices cuyas entradas y salidas respectivamente son polinomios. Por su parte, el modelo completo lineal e invariante en el tiempo, de tiempo discreto, está dado por:

$$A(q)y(t) = \sum_{i=1}^{nu} \frac{B_i(q)}{F_i(q)} u_i(t - nk_i) + \frac{C(q)}{D(q)} e(t)$$
(1.9)

<u>Donde:</u> u_i : la *i*-ésima entrada del sistema.

nu: el número total de entradas.

 nk_i : el retardo de la *i*-ésima entada que caracteriza el tiempo de retardo de la respuesta.

También inmediatamente análogo a (1.6) puede ser definida entonces la matriz siguiente:

$$\theta = \begin{bmatrix} A_1 \dots A_{na} & B_1 \dots B_{nb} & F_1 \dots F_{nf} & C_1 \dots & C_{nc} & D_1 \dots & D_{nd} \end{bmatrix}^T$$
(1.10)
Por ejemplo, para un sistema con *nu* entardas y *ny* salidas, *A(q)* es una matriz de dimensión *ny* **x** *ny* y puede ser representada como un polinomio en el operador de retardo q^{-1} :

CAPÍTULO 1: Modelos y métodos aplicables a la modelación matemática.

$$A(q) = I_{ny} + A_1 q^{-1} + A_2 q^{-2} + \dots + A_{na} q^{-na}$$
(1.11)

Además puede ser representada como una matriz:

$$\begin{pmatrix} a_{11}(q) & a_{12}(q) & \cdots & a_{1ny}(q) \\ a_{21}(q) & a_{22}(q) & \cdots & a_{2ny}(q) \\ \cdots & \cdots & \cdots & \cdots \\ a_{ny1}(q) & a_{ny2}(q) & \cdots & a_{nyny}(q) \end{pmatrix}$$
(1.12)

Donde los elementos de la matriz α_{kj} son polinomiales en el operador de retardo q^{-1} :

$$\alpha_{kj}(q) = \delta_{kj} + \alpha_{kj}^1 q^{-1} + \dots + \alpha_{kj}^{na_{kj}} q^{-na_{kj}}$$
(1.13)

 α_{kj} representa la delta Kronecher, la cual es igual a uno (1) para k = j e igual a cero (0) para $k \neq j$. Estos polinomios describen como los valores anteriores de las *j*-ésima salidas son afectados por la *k*-ésima salida. La *i*-ésima fila de A(q) representa la contribución de los valores de las salidas pasadas para predecir el valor presente de la *i*-ésima salida (Ljung., 2007b).

Por su parte B(q) es una matriz ny x nu que puede ser representada de forma polinomial en el operador de retardo q^{-1} .

$$B(q) = B_0 + B_1 q^{-1} + B_2 q^{-2} + \dots + B_{nb} q^{-nb}$$
(1.14)

B(q) además puede ser representada como una matriz:

$$\begin{pmatrix} b_{11}(q) & b_{12}(q) & \cdots & b_{1nu}(q) \\ b_{21}(q) & b_{22}(q) & \cdots & b_{2nu}(q) \\ \cdots & \cdots & \cdots & \cdots \\ b_{ny1}(q) & b_{ny2}(q) & \cdots & b_{nynu}(q) \end{pmatrix}$$
(1.15)

Donde los elementos de la matriz b_{kj} son polinomiales en el operador de retardo q^{-1} .

$$b_{kj}(q) = \alpha_{kj}^1 q^{-nb_{kj}} + \dots + \alpha_{kj}^{nk_{kj}} q^{-nb_{kj}-nb_{kj}+1}$$
(1.16)

 nk_{kj} es el retardo de la *j*-ésima entrada en la *k*-ésima salida. B(q) representa la contribución de las entradas para la predección de todos los valores de las salidas (Ljung., 2007b).

1.3.2 Modelos en el espacio de estados.

Los modelos en el espacio de estados son aquellos que usan las variables de estado para describir un sistema por un set de ecuaciones diferenciales de primer orden, en vez de por una o más ecuaciones diferenciales de n-ésimo orden. Las variables de estado pueden ser reconstruidas a partir de datos medibles de entradas y salidas del sistema, pero no siempre

19

se les puede medir a sí mismas durante un experimento práctico (Paraskevopoulos., 2002); (Ljung., 2007b).

Ogata (1998) plantea que dependiendo de algunas circunstancias, por ejemplo, en problemas de control óptimo, es provechoso usar representaciones en el espacio de estados. Según este mismo autor, el concepto de estado de ningún modo está limitado a los sistemas físicos, pues se puede aplicar a diferentes sistemas como los biológicos, económicos y sociales.

Según Ogata (1998), el estado de un sistema dinámico es el conjunto más pequeño de variables (denominadas variables de estado) de modo que el conocimiento de estas variables en un tiempo $t = t_0$, junto con el conocimiento de la entrada para $t \ge t_0$, determina por completo el comportamiento del sistema para cualquier tiempo $t \ge t_0$. La cantidad de variables de estado necesarias para definir completamente la dinámica del sistema es igual a la cantidad de integradores que contiene el sistema. El conjunto de n variables de estado que se necesitan para describir por completo el comportamiento de un sistema determinado, se consideran las componentes de un vector, tal vector se denomina vector de estado; el cual determina de manera única el estado del sistema x(t) para cualquier tiempo $t \ge t_0$, una vez que se obtiene el estado en $t = t_0$ y se especifica la entrada u(t) para $t \ge t_0$. El espacio de las n dimensiones cuyos ejes de coordenadas están formados por *el eje* x_1 , *el eje* x_2 , ..., *el eje* x_n ; se denomina espacio de estados. Cualquier estado puede representarse mediante un punto en el espacio de estados (Ogata., 1998).

Para un sistema que contiene *n* integradores, *r* entradas $u_1(t), u_2(t), ..., u_r(t)$ y *m* salidas $y_1(t), y_2(t), ..., y_m(t)$; una vez definidas las salidas de los integradores como variables de estado $x_1(t), x_2(t), ..., x_n(t)$ el sistema se puede describir como:

$$\dot{x}_{1}(t) = f_{1}(x_{1}, x_{2}, \dots, x_{n}; u_{1}, u_{2}, \dots, u_{r}; t)$$

$$\dot{x}_{2}(t) = f_{2}(x_{1}, x_{2}, \dots, x_{n}; u_{1}, u_{2}, \dots, u_{r}; t)$$

$$\dot{x}_{n}(t) = f_{n}(x_{1}, x_{2}, \dots, x_{n}; u_{1}, u_{2}, \dots, u_{r}; t)$$
(1.17)

Y las salidas $y_1(t), y_2(t), ..., y_m(t)$ del sistema se obtienen mediante:

$$y_{1}(t) = g_{1}(x_{1}, x_{2}, ..., x_{n}; u_{1}, u_{2}, ..., u_{r}; t)$$

$$y_{2}(t) = g_{2}(x_{1}, x_{2}, ..., x_{n}; u_{1}, u_{2}, ..., u_{r}; t)$$

$$y_{m}(t) = g_{m}(x_{1}, x_{2}, ..., x_{n}; u_{1}, u_{2}, ..., u_{r}; t)$$
(1.18)

Luego definiendo:

$$\begin{aligned} x(t) &= \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} \qquad y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix} \qquad u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ y_m(t) \end{bmatrix} \\ f(x, u, t) &= \begin{bmatrix} f_1(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_r; t) \\ f_2(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_r; t) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_r; t) \\ \vdots \\ g_1(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_r; t) \\ \vdots \\ g_n(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_r; t) \end{bmatrix} \end{aligned}$$

Las ecuaciones (1.17) y (1.18) se convierten en:

$$\dot{x}(t) = f(x, u, t)$$
 (1.19)
 $y(t) = g(x, u, t)$ (1.20)

Donde la ecuación (1.19) es la ecuación de estado y la ecuación (1.20) es la ecuación de salida, siendo importante señalar que las funciones vectoriales f y g no involucren explícitamente el t para que el sistema sea invariante con el tiempo. Si se linealizan las ecuaciones (1.19) y (1.20) alrededor del estado de operación, se obtienen las ecuaciones de estado y de salida linealizadas:

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{1.21}$$

$$y(t) = Cx(t) + Du(t)$$
 (1.22)

Donde: A: matriz de estado.

- *B*: matriz de entrada.
- *C*: matriz de salida.
- D: matriz de transmisión directa.

Los modelos en el espacio de estados proveen un espectro de posibilidades de modelación, dentro de tal espectro tiene lugar la representación en tiempo discreto, la cual proporciona el mismo tipo de relación en diferencia lineal entre las entradas y salidas del sistema que un modelo polinomial lineal, pero es cambiado de forma tal que hay un solo retardo en la expresión (Ljung., 2007b). La estructura del modelo en el espacio de estados, de tiempo discreto que describe al ruido del sistema, escrita en la forma de innovación es:

$$x(kT + T) = Ax(kT) + Bu(kT) + Ke(kT)$$
(1.23)

$$y(kT) = Cx(kT) + Du(kT) + e(kT)$$

$$x(0) = x0$$

Donde: *T*: es el intervalo de muestreo.

u(kT): es la entrada en el instante de tiempo (kT).

y(kT): es la salida en el instante de tiempo (kT)

x0: es el estado inicial del sistema

K: es la ganancia de Kalman.

Entre las múltiples aplicaciones de los modelos en el espacio de estados se pueden citar:

La aplicación de Weiyong & Dexian (2008) para un Horno de Coquización Retardada, equipo fundamental para el mejoramiento de los productos pesados del petróleo en el proceso de coquización retardada, un importante proceder de las refinerías que mejora los beneficios económicos. El objetivo del modelo matemático, es ser utilizado en un Modelo de Control Predictivo para la temperatura de salida del horno, la cual tiene una influencia directa en el rendimiento del producto final. Basado en los conocimientos físicos del sistema logran un modelo no-lineal simplificado, a partir del cual, una vez que linealizan; obtienen el requerido modelo en el espacio de estados.

1.3.3 Correlación entre funciones transferenciales y ecuaciones en el espacio de estados.

Retomando el modelo descrito por la ecuación 1.1 donde *G* convierte las entradas u en salidas y y captura la dinámica del proceso, H es un operador que describe las propiedades del modelo de ruido de salida aditivo; tanto *G* como *H* son llamadas funciones de transferencia. También teniendo en cuenta la representación del modelo en el espacio de estados de tiempo discreto visto en las ecuaciones de la 1.23 - 1.25, la relación entre las matrices de uno y las funciones de transferencia del otro están dadas por las siguientes ecuaciones:

$$G(q) = C(qI_{nx} - A)^{-1}B + D$$
(1.26)

$$H(q) = C(qI_{nx} - A)^{-1}K + I_{ny}$$
(1.27)

<u>Donde:</u> I_{nx} : es la matriz identidad nx x nx

 I_{ny} es la matriz identidad nx x nx

ny es la dimensión de y en e.

(1.24)

(1.25)

Se puede observar en 1.26 que |qI - A| es igual al polinomio característico de G(q), lo que quiere decir que, los valores específicos de A son idénticos a los polos de G(q) (Ogata., 1998). El hecho poder obtener un modelo de estructura a partir de otro por medio de estas relaciones es de gran utilidad práctica según el criterio del autor de la presente investigación. En determinadas situaciones es más conveniente un tipo de representación en dependencia de la aplicación y el uso que se pretenda, entonces al disponer de un tipo de estructura se puede obtener otra sin necesidad de tener que iniciar un nuevo proceso de modelación matemática.

1.4. Métodos para obtener modelos de sistemas lineales e invariantes en el tiempo.

De los métodos disponibles en la bibliografía consultada, empleados con el objetivo de obtener modelos matemáticos de sistemas lineales e invariantes en el tiempo, la presente investigación se enfoca específicamente en los siguientes, los mismos son clasificados según Ljung (2007b) en:

- Algoritmos de modelación no-paramétrica (Análisis de correlación): proveen una estimación directa del estado transitorio y solo asume que el sistema es lineal y no requiere un modelo de estructura específico. Es una estimación no paramétrica de la respuesta al impulso y al paso de sistemas dinámicos.
- Algoritmos no iterativos para modelos en el espacio de estados y polinomiales: Incluye mínimos cuadrados lineales, variables instrumentales y métodos de subespacios.
- Algoritmos de predicción de error para modelos paramétricos.

1.4.1 Algoritmos de modelación no-paramétrica (Análisis de correlación):

En el análisis de correlación para obtener la respuesta al impulso y al paso de los sistemas dinámicos, se calcula un modelo de respuesta al impulso finito (FIR siglas en inglés) de un conjunto de datos dados.

Para un modelo como el analizado en la ecuación 1.1 solo con la diferencia de representar el término que representa el ruido adicionado en una forma más compacta, siendo la misma v(t); como se puede ver a continuación:

y(t) = G(q)u(t) + v(t)

La notación G(q)u(t) representa la operación:

$$G(q)u(t) = \sum_{k=1}^{\infty} g(k)u(t-k)$$
(1.28)

Además G(q) y el operador de retardo q están definidos como lo visto en las ecuaciones (1.2) y (1.7) respectivamente.

La respuesta al impulso de un sistema dinámico no es más que la señal de salida que resulta cuando la entrada es una señal impulso u(t) de la siguiente forma: u(t) = 1 cuando t = 0y u(t) = 0 cuando t > 0 para datos de tiempo discreto. Por su parte, la respuesta al paso es la señal de salida que resulta de una entrada paso con determidada amplitud, según sea de interés para el estudio que se realice. Por ejemplo, la respuesta al paso unitario es la señal de salida que resulta de una señal de entrada paso u(t) (con amplitud del paso igual a uno) de la siguiente forma: u(t < 0) = 0 y $u(t \ge 0) = 1$.

Para la respuesta, tanto al impulso como al paso, el algoritmo estima los coeficientes g de la respuesta según corresponda, ya sea para datos de salidas únicas o múltiples. De esta forma la respuesta es estimada como un modelo de elevado orden, medelo FIR no causal:

$$y(t) = g(-m)u(t+m) + \dots + g(-1)u(t+1) + g(0)u(t) + g(1)u(t-1) + \dots + g(n)u(t-n)$$
(1.29)

El término no causal en el contexto aplicado a los modelos de sistemas dinámicos significa que la salida del sistema en cierto instante de tiempo depende no solo de la entrada en ese instante de tiempo (Ljung., 1999)

El algoritmo de estimación pre-filtra los datos tal que las entradas son tan blancas como sea posible, luego calcula las correlaciones de los datos pre-filtrados para obtener los coeficientes FIR; usando el método de los mínimos cuadrados (LSM siglas en inglés) (Bjorklund., 2003); (Ljung., 2007b). Estos coeficientes g también son estimados para retrasos negativos, los cuales toman en cuenta cualquier efecto no causal de las entradas en las salidas. Efectos no causales puede resultar de las retroalimentaciones propias en los sistemas.

Para sistemas de múltiples entradas y múltiples salidas, la respuesta g_k es una matriz ny x nu, donde ny es el número de salidas y nu el número de entradas. El ij -ésimo elemento de la matriz respuesta al impulso describe la conducta de la i -ésima salida después de un impulso en la j -ésima entrada. Según Ljung (1999), las principales ventajas de este método son que no requiere de un procesamiento complejo de los datos, ni de ningún tipo de conocimiento previo de la planta, a excepción de que la misma tenga un comportamiento lineal. Su principal inconveniente es que no puede usarse directamente para la simulación (López Guillén., 1999).

1.4.2 Algoritmos no iterativos (mínimos cuadrados lineales):

Sin contar con información a priori exacta sobre la relación entre y(t) y $\varphi(t)$ y en su lugar tener una colección de observaciones previas Z^N (datos históricos) de valores relacionados de y y φ descritos dentro de un marco estocástico, donde es conveniente enumerar estos valores usando un argumento t:

$$y(t), \varphi(t), t = 1, ..., N$$
 (1.28)

Y $\varphi(t)$ vector de regresión en el tiempo t.

Planteándose el problema de buscar una función de los regresores $g(\varphi)$ tal que la diferencia $y - g(\varphi)$ se convierta tan pequeña como sea posible (por ejemplo: tal que $\hat{y} = g(\varphi)$ sea una buena predicción de y), se puede apostar por minimizar:

$$E[y - g(\varphi)]^2 \tag{1.29}$$

La ecuación 1.29 es conocida como el valor esperado de la varianza implicada en la estimación. Con datos históricos, como a los referidos anteriormente, se puede remplazar la ecuación 1.29 por la varianza muestral:

$$\frac{1}{N} \sum_{t=1}^{N} [y(t) - g(\varphi(t))]^2$$
(1.30)

Tratando entonces de ajustar y a una cambinación lineal de φ_i

$$g(\varphi) = \theta_1 \varphi_1 + \theta_2 \varphi_2 + \dots + \theta_d \varphi_d$$
(1.31)
Con el vector $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix}$

Entonces 1.31 su puede escribir como:

$$g(\varphi) = \varphi^T \theta \tag{1.32}$$

De esta forma 1.30 representándose por $V_N(\theta)$ (función criterio a ser minimizada) es:

$$V_N(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^{N} [y(t) - \varphi^T(t)\theta]^2$$
(1.33)

Donde una apropiada selección de θ es la minimización del argumento de 1.32. Esto es el criterio de los mínimos cuadrados para una regresión lineal, desarrollado de la parametrización lineal y el criterio cuadrático; con la única característica de que es una función cuadrática en θ . Sin embargo, según Ljung (1999), esta puede ser minimizada analíticamente, la cual da como resultado, proporcionando que la inversa indicada exista:

$$\hat{\theta}_N^{LS} = \arg\min V_N(\theta, Z^N) = \left[\frac{1}{N} \sum_{t=1}^N \varphi(t) \varphi^T(t)\right]^{-1} \frac{1}{N} \sum_{t=1}^N \varphi(t) y(t)$$
(1.34)

La conocida estimación de mínimos cuadrados LSE (siglas en inglés) y $\hat{\theta}_N^{LS}$, es simplemente el valor que toma la mejor conformación del estimador cuando es aplicado a datos históricos.

Caso multivariable

En el caso de que se use una parametrización similar a 1.10, con θ como una matriz y:

$$y(t \mid \theta) = \theta^T \varphi(t) \tag{1.35}$$

El criterio de mínimos cuadrados se convierte en:

$$V_N(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^N ||y(t) - \theta^T \varphi(t)||^2$$
(1.36)

Con el estimador:

$$\hat{\theta}_{N}^{LS} = \left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^{T}(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y^{T}(t)$$
(1.37)

1.4.3 Algoritmos no iterativos (variables intrumentales):

Idealmente el error de predicción $\varepsilon(t,\theta)$ para un buen modelo debe ser independiente de datos antiguos Z^{t-1} , pues esta es una condición inherente para modelos probabilísticos. Otra forma y aún más pragmática de ver esta condición es que si $\varepsilon(t,\theta)$ está correlacionado con Z^{t-1} entonces había más información disponible en Z^{t-1} sobre y(t) que la explicada por $\hat{y}(t | \theta)$, por lo que el predictor resulta no ser ideal. Un test para probar si $\varepsilon(t,\theta)$ es independiente de un set de datos completo Z^{t-1} puede ser suficiente probar si todas las transformaciones no-lineales de $\varepsilon(t,\theta)$ fueran no correlacionadas con todas las posibles funciones de Z^{t-1} . Según Ljung (1999), esto de seguro no es posible en la práctica y en su lugar se puede seleccionar cierta secuencia de un vector finito dimensional $\{\zeta(t)\}$, derivado de Z^{t-1} y demandando una cierta transformación de $\varepsilon(t, \theta)$ que sea no correlacionada con esta secuencia. Existen varios procedimientos para tal fin, los cuales toman diferentes formas, dependiendo de cuál modelo de estructura haya sido aplicada y la selección particular de ζ . El método de las variables instrumentales (IV siglas en inglés) quizás sea uno de los mejores conocidos en opinión de Ljung (1999), el mismo se describe a continuación.

Considerando un modelo de regresión lineal

$$\hat{y}(t \mid \theta) = \varphi^{T}(t)\theta + \mu(t)$$
(1.38)

Modelo que puede contener a su vez varios modelos típicos de sistemas tanto lineales como no-lineales, en el que la estimación de mínimos cuadráticos (LSE siglas en inglés) para θ viene dada por la ecuación 1.34 y puede ser expresada como (Ljung., 1999):

$$\hat{\theta}_N^{LS} = sol\left\{\frac{1}{N}\sum_{t=1}^N \varphi(t)[y(t) - \varphi^T(t)\theta] = 0\right\}$$
(1.39)

Ahora, suponiendo que los datos pueden ser descritos por:

$$y(t) = \varphi^{T}(t)\theta_{0} + v_{0}(t)$$
(1.40)

Teniendo en cuenta que en algunos casos típicos la LSE para θ_N no tiende a θ_0 por razones de correlación entre $w_0(t)$ y $\varphi(t)$; se selecciona una secuencia de vectores de correlación, en el que los elementos ζ son llamados instrumentos o variables instrumentales, de la siguiente manera (Ljung., 1999):

$$\zeta(t,\theta) = \zeta(t, Z^{t-1}, \theta) \tag{1.41}$$

Luego, proporcionando que la inversa que se indicará existe, esto convierte la LSE de θ en:

$$\hat{\theta}_N^{IV} = sol\left\{\frac{1}{N}\sum_{t=1}^N \zeta(t)[y(t) - \varphi^T(t)\theta] = 0\right\}$$
(1.42)

0

$$\hat{\theta}_{N}^{IV} = \left[\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\varphi^{T}(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\zeta(t)y(t)$$
(1.43)

Para que el método de la ecuación 1.42 sea exitosamente aplicable al sistema descrito por 1.40, se requieren las siguientes propiedades de las variables instrumentales $\zeta(t)$ (remplazando la media muestral por el valor esperado).

$$E\zeta(t)\varphi^T(t)$$
 sea no singular (1.44)

$$\bar{E}\zeta(t)\boldsymbol{v}_0(t) = 0 \tag{1.45}$$

En realidad lo que se trata de conseguir con estas condiciones es que los instrumentos deben ser correlacionados con las varibles de regresión, pero no correlacionados con el ruido. Se evidencia entonces que la selección de los instrumentos es clave para que se puedan suplir las demandas de los requisitos planteados.

Selección de los instrumentos.

Loa autores Ljung (2007b) y Gilson, Garnier, Young, & Van den Hof (2009) consideran que una idea natural es generar los instrumentos similares al modelo que se desea obtener, pero al mismo tiempo evitar que estén influenciados por el ruido del sistema $v_0(t)$. Un ejemplo de ello puede ser, suponiendo que la descripción real de la ecuación 1.40 corresponde con un modelo ARX como el explicado en secciones anteriores, representado como:

$$y(t) + a_1 y(t-1) + \dots + a_{na} y(t-n_a)$$

= $b_1 u(t-1) + \dots + b_{nb} u(t-n_b) + v(t)$ (1.46)

Obtener instrumentos como:

$$\zeta(t) = K(q)[-x(t-1) - x(t-2) ... - x(t-n_a) u(t-1) ... u(t-n_b)]^T$$
(1.47)

Donde: *K* es un filtro lineal.

x(t) es generado de las entradas a través del sistema lineal

$$N(q)x(t) = M(q)u(t)$$
(1.48)

$$N(q) = 1 + n_1 q^{-1} + \dots + n_{n_n} q^{-n_n}$$
(1.49)

$$M(q) = m_0 + m_1 q^{-1} + \dots + m_{n_m} q^{-n_m}$$
(1.50)

En la práctica la mayoría de los instrumentos son generados de esta forma, según Ljung (1999) y Gilson. et al (2009). Obviamente $\zeta(t)$ es obtenido a partir de entradas antiguas por filtrado lineal y puede ser escrito de manera conceptual como:

$$\zeta(t) = \zeta(t, u^{t-1}) \tag{1.51}$$

Si las entradas son generadas en lazo abierto, tal que no dependan del ruido del sistema $v_0(t)$, entonces la condición dada en la ecuación 1.45 claramente se cumple. Como que el vector φ y el vector ζ son generados de la misma secuencia de entradas (φ contiene en adición el efecto de v_0) se puede esperar que la ecuación 1.44 se cumpla en general.

Una simple y atractiva selección de los instrumentos es primero aplicar LSM a la ecuación 1.46 y luego utilizar el modelo de mínimos cuadrados estimados para N y M en la ecuación 1.48. Los instrumentos entonces son seleccionados en la forma de la ecuación 1.47 con K(q) = 1

La calidad del LSE del método de variables instrumentales ($\hat{\theta}_N^{IV}$) depende de la selección de los instrumentos $\zeta(t)$, por esto es recomendable seleccionar el filtro de la ecuación 1.48 igual a los del sistema real (Ljung., 1999). Para el caso que se ha supuesto, un sistema representado por un modelo ARX, es:

$$N(q) = A_0(q)$$
 (1.52)

$$M(q) = B_0(q)$$
 (1.53)

Los cuales son ciertamente desconocidos, pero se pueden hacer depender estos instrumentos en los parámetros de la forma obvia:

$$\zeta(t,\theta) = K(q)[-x(t-1,\theta) \dots - x(t-n_a,\theta) \quad u(t-1) \dots u(t-n_b)]^T$$
(1.54)

$$N(q)x(t) = M(q)u(t)$$
(1.55)

Y de forma general se puede escribir la generación de $\zeta(t, \theta)$ como sigue:

$$\zeta(t,\theta) = K_u(q,\theta)u(t) \tag{1.56}$$

Siendo $K_u(q, \theta)$ un vector columna de dimensión d del filtro lineal.

En fin, incluyendo un prefiltro lineal L(q) y haciendo

$$\varepsilon_F(t,\theta) = L(q)\varepsilon(t,\theta) \tag{1.57}$$

Y conformando una función $\alpha(\cdot)$ para los errores de predicción, el método de la variables instrumentales puede ser resumido por:

$$\varepsilon_F(t,\theta) = L(q)[y(t) - \varphi^T(t)\theta]$$
(1.58)

$$\hat{\theta}_N^{IV} = \sup_{\substack{\theta \in D_M}} [f_N(\theta, Z^N) = 0]$$
(1.59)

$$f_N(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^N \zeta(t, \theta) \alpha \left(\varepsilon_F(t, \theta) \right)$$
(1.60)

$$\zeta(t,\theta) = K_u(q,\theta)u(t) \tag{1.61}$$

Con D_M un set de valores dentro del cual θ es comprendido en un modelo de estructructura. Caso multivariable.

Suponiendo que el sistema está conformado por *p*-salidas y *m*-entradas, entonces los instrumentos $\zeta(t)$ son una matriz de dimensión $d \mathbf{x} p$. La generación lineal de $\zeta(t, \theta)$ puede hasta ser dada por la ecuación 1.56; con la interpretación que la *i*-ésima columna de $\zeta(t, \theta)$ está dada por:

$$\zeta^{(i)}(t,\theta) = K_u^{(i)}(q,\theta)u(t) \tag{1.62}$$

Siendo $K_u^{(i)}(q,\theta)$ un filtro matricial de dimensión $d \mathbf{x} m$. [$K_u(q,\theta)$ en la ecuación 1.56 es como tal un tensor, una "entidad de triple índice"]. Con $\alpha(\varepsilon)$ siendo una función de R^p a R^p y L(q) un filtro matricial de dimensión $p \mathbf{x} p$, el método de las variables instrumentales es todavía dado por la ecuaciones de la 1.58 - 1.61.

En opinión de Ljung (2007b), el método de las variables instrumentales, a diferencia de otros métodos de estimación de parámetros posee la ventaja de que no es sensible al tipo de ruido que afecta la respuesta del sistema, o sea, no asume que el ruido es blanco; por lo que es recomendable su uso cuando se desconocen las particularidades del tipo de ruido. Además para tamaños muestrales lo suficientemente grandes $(N \rightarrow \infty)$, el error cuadrático medio de los estimados por variables instrumentales tiende a cero, mientras que los LSE tienden asintóticamente al cuadrado de sus desviaciones; en tal sentido el método de IV se desarrolla mejor que el LSM para *N* suficientemente grande (R. Pintelon. & J. Schoukens., 2001). Según Gilson. et al (2009), la técnica de las variables instrumentales presenta la mayor ventaja de ser capaz de identificar consistentemente modelos de plantas en lazo cerrado y es aplicable ya sea conocida o no la estructura del modelo lineal que describe el controlador implementado en el proceso.

1.4.4 Algoritmos no iterativos (método de subespacios).

Los modelos en el espacio de estados pueden ser estimados usando varias técnicas, como son los métodos de caja negra en forma canónica y la estimación de máxima probabilidad (Ljung., 1999).

Se puede utilizar, además de estas técnicas, el método de los subespacios. Para ello se asume que no se tiene una percepción de la estructura particular y que se pueden justamente obtener cualquiera de las matrices *A*, *B*, *C*, y *D*, que den una buena descripción de la

conducta de entradas y salidas del sistema. Como en realidad existen un número infinito de tales matrices que describen el mismo sistema, el método de los subespacios se ocupa de encontrar la solución de las coordenadas de la base para la representación en el espacio de estados.

En presencia de un sistema lineal descrito en el espacio de estados, para el cual, según Ljung (1999), es más común separar el término de ruido compacto v(t) en la contribución de ruido medible v(t) y en el ruido de proceso actuando en los estados w(t). También se asume que v(t) y w(t) son secuencias de variables aleatorias independientes con valores de las medias y las covarianzas iguales a cero:

$$x(t+1) = Ax(t) + Bu(t) + w(t)$$
(1.63)

$$y(t) = Cx(t) + Du(t) + v(t)$$
(1.64)

Si u, x, e y fueran conocidos, el modelo de las ecuaciones 1.63 y 1.64 acabadas de plasmar se convierte en una regresión lineal, donde los parámetros desconocidos son todas las entradas de las matrices, mezcladas con señales medidas en combinaciones lineales. Una clara representación de ello es:

$$Y(t) = \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix}, \qquad \Theta = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

Luego, según Ljung (1999), el modelo de las ecuaciones 1.63 y 1.64 se puede escribir como:

$$Y(t) = \Theta\Phi(t) + E(t)$$
(1.65)

Para lo cual, todos los elementos en Θ pueden ser estimados por el LSM como fue descrito en las ecuaciones 1.63 – 1.64. La matriz de covarianza para E(t) es estimada facilmente como la suma de los cuadrados muestrales de los residuos del modelo (Ljung., 1999). Plantea este mismo autor que todo este procedimiento antes expusto, sin tener que sufrir cambios, se ajusta a sistemas multivariables; cuando las salidas y entradas son vectores. Como en realidad x no es conocido, el problema está en cómo obtener la secuencia x del vector de estado. Con tal fin se retoma un sistema dado por la representación de la respuesta al impulso:

$$y(t) = \sum_{j=0}^{\infty} [h_u(j)u(t-j) + h_e(j)e(t-j)]$$
(1.66)

Donde *u* es la entrada y *e* las innovaciones. Entonces el predictor formal *k*-pasos hacia delante es definido por la justa eliminación de las constribuciones a y(t) de e(j), u(j); j = t, ..., t - k + 1:

$$\hat{y}(t \mid t - k) = \sum_{j=k}^{\infty} [h_u(j)u(t - j) + h_e(j)e(t - j)]$$
(1.67)

Definido:

$$\hat{Y}_{r}(t) = \begin{bmatrix} \hat{y}(t \mid t-1) \\ \vdots \\ \hat{y}(t+r-1 \mid t-1) \end{bmatrix}$$
(1.68)

$$\hat{Y} = \left[\hat{Y}_r(1) \dots \hat{Y}_r(N)\right]$$
(1.69)

Según Ljung (1999), cuando $N \to \infty$ el vector de estado de una realización mínima en forma de innovación puede ser seleccionado como una combinación lineal de \hat{Y}_r que forme una fila base para \hat{Y} , como por ejemplo:

$$x(t) = L\,\hat{Y}_r(t) \tag{1.70}$$

<u>Donde:</u> *L* matriz de dimensión $n \mathbf{x} pr$ tal que $L\hat{Y}$ rota a \hat{Y} .

p: es la dimensión del vector de salida y(t).

De esta manera una representación canónica en el espacio de estados, de forma particular, puede corresponder a L matrices que justamente seleccionen ciertas filas de \hat{Y}_n . En general no se está confinado a tales selecciones, pero se permite escoger L tal que x(t) se convierta en una base bien acondicionada.

Una vez buscado el vector apropiado de los datos, el único problema que permanece es estimar el predictor k-pasos hacia delante. El predictor real es dado por $\hat{y}(t + k - 1 | t - 1)$ como se ve en 1.67 y la innovación e(t) puede ser escrita como una combinación lineal de datos pasados de entadas y salidas, por lo tanto el predictor puede ser expresado como una combinación lineal de u(i), y(i), $i \le t - 1$. Sin embargo por razones prácticas el predictor es aproximado tal que solo dependa de una cantidad fijada y finita de datos pasados como las s_1 salidas pasadas y las s_2 entradas pasadas, lo que quiere decir que toma la forma:

$$\hat{y}(t+k-1 \mid t-1) = \alpha_1 y(t-1) + \dots + \alpha_{s1} y(t-s_1) + \beta_1 u(t-1) + \dots + \beta_{s2} u(t-s_2)$$
(1.71)

Ljung (1999) plantea que este predictor puede ser eficientemente estimado por otro LSM lineal, proyectado directamente en datos de entradas y salidas del sistema; y como resultado se obtiene el modelo:

$$y(t + k - 1) = \theta_k^T \varphi_s(t) + \gamma_k^T U_\ell(t) + E(t)$$
(1.72)

O tratando con todos los r predictores simultáneamente:

$$Y_r(t) = \Theta\varphi_s(t) + \Gamma U_\ell(t) + E(t)$$
(1.73)

Donde:
$$\varphi_s(t) = [y^T(t-1) \dots y^T(t-s_1) \quad u^T(t-1) \dots u^T(t-s_2)]^T$$
 (1.74)

$$U_{\ell}(t) = [u^{T}(t) \dots u^{T}(t+\ell-1)]^{T}$$
(1.75)

$$Y_r(t) = [y^T(t) \dots y^T(t+r-1)]^T$$
(1.76)

$$\Theta = [\theta_1 \dots \theta_r]^T , \qquad \Gamma = [\gamma_1 \dots \gamma_r]^T$$
(1.77)

$$E(t) = [\varepsilon^{T}(t) \dots \varepsilon^{T}(t+r-1)]^{T}$$
(1.78)

 ℓ : es un número, tipicamente igual a r, de valores de entrada, los cuales influencian en la $Y_r(t)$ que está siendo considerada.

Además Θ y Γ en 1.73 puede ser estimada usando LSM, dando $\widehat{\Theta}_N$ y $\widehat{\Gamma}_N$, así los predictores *k*-pasos hacia delante están dados por:

$$\widehat{Y}_r(t) = \widehat{\Theta}_N \varphi_s(t) \tag{1.79}$$

Los cuales, para valores suficiente grandes de *s* dan una buena aproximación a los predictores reales.

Ljung (1999) y R. Pintelon & J. Schoukens (2001) resumen el enfoque del método de los subespacios para obtener las matrices del modelo en espacio de estados, incluyendo las bases para la representación y las matrices de covarianza del ruido; en los siguientes pasos:

- 1- Seleccionar s_1, s_2, r, ℓ , la forma de $\hat{Y}_r(t)$ en 1.51 e Y como en 1.68 y 1.69.
- 2- Estimar el rango n de Y, ya que el sistema de la ecuación 1.66 tiene una descripción mínima en el espacio de estados de n-ésimo orden si y solo si el rango de Ŷ es igual a n todo r ≥ n; y determinar L en 1.70 tal que x(t) corresponda con una base bien aconcionada para esto.

3- Obtener A, B, C, D, y la matriz de covarianza por aplicación de LSM a la regresión lineal 1.65.

Según Ljung (1999), las ventajas del método de los subespacios (N4SID) son varias, entre ellas se pueden citar las siguientes: Es particularmente apropiado para sistemas multivariables. Su algoritmo permite implementaciones confiables numéricamente y típicamente produce estimación de modelos con buena calidad. Sus resultados pueden ser complementados con otros métodos, como por ejemplo, el método de minimización del error de predicción (PEM siglas en inglés) en aras de mejorar la calidad de los resultados. El algoritmo contiene un número de alternativas y opciones en la forma de desarrollarlo, proporcionándole flexibilidad, variantes del enfoque pueden verse en Verheagen (1994); Van Overschee & DeMoor (1996); Borjas & Garcia (2011). En opinión de Ljung (2007b), es más rápido que el propio PEM, aunque de aquí una de sus desventajas, es menos preciso y robusto que el PEM aparte de requerir argumentos adicionales que pueden ser difíciles de especificar.

1.4.5 Algoritmos de estimación del error de predicción.

Teniendo en cuenta que un buen modelo es aquel que sea bueno al predecir un set de datos conocidos Z^N , o sea, que produzca errores pequeños de predicción cuando es aplicado a los datos observados; surge la incógnita de cómo calificar lo que puede significar pequeño en este contexto. Como respuesta a tal incógnita es conocido el método de minimización del error de predicción (PEM siglas en inglés).

El error de predicción dado por un cierto modelo $\mathcal{M}(\theta_*)$ se puede calcular a partir de:

$$\varepsilon(t,\theta_*) = y(t) - \hat{y}(t \mid \theta_*) \tag{1.80}$$

La secuencia del error de predicción puede ser vista como un vector en \mathbb{R}^N . La magnitud de este vector puede ser medida usando una norma en \mathbb{R}^N , cuadrática o no (Ljung., 1999); (Van den Hof., Dankers., Heuberger., & Bombois., 2013). Al ser filtrada, la secuencia del error de predicción, a través de un filtro lineal estable L(q):

$$\varepsilon_F(t,\theta) = L(q)\varepsilon(t,\theta) \qquad 1 \le t \le N \tag{1.81}$$

Además usando la siguiente norma, se tiene que:

$$V_N(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^N \ell(\varepsilon_F(t, \theta))$$
(1.82)

Donde: $\ell(\cdot)$ es una función de valor escalar (típicamente positiva).

 $V_N(\theta, Z^N)$ es, para un Z^N dado, una función bien definida de valor escalar de los parámetros del modelo θ .

Luego los parámetros estimados del modelo $\hat{\theta}_N$ están definidos por minimización de la ecuación acabada de plasmar 1.82.

$$\hat{\theta}_N = \hat{\theta}_N(Z^N) = \underset{\theta \in D_{\mathcal{M}}}{\operatorname{arg\,min\,}} V_N(\theta, Z^N)$$
(1.83)

Aquí arg min significa la minimización del argumento de la función.

Selección del filtro L:

El efecto del filtro L es permitir libertad extra en el tratamiento con las propiedades no constantes del error de predicción. Según Ljung (1999), si el predictor es lineal, el sistema de tiempo invariante y las y - u son escalares; entonces el resultado de filtrar ε , es el mismo que se obtiene si primero se filtran los datos de entradas y salidas y luego se aplicara el predictor. Como consecuencia de usar un filtro L pueden ser removidos efectos de disturbios de alta frecuencia no esenciales para el problema de modelación o cambios incontrolados lentos.

En el caso de que sea usado un modelo lineal e invariante en el tiempo, de tiempo discreto, como el de la ecuación 1.1; el aspecto particular de filtración de la ecuación 1.81 puede ser empleado, manifestándose de la siguiente manera (Ljung., 1999):

$$\varepsilon_F(t,\theta) = L(q)\varepsilon(t,\theta) = [L^{-1}(q)H(q,\theta)]^{-1}[y(t) - G(q,\theta)u(t)]$$
(1.84)

Acá el efecto de la pre-filtración es como tal idéntico al cambio del modelo del ruido de $H(q, \theta)$ a:

$$\overline{H}_{L}(q,\theta) = L^{-1}(q)H(q,\theta) \tag{1.85}$$

Cuando se describen y analizan métodos que emplean modelos de ruido generales en sistemas lineales, se restringe normalmente a $L(q) \equiv 1$, debido a que la opción de prefiltrado es tomada con cuidado por la libertad en la selección de $H(q, \theta)$ (Ljung., 1999).

Selección de la norma (l).

Para la selección de la norma $\ell(\cdot)$, el primer candidato debe ser una norma cuadrática:

$$\ell(\varepsilon) = \frac{1}{2}\varepsilon^2 \tag{1.86}$$

Esta es ciertamente una selección estandar, conveniente tanto para el cálculo como el análisis (Ljung., 1999). Cuestiones como la robustez contra los datos incorrectos justifican

el uso de otras normas. Se pueden además concebir situaciones donde la mejor norma no se conoce de antemano y se hace rasonable parametrizar la norma en si misma $\ell(\varepsilon, \theta)$.

A menudo la parametrización de la norma es independiente de la parametrización del modelo resultando:

$$\theta = \begin{bmatrix} \theta \\ \alpha \end{bmatrix} : \ell(\varepsilon(t,\theta),\theta) = \ell(\varepsilon(t,\theta'),\alpha)$$
(1.87)

Retomando la idea de la minimización de la ecuación acabada 1.82, la forma de obtener los $\hat{\theta}_N$ comprende procedimientos bien conocidos y muy usados. Métodos particulares y con nombres específicos, son obtenidos como casos especiales en dependencia de la selección de la norma $\ell(\cdot)$, el prefiltro $L(\cdot)$, la estructura del modelo y el propio método por el cual es realizada la minimización (Ljung., 1999). En la presente investigación se presta particular atención a dos algoritmos bien conocidos destinados a tal fin: el método de los mínimos cuadrados (visto en la sección 1.3.2) y el método de la máxima verosimilitud, el cual es abordado en la próxima sección.

Los métodos del error de predicción son un enfoque básico para la modelación matemática, gracias a tres importantes ventajas según criterio de Ljung (1999).

- 1- Aplicabilidad a estructuras de modelos generales.
- 2- Óptima exactitud asintótica cuando el sistema real puede ser representado dentro del modelo de estructura analizado.
- 3- Propiedades de aproximación razonable cuando el sistema real no puede ser representado dentro del modelo de estructura analizado.

1.4.5.1-<u>Algoritmo iterativo de minimización del error de predicción (método de máxima verosimilitud).</u>

La modelación matemática y la estimación de parámetros, tratan con el problema de la extracción de información de observaciones que por sí solas pueden ser no fiables (Ljung., 1999). Según el mismo autor, las observaciones son entonces descritas como muestras de variables estocásticas y se puede suponer que son representadas por la variable aleatoria $y^N = (y(1), y(2), ..., y(N))$ que toma valores en \mathbb{R}^N . Su función de densidad de probabilidad (PDF siglas en inglés) se puede suponer como:

$$f(\theta; x_1, x_2, \dots, x_N) = f_{\mathcal{V}}(\theta; x^N) \tag{1.88}$$

Donde
$$P(y^N \epsilon A) = \int_{x^N \epsilon A} f_y(\theta; x^N) dx^N$$
 (1.89)
El vector de los parámetros de dimensión d de la ecuación 1.88, que describe las propiedades de la variable observada, se puede suponer que no es conocido; por lo que el propósito de las observaciones como tal puede ser obtener dicho vector usando y^N . Esto puede ser logrado mediante el uso de un estimador $\hat{\theta}(y^N)$ que es función de \mathbf{R}^N a \mathbf{R}^d . Suponiendo que el valor de y^N es y_*^N , entonces consecuentemente el resultado del estimador antes definido queda:

$$\hat{\theta}_* = \hat{\theta}(y_*^N) \tag{1.90}$$

Existen varios estimadores que se pueden seleccionar, uno en particular que maximiza la probabilidad del evento observado, además de los más utilizados en opinión de Hair (1999) y Ljung (1999); es el estimador de máxima verosimilitud, introducido por Fisher (1912). Según Ljung (1999), se puede definir este estimador como la función de densidad de probabilidad conjunta del vector aleatorio que es observado, dado en la ecuación 1.88. La probabilidad de que una observación ciertamente deba tomar el valor y_*^N es como tal proporcional a $f_y(\theta; y_*^N)$. Esta es una función determinística de θ una vez que el valor numérico de y_*^N es evaluado en la misma y se le conoce como la función de verosimilitud; refleja la probabilidad de que el evento observado deba ciertamente ocurrir. Así de una manera razonable, el estimador de máxima verosimilitud de θ se encarga de que su selección sea tal que el evento observado se convierta tan probable como sea posible, visto entonces como:

$$\hat{\theta}_{ML}(y_*^N) = \arg\max_{\theta} f_y(\theta; y_*^N)$$
(1.91)

La maximización es llevada a cabo para un y_*^N prefijado.

Función de probabilidad para modelos probabilísticos de sistemas dinámicos:

Dado un modelo probabilístico completo $\mathcal{M}(\theta)$, que incluye tanto una función predictora como una función de densidad de probabilidad; asumida para los errores de predicción asociados:

$$\mathcal{M}(\theta): \quad \hat{y}(t \mid \theta) = g(t, Z^{t-1}; \theta) \tag{1.92}$$

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t \mid \theta) \quad \text{sean independientes}$$

$$Y \text{ con PDF } f_e(x,t;\theta)$$
(1.93)

Las salidas del mismo son generadas por:

$$y(t) = g(t, Z^{t-1}; \theta) + \varepsilon(t, \theta)$$
(1.94)

Una vez conocida la PDF conjunta para las observaciones y^N (dada la secuencia determinística u^N) es posible obtener la función de probabilidad:

$$f_{y}(\theta; y^{N}) = \prod_{t=1}^{N} f_{e}(y(t) - g(t, Z^{t-1}; \theta), t; \theta) = \prod_{t=1}^{N} f_{e}(\varepsilon(t, \theta), t; \theta)$$
(1.95)

Maximizar esta función es lo mismo que maximizar:

$$\frac{1}{N}\log\bar{f}_{y}(\theta;y^{N}) = \frac{1}{N}\sum_{t=1}^{N}\log f_{e}(\varepsilon(t,\theta),t;\theta)$$
(1.96)

Si se define $\ell(\varepsilon, t, \theta) = -log f_e(\varepsilon, t, \theta)$.

Caso especial Gaussiano:

Cuando los errores de predicción son asumidos de tipo Gaussiano, con valores de media cero y covarianza λ se tiene que:

$$\ell(\varepsilon, t, \theta) = -\log f_e(\varepsilon, t, \theta) = const + \frac{1}{2}\log\lambda + \frac{1}{2}\frac{\varepsilon^2}{\lambda}$$
(1.97)

La cual, para un λ conocido, se convierte en el criterio cuadrático de la ecuación 1.86 $\ell(\varepsilon) = \frac{1}{2}\varepsilon^2$; si λ es desconocida se está en presencia de un ejemplo de criterio de norma parametrizada $\ell(\varepsilon, \theta)$ pudiendo ser útil la ecuación 1.87.

Los estimadores de máxima verosimilitud poseen una gran ventaja como consecuencia de sus propiedades asintóticas. Un resultado clásico en el caso de observaciones independientes, el obtenido por Cramér (1946) y Wald (1949) para las propiedades del límite cuando el tamaño muestral (en este caso el número de N) tiende a infinito; comprende una clara evidencia de ello, en el que se plantea que:

Si las variables aleatorias $\{y(i)\}$ son independientes e idénticamente distribuidas, tal que:

$$f_{y}(\theta; x_{1}, x_{2}, \dots, x_{N}) = \prod_{t=1}^{N} f_{y(t)}(\theta; x_{t})$$
(1.98)

Suponiendo que la distribución de y^N está dada por $f_y(\theta_0; x^N)$ para algunos valores de θ_0 . Entonces la variable aleatoria $\hat{\theta}_{ML}(y^N)$ tiende a θ_0 con probabilidad igual a uno cuando $N \to \infty$ y la variable aleatoria $\sqrt{N}[\hat{\theta}_{ML}(y^N) - \theta_0]$ converge en la distribución a la distribución normal con media igual a cero. En este sentido, el estimador de máxima verosimilitud, es como tal el mejor estimador posible. Según Ljung (1999), este resultado también se cumple cuando el estimador de máxima verosimilitud es aplicado a sistemas dinámicos.

1.5. Modelos de sistemas no-lineales:

Una relación no-lineal entre las secuencias de entradas y de salidas da posibilidades mucho más ricas para describir sistemas, y al mismo tiempo, la situación se convierte mucho más flexible (Ljung., 1999). La mayoría de los sistemas encontrados en el mundo real son nolineales S. Chen, S. Billings & Luo (1989); R. Pintelon, J. Schoukens & Rolain (2003); Tóth, Heuberger & Van den Hof (2009); Sakthivel, Anandhi & Natarajan (2011); de ahí que a menudo los modelos lineales no pueden capturar la conducta dinámica que manifiestan. El campo de la modelación matemática de sistemas no-lineales está progresando rápidamente en la actualidad, con varios algoritmos que están disponibles en paquetes de software estándar como MATLABTM System Identification Toolbox (Wigren. & J. Schoukens., 2013). Según Ljung & Vicinio (2005), la obtención de modelos nolineales es probablemente el área más activa en la modelación de sistemas en el presente. Si bien una relación no-lineal convierte la situación mucho más flexible según Ljung (1999); en el sentido de lograr una representación más ajustada a los sistemas reales; la tarea desde el punto de vista de la obtención de los modelos no-lineales se torna verdaderamente compleja. Autores como Van Donkelaar (2000); Lyzell, Glad, Enqvist & Ljung (2009) plantean que la razón está en el reto de la estimación de los parámetros, que con frecuencia cae en la necesidad de dar solución a problemas de optimización no-convexos; y según T. Chen, Ohlsson, Goodwin & Ljung (2011), no se puede garantizar de manera segura que el óptimo global de un criterio no-convexo siempre sea logrado; pues en opinión de Luenberger & Ye (2008), en la teoría básica asociada con la optimización en la programación no-lineal se requiere del cumplimiento de una serie de condiciones (necesarias y/o suficientes) para lograr un punto de solución (óptimo local y/o global).

Como se ha visto en las secciones anteriores, un modelo es una función de datos antiguos en el espacio de las salidas. En el caso no-lineal, según Ljung (1999), esta función puede ser descrita por la estructura general:

$$\hat{y}(t \mid \theta) = g(t, Z^{t-1}; \theta) \tag{1.99}$$

Entre las diferentes estructuras de modelos de sistemas no-lineales se encuentran los modelos polinomiales no-lineales y los modelos Weiner & Hammerstein. De una breve descripción de los mismos, así como de algunas de sus posibles parametrizaciones se encargan las dos siguientes secciones del presente capítulo.

1.5.1 Modelos polinomiales no-lineales:

Si se omite la dependencia del tiempo de manera explícita la ecuación (1.99) también puede ser expresada como:

$$\hat{y}(t \mid \theta) = g(Z^{t-1}; \theta) \tag{1.100}$$

En casos de no poseer una percepción particular de las propiedades del sistema como tal descrito por la ecuación 1.100; debe entonces lograrse una parametrización de g que sea flexible y cubra todo tipo de conducta de sistema razonable, dando como resultado una estructura de modelo polinomial de caja negra no-lineal (Ljung., 1999).

El autor antes mencionado establece que la ecuación 1.100 describe una familia de estructura de modelo que es general también y a su vez útil para los modelos no-lineales, donde se escribe g como una concatenación de dos funciones: una que toma el número incrementado de observaciones pasadas Z^t y las relaciona dentro de un vector dimensional finito $\varphi(t)$ de dimensión fijada; la otra transporta este vector al espacio de las salidas.

$$g(Z^{t-1};\theta) = g(\varphi(t),\theta) \tag{1.101}$$

Donde:

$$\varphi(t) = \varphi(Z^{t-1}) \tag{1.102}$$

La selección de la función no-lineal en la ecuación 1.100 ha sido como tal descompuesta en dos problemas parciales para sistemas dinámicos (Ljung., 1999):

- 1- Cómo seleccionar el vector de regresión $\varphi(t)$ de entradas y salidas antiguas.
- 2- Cómo seleccionar el la función no-lineal $g(\varphi, \theta)$ del espacio del regresor al espacio de la salida.

La selección del regresor es de seguro dependiente de la aplicación, o sea, de las condiciones propias del sistema que se trate. Según Ljung (1999), típicamente los regresores son seleccionados en la misma forma que para modelos lineales: mediciones antiguas, posiblemente salidas de modelos antiguos y errores de predicción antiguos. Por esta razón se puede estar hablando entonces, análogamente a la **Tabla 1.1**, sobre estructuras

de modelos NARX, NARMAX, NOE y NBJ; donde la letra N da el sentido de no-lineales a cada una de las variantes ya antes vistas.

Con respecto a la función no-lineal $g(\varphi, \theta)$ se tiene que para un θ dado convierte $\mathbf{R}^{\mathbf{d}}$ a $\mathbf{R}^{\mathbf{p}}$ y puede incluir tanto funciones lineales como no-lineales (Ljung., 2007b). La función $g(\varphi, \theta)$ puede además ser expresada como expansión de funciones en la forma siguiente (Ljung., 1999):

$$g(\varphi, \theta) = \sum_{k=1}^{n} \alpha_k g_k(\varphi), \qquad \theta = [\alpha_1 \dots \alpha_n]^T$$
(1.103)

Donde se refiere a g_k como funciones bases, ya que el papel que juegan en la ecuación 1.103 es el de una base del espacio funcional. La expansión que se acaba de mostrar, con el uso de diferentes funciones bases, juega el papel de un marco unificado para la investigación de los más conocidos modelos de estructuras no-lineales polinomiales de caja negra (Ljung., 1999). Luego una cuestión clave es el cómo seleccionar las funciones bases g_k .

Un modo de descripción general de $g(\varphi, \theta)$ está comprendido por la siguiente ecuación:

$$g(\varphi, \theta) = \sum_{k=1}^{n} \alpha_k \mathcal{K}(\beta_k (x - \gamma_k))$$
(1.104)

Donde: \mathcal{K} : es la función no-lineal unitaria.

n: es el número de unidades no-lineales.

 α_k , β_k y γ_k : son los parámetros del estimador no-lineal.

Estos tres tipos de parámetro afectan el modelo en formas completamente diferentes. Las coordenadas α_k entran linealmente, lo cual significa que la ecuación 1.104 en una regresión lineal para parámetros previamente prefijados de escala o dilatación β_k y de localización γ_k (Ljung., 1999).

Como $g(\varphi, \theta)$ es una función de regresión no-lineal, la cual es aproximada por estimadores no-lineales en opinión de Ljung (1999); un tema importante es conocer sobre estimadores no-lineales disponibles para los modelos polinomiales no-lineales de caja negra.

1.5.2 Modelos Hammerstein & Weiner:

Es una situación completamente común que, mientras la dinámica en sí misma puedes ser correctamente descrita por un sistema lineal, existen no-linealidades estáticas en las entradas y/o salidas del sistema. Entre las posibles causas que originan estas

particularidades están la presencia de actuadores no-lineales debido a saturación y las características no-lineales de los sensores (Ljung., 1999). Un modelo con no-linealidades estáticas en las entradas es llamado Hammerstein, mientras que un modelo con no-linealidades en las salidas es llamado Weiner; luego la combinación de ellos es un modelo Hammerstein & Weiner. Autores como (J. Schoukens, R. Pintelon & Rolain (2006) y Ljung (2007a) se refieren a este tipo de modelos como una extensión de los modelos lineales; modelos orientados en bloques que consisten en una cascada de dinámicas lineales y no-linealidades estáticas.

La siguiente ecuación general describe la estructura Hammerstein & Weiner:

$$w(t) = f(u(t)) \tag{1.105}$$

$$x(t) = \frac{B_{j,i}(q)}{F_{j,i}(q)}w(t)$$
(1.106)

$$y(t) = h(x(t)) \tag{1.107}$$

Donde: $u(t) \neq y(t)$ son las entradas y salidas del sistema respectivamente.

f y h: son funciones no-lineales que corresponden a las entradas y salidas nolineales respectivamente. Para múltiples entradas y/o salidas, f y/o h están definidas en forma de componentes.

w(t) y x(t): son variables internas. w(t) tiene la misma dimensión que u(t) y x(t) tiene la misma dimensión que y(t):

B(q) y F(q): en el bloque dinámico lineal son polinomios en el operador de tiempo de retardo, como los descritos en los modelos polinomiales de la sección 1.3.1.

Para *ny* salidas y *nu* entradas el bloque lineal es una matriz de funciones transferencias que contiene entradas en la forma $\frac{B_{j,i}(q)}{F_{j,i}(q)}$ con j = 1, 2, ..., ny y con i = 1, 2, ..., nu:

1.5.3 Estimadores para modelos no-lineales:

A pesar de que en la teoría de la estimación, las propiedades asintóticas y los algoritmos básicos son los mismos para algunos modelos no-lineales como para otros de los modelos tratados anteriormente en este capítulo, existen algoritmos y métodos que son particulares para los modelos no-lineales según Ljung (1999); a continuación se hace referencia a dos de ellos; las redes en forma de "S" y las redes onduladas.

1.5.3.1-<u>Redes en forma de "S":</u>

Las redes en forma de "S" son usadas para obtener modelos no-lineales polinomiales y modelos Hammerstein & Weiner. Definen una función no-lineal y = F(x) donde y es un escalar y x es un vector columna de dimensión m. La función es basada en la siguiente expansión:

 $F(x) = (x - r)PL + \alpha_1 f((x - r)Qb_1 - c_1) + \dots + \alpha_n f((x - r)Qb_n - c_n) + d \quad (1.108)$ **Donde**: *P* y *Q*: son matrices de proyección de dimensiones *m x p* y *m x q* respectivamente; además son estimadas por el análisis de componentes principales de los datos empleados para la estimación del modelo.

Usualmente p = m.

Si los componentes de x en los datos de estimación son linealmente dependientes, entonces p < m.

El número de columnas de Q, q, corresponden al número de componentes de x usados en la función en forma de "S".

f: la función forma de "S" definida por la ecuación $f(z) = \frac{1}{e^{-z}+1}$:

r: es un vector de dimensión 1 x m y representa el valor medio del vector regresor calculado de los datos de estimación.

 $d, \alpha_k, y c_k$: son escalares.

L: es un vector de dimensión $p \ge 1$.

 b_k : son vectores de dimensión $q \ge 1$.

1.5.3.2-<u>Redes onduladas.</u>

Las redes onduladas son usadas para obtener modelos no-lineales polinomiales y modelos Hammerstein & Weiner. Definen una función no-lineal y = F(x) donde y es un escalar y x es un vector columna de dimensión m. La función es basada en la siguiente expansión:

$$F(x) = (x - r)PL + \alpha_{s1}f((b_{s1}(x - r))Q - c_{s1}) + \cdots + \alpha_{sn}f((b_{sn}s(x - r))Q - c_{sn}) + \cdots + \alpha_{w1}g((b_{w1}(x - r))Q - c_{w1}) + \cdots + \alpha_{wn}wg((b_{wn}w(x - r))Q - c_{wn}) + d$$
(1.109)

Donde: *f* : es una función de escala.

g: es la función ondulada.

 $P ext{ y } Q$: son matrices de proyección de dimensiones $m ext{ x } p ext{ y } m ext{ x } q$ respectivamente; además son estimadas por el análisis de componentes principales de los datos empleados para la estimación del modelo.

Usualmente p = m.

Si los componentes de x en los datos de estimación son linealmente dependientes, entonces p < m.

El número de columnas de Q, q, corresponden al número de componentes de x usados en la funciones de escala y ondulada.

r: es un vector de dimensión 1 x m y representa el valor medio del vector regresor calculado de los datos de estimación.

 $d, \alpha_s, b_s, \alpha_w y b_w$: son escalares. Los parámetros con el subíndice *s* son los parámetros de escala y los parámetros con el subíndice *w* son los parámetros de la onda.

L: es un vector de dimensión $p \ge 1$.

 c_s y c_w : son vectores de dimensión 1 x q.

La función de escala f y la función ondulada g son ambas funciones radiales de la siguiente forma:

$$f(x) = e^{-0.5x'x} \tag{1.1010}$$

 $g(x) = (\dim(x) - x'x)e^{-0.5x'x}$

1.6. Conclusiones del capítulo.

- 1- La modelación matemática, empleando métodos de parametrización de caja negra, se convierte en una herramienta potente, pues contando con poco más que un record de entradas y salidas del sistema, es posible describir la dinámica del proceso y las perturbaciones que le interfieren.
- 2- El método de los mínimos cuadrados lineales constituye una herramineta básica aplicable a la parametrización de modelos polinomiales y de espacio de estados para los sistemas dinámicos.
- 3- Los modelos no-lineales presentan mayor complejidad de estrutura que los lineales, debido a las características de las funciones que los integran, además sus métodos de estimación son más costosos desde el punto de vista operacionanal.

(1.1011)

CAPÍTULO 2. Procedimiento para modelar sistemas en procesos industriales a gran escala.

Introducción.

En procesos industriales a gran escala, la modelación matemática basada en pruebas a lazo abierto, generalmente no es permitida (Pitta. & Odloak., 2012). Dada la necesidad de involucrar gran número de operadores, técnicos de instrumentación e ingenieros, desarrollar pruebas a lazo abierto de larga duración, puede tomar semanas de preparación, sin embargo, la posibilidad de incurrir en graves errores operacionales sigue siendo elevada. Según los autores antes mencionados, por lo general, en el emplazamiento de una planta no se autoriza a esta clase de pruebas en un escenario competitivo; luego en estos casos, la modelación a lazo cerrado es una práctica necesaria y requiere ser tomada con seriedad. Producto de esta situación, sumado a las peculiaridades del proceso estudiado, el presente capítulo se centra en el diseño de un procedimiento para la modelación matemática de sistemas que operan en configuración de lazo cerrado. El mismo consta de dos epígrafes, el primero describe el sistema bajo análisis y en el segundo se sientan las pautas fundamentales que permitan identificarlo de una forma correcta.

2.1 Descripción del proceso.

La Sección 400, o Bloque 400 de la Refinería de Petróleos de Cienfuegos, está diseñada para el fraccionamiento de gases provenientes de la planta de destilación atmosférica Sección 100 y de la planta de reformación catalítica Sección 200. Su objetivo principal es la obtención de propano, n-butano, isobutano, gas seco y 70°C estable.

Se cuenta en este proceso con flujos de entradas estándares, 27.98 m³/h con punto inicial de ebullición (pie) 70°C inestable de la Sección 100 y 1.77 m³/h de una mezcla liviana inestable de la Sección 200.

En la actualidad el sistema funciona con una sola columna de fraccionamiento, la columna desbutanizadora T-401/1, la cual presenta las siguientes características de operación. La materia prima debe entrar a una temperatura de 124°C a la columna por los platos 41, 44 y 47 a una presión 19,5 Kgf/cm² y así comienza el proceso de destilación. Por el tope de la columna deben salir los compuestos más ligeros de hidrocarburos, dígase (C1, C2, C3 y C4) metano, etano, propano y butano; a una temperatura de 97°C y a una presión de 19 Kgf/cm² hacia los enfriadores de aire y de estos al tambor de reflujo del tope, a una

temperatura de 61°C y a una presión de 15 Kgf/cm². Dentro del tambor, a estos parámetros operacionales, ocurre la condensación del propano, isobutano y n-butano; que son los componentes principales del GLP. Mientras que los compuestos metano y etano, que permanecen como incondensables para esas condiciones de operación, pasan en forma gaseosa a las fuentes de suministro de gases de los quemadores de los hornos de la Sección 100. El líquido contenido en dicho tambor es considerado como producto terminado o GLP, del cual se utiliza una porción como reflujo del tope de la columna a través de una bomba; con una temperatura de 60 °C y 27 Kgf/cm² de presión.

Por el fondo de la columna se obtienen los elementos más pesados de la mezcla entrante, conocidos como pie 70 °C estable. Parte de este pie 70 °C estable se envía a razón de 19.39 m³/h hacia el horno F-401/1, a una temperatura de 167 °C y 19.5 Kgf/cm² de presión. Del horno vuelven ser re-inyectados a la columna por debajo del plato 1. Desde el fondo de la columna también sale una fracción de pie 70 °C estable hacia los intercambiadores de calor, donde le ceden parte de su energía al pie 70 °C inestable. Luego va al intercambiador por agua para ser enfriado hasta una temperatura de 34 °C y una presión de 19.4 Kgf/cm², siguiendo camino al patio de tanques. Un diagrama de flujo del proceso puede ser visto en el Anexo I, en el cual se representan equipos y líneas fundamentales así como los principales parámetros a controlar durante la operación.

El control de temperatura del tope es fundamental, ya que a través de mismo se puede modificar la composición del gas que sale por el tope de la columna. El sistema que en la actualidad está en funcionamiento y que es el objeto de la presente investigación puede ser representado de forma esquemática como se muestra en la figura 2.1:



Figura 2.1: Esquema de control del tope de la columna desbutanizadora T-401/1. Fuente: (Alonso Díaz. et al., 2012)

2.2 Procedimiento para la modelación matemática de sistemas en procesos industriales a gran escala.

Como se ha dicho anteriormente, la modelación matemática es un proceso cíclico lógico, donde se destacan etapas claves según criterio de autores consultados, Zill (1997); Ljung (1999); R. Pintelon & J. Schoukens (2001); Liu & Gao (2012). Por esta razón el procedimiento propuesto, está basado en las fases referenciadas por el conjunto de los autores antes mencionados. Comprenden las mismas la planificación del experimento, la recolección y pre-procesamiento de los datos, la selección de un set de modelos, la obtención del modelo mediante el método seleccionado y la validación del mismo, la cual puede ser una última etapa una vez que haya sido probado que el modelo realmente es bueno para el fin que ha sido previsto, de lo contrario tiene que retomarse alguna de las etapas previas en busca de mejorar el proceso de modelación. A continuación se explican, los pasos a realizar en el procedimiento diseñado en la presente investigación.

2.2.1 Fase I - Planificación del experimento.

Dado el hecho de que los modelos estén relacionados con el experimento, es importante planificar experimentos que cumplan con la intención de uso del modelo. De ahí que un objetivo principal sea planificar experimentos que conduzcan a datos mediante los cuales se pueda discriminar entre diferentes modelos en un conjunto de modelos planificados (Ljung., 1999). La planificación del experimentos se convierte en un tema crucial en la modelación matemática por la presencia errores en los modelos (J. Schoukens. et al., 2006). La planificación del experimento incluye:

2.2.1.1 - Selección de las señales a medir.

La selección de las señales a medir no es más que la elección de las variables de entradas y salidas que resulten de interés para el estudio del sistema. Además de seleccionar cuáles van hacer medidas, es importante decidir cuándo medirlas, cuáles van a ser manipuladas y cómo manipularlas.

Selección de las variables de entradas y salidas.

Al confrontarse con un sistema físico como el de una columna desbutanizadora, una vez que la dinámica del mismo haya sido identificada, hay un número de cuestiones que son respondidas, por ejemplo: qué es considerado como entrada, qué es considerado como salida y dónde en el proceso deben estar situados los sensores. Al tratarse de un sistema complejo, en el que intervienen un gran número de variables, realmente un área importante es la reducción de la dimensión de los datos, para aplicar variantes no-lineales o suavizar el problema de la parametrización no-lineal del modelo. La optimización no convexa que se origina producto de la parametrización no-lineal es difícil de resolver para la modelación matemática de procesos industriales, los cuales a menudo tienen un gran número de variables (Van Donkelaar., 2000). Esta es una contraparte no-lineal del análisis de datos multivariantes, en la cual puede ser aplicado el Análisis de Componentes Principales (Ljung., 2010). Desde la perspectiva del resumen de datos, según Hair (1999), el método de los componentes principales como una alternativa del análisis factorial, proporciona al investigador una comprensión clara de cuáles variables realmente se puede esperar que tengan un impacto en el análisis.

Cuándo medir las señales.

Cuando ha sido decidido dónde y qué medir, la siguiente cuestión es cuándo medir. El procedimiento de muestrear los datos que son producidos por los sistemas es inherente en la adquisición de datos basados en computadoras. Es inevitable que el muestreo como tal conlleve a pérdidas de información, razón por la cual es importante seleccionar instancias de muestreo que hagan estas pérdidas insignificantes. Con frecuencia las señales son medidas usando un intervalo de muestreo constante T (equidistantes), entonces esta cantidad tiene que ser seleccionada. Varios criterios según (Ljung., 1999) pueden ser aplicados, entre ellos se tienen:

- Si el tiempo total del experimento 0 ≤ t ≤ T_N es limitado, pero la adquisición de los datos dentro de este tiempo es poco costosa, es claramente ventajoso desde un punto de vista basado en la teoría de la información, muestrear tan rápido como sea posible.
- Seleccionar *T* tal que sea una especie de negocio entre la reducción del ruido y la relevancia de la dinámica del sistema. Lo anterior se debe a que intervalos de muestreos que sean más largos que las constantes de tiempos relevantes del sistema, pueden entonces proporcionar datos con poca información sobra las dinámicas; mientras que intervalos de muestreos pequeños, por otra parte, no dan margen para mucha reducción del ruido.

- Si el modelo puede ser usado para propósitos de control, otros ciertos aspectos deben tenerse en cuenta. El intervalo de muestreo para el cual se construya el modelo debe ser el mismo que para la aplicación de control (a menos que se desee volver a calcular este de un intervalo de muestreo a otro). Además un modelo obtenido de un muestreo rápido con frecuencia va a ser de fase mínima, y sistemas con tiempos muertos, pueden ser modelados con retrasos de muchos períodos de muestreo; tal efecto puede causar problemas para el diseño del control e influir además en la selección de *T*.
- Una selección óptima de *T* para un número fijado de muestras cae en el rango de las constantes de tiempo del sistema. Sin embargo, no es exactamente conocido y sobreestimarlo puede conducir a malos resultados. De esta forma una apropiada selección de la frecuencia de muestreo cae en el rango de diez veces el ancho de banda predicho del sistema. En la práctica, es útil primero registrar una respuesta al paso del sistema y luego seleccionar el intervalo de muestreo tal que este tome entre cuatro y seis muestras durante el tiempo de subida.

2.2.1.2 - Selección de las señales a manipular y cómo manipularlas.

Una ideal situación en la modelación matemática a lazo cerrado es el caso donde los datos de la rutina de operación pueden ser usados en el algoritmo de modelación sin la necesidad de la introducción de disturbios externos adicionales, sin embargo, este planteamiento carece de aplicabilidad práctica. La selección de las señales a manipular se refiere específicamente a seleccionar cuáles entradas del sistema deben ser excitadas durante el experimento. Las señales de entrada determinan el punto de operación del sistema, cuáles partes y de qué modos va a ser excitado el sistema (Ljung., 1999). La libertad de los usuarios en la selección de las características de las entradas puede variar considerablemente con la aplicación. En la industria de procesos, puede que no sea permitido del todo manipular un sistema en modo de producción continua, pues debido a las especificaciones de operación y las estrictas restricciones de seguridad, es probable incurrir en pérdidas inaceptables en la producción y la calidad de los productos (Pitta. & Odloak., 2012); además se puede exponer el personal de operación, la propia instalación y el medio ambiente a elevados niveles de riesgo. Esto puede entonces presionar a que algunas señales asociadas al proceso, que siendo correctamente consideradas como

entradas, en el sentido que ellas afectan al sistema; sin embargo no sea posible, factible o permitido manipularlas. Según Ljung (1999), si estas señales son medibles, es aún muy favorable incluirlas entre la señales de entradas medibles y tratarlas como tal en la construcción del modelo; aunque desde un punto de vista operacional se debe preferir considerarlas como disturbios (ruidos) medibles.

Para el resto de las señales de entrada que no ocurra una situación similar, si es de vital importancia excitar las mismas. A pesar de la presencia de varios disturbios que afectan al sistema durante la operación normal, usualmente estas excitaciones no son suficientes para permitir la correcta obtención de un modelo del sistema. Esto ocurre por dos razones principales, la primera, el nivel de estas exitaciones no son lo suficientemente largas comparadas con el nivel de ruido de los instrumentos; la segunda, a menos que una de las salidas alcance uno de sus límites en la zona de la estrategia de control, el controlador no reacciona a los disturbios. Tales evidencias han sido encontradas en publicaciones de autores como Ogata (1998); Ljung (1999); R. Pintelon & J. Schoukens (2001), Pitta & Odloak (2012); Dankers, Van den Hof, Heuberger & Bombois (2012) y Van den Hof. et al (2013). Para resolver este problema de carencia de excitación natural se debe garantizar la necesaria excitación de las entadas mediante la introducción de una señal de prueba persistentemente excitante. Según Ljung (1999), una señal cuasi-estacionaria {u(t)}, con espectro $\Phi_u(\omega)$, se dice que es persistentemente excitante de orden *n* si, para todo filtro de la forma:

$$M_n(q) = m_1 q^{-1} + \dots + m_n q^{-n}$$
(2.1)

La relación:

$$\left|M_n(e^{i\omega})\right|^2 \Phi_{\rm u}(\omega) \equiv 0 \tag{2.2}$$

Implica que:

$$\left|M_n(e^{i\omega})\right|^2 \equiv 0 \tag{2.3}$$

Por esta razón u(t) es persistentemente excitante de orden n, si $\Phi_u(\omega)$ es diferente de cero como mínimo en n puntos en el intervalo de frecuencia $-\pi < \omega < \pi$.

Las señales de excitación pueden ser introducidas por el punto de ajuste de la variable controlada o adicionadas a la(s) variable(s) manipulada(s) (Pitta. & Odloak., 2012). Entonces surgen una serie de cuestiones: qué niveles de excitación deben ser aplicados y

qué bandas de frecuencia deben ser excitadas. Algunos criterios resultan útiles para seleccionarlos. Según R. Pintelon & J. Schoukens (2001):

- Para procesos con la configuración operador-controlador, los operadores deben tener buen conocimiento de los valores aceptables.
- Para otros dispositivos los valores nominales dados en los manuales para el usuario pueden suministrar algunas indicaciones.
- Si no-linealidades significantes han sido detectadas o incluso sospechadas, y solo se intenta extraer el sistema lineal subyacente, la amplitud de la señal de excitación es hecha tan pequeña como sea posible; aunque algunas no-linealidades estén pronunciadas en este caso.
- Si no-linealidades significantes han sido detectadas o incluso sospechadas, y se está interesado en identificar la mejor aproximación lineal del sistema o identificar el modelo no-lineal que lo describe, las excitaciones deben ser representativas de la clase de excitaciones que sean aplicadas posteriormente en los dispositivos. En otras palabras, se deben seleccionar rangos de amplitud y bandas de frecuencia para las señales de excitación que cubran las bandas de frecuencia de interés.

2.2.2 Fase II: Pre-procesamiento de los datos.

Una vez que los datos han sido colectados por medio del experimento planificado, es probable que no estén en condiciones para ser usados de inmediato en el algoritmo de modelación. En opinión de Ljung (1999), en aplicaciones que lo permitan, tal como las offline, se debe inspeccionarlos siempre, en busca de deficiencias. Según el mismo autor, el pre-procesamiento de los datos debe responder a los tipos de modelos que se pretenda usar, ya sean lineales o no-lineales. En las dos siguientes secciones se aborda cómo pre-procesar los datos para evitar problemas en los procedimientos en dependencia del modelo.

2.2.2.1- Pre-procesamiento de los datos para modelos lineales de sistemas:

Varias son las posibles deficiencias contenidas en los set de datos que pueden perjudicar la tarea de extraerles los modelos paramétricos lineales. Con el objetivo de solucionarlas, el pre-procesamiento incluye: tratamiento de datos perdidos, valores atípicos (outliers), y remoción de tendencias.

Tratamiento de datos perdidos:

Mal funcionamiento del equipamiento de adquisición de datos, en ocasiones, resulta en la pérdida de algunos de ellos durante la realización del experimento, trayendo consigo brechas al representarlos. Según Ljung (1999), tratar datos perdidos tiene pocas posibilidades, una es hacer segmentos de corte en la secuencia de datos tal que las porciones malas sean evitadas, luego los segmentos buenos pueden ser juntados. En otros casos, para datos con muchas entradas y/o salidas, puede resultar difícil buscar aquellas secciones vacías. Ante estas situaciones se debe distinguir entre entradas perdidas y salidas perdidas, ya que se deben tratar de forma diferente. Para un modelo $\mathcal{M}(\theta)$ que describe la relación entre las entradas y salidas del sistema, un predictor lineal básico está dado por Ljung (1999):

$$\hat{y}(t \mid \theta) = \sum_{k=1}^{t} g(t - k, \theta) u(k) + \sum_{k=1}^{t} h(t - k, \theta) y(k)$$
(2.4)

Datos perdidos de entradas:

Es natural considerar las entradas perdidas como parámetros desconocidos. Como la expresión de las ecuación 2.4 es lineal en los datos, está claro que para un modelo dado $\mathcal{M}(\theta)$, las entradas desconocidas pueden ser estimadas usando la regresión lineal por el procedimiento de mínimos cuadrados (Ljung., 1999). Al denotarlas con el vector η se tiene:

$$\hat{y}(t \mid \theta, \eta) = \sum_{k \in K_u} g(t - k, \theta) u(k) + \varphi^T(t, \theta) \eta + \sum_{k=1}^t h(t - k, \theta) y(k)$$
(2.5)

Donde: $k \in K_u$: es el set de entradas no perdidas u(k).

 $\varphi(t,\theta)$: es hecho de $g(t-k_i,\theta)$; $k_i \notin K_u$ de una manera obvia.

Los parámetros θ y η entonces pueden ser calculados por un criterio del error de predicción por la via usal. Para valores de θ fijados, la ecuación 2.5 es una regresión lineal para η ; luego tales datos de entradas perdidas pueden ser estimados para cualquier modelo dado. Es un proceder normal, aunque no necesariamente eficiente numéricamente, iterar entra la estimación de los datos perdidos usando el modelo seleccionado (estimar η para valores de θ fijados) y la obtención del modelo θ usando los datos perdidos reconstruidos a partir del modelo seleccionado. Para comenzar las iteraciones, el primer modelo puede ser construido usando valores interpolados linealmente para los datos perdidos.

Datos perdidos de salidas:

Los dados perdidos de las salidas normalmente no son considerados como parámetros desconocidos, dado que son tratados como variables aleatorias en la estructura de la predicción. El criterio del error de predicción correcto es la minimización del error entre $y(t) \ y \ \hat{y}(t \mid \theta, Y_{K_y})$, donde la predicción es basada en aquellos datos antiguos y(t) que realmente han sido observados $k \ \epsilon \ K_y$. Para calcular esta predicción correctamente se puede hacer uso del filtro variante en el tiempo de Kalman, dado por Anderson & Moore (1979) y retomado por Ljung (1999); tratando con los datos perdidos como muestras irregulares. Según Ljung (1999), trabajar con predictores variantes en el tiempo de seguro conduce a muchos más cálculos, por lo que alternativas aproximadas puede ser preferibles, una de ellas es reemplazar cualquier valor perdido y(k) en la ecuación 2.4 por $\hat{y}(k \mid \theta)$ usando estimaciones suavizadas de y(k) empleando datos medidos hasta el tiempo t - 1. Otra aproximación es tratar además las salidas perdidas como parámetros desconocidos. Esto corresponde a reemplazar las y(k) perdidas en la ecuación 2.4 por sus estimaciones suavizadas, usando para este caso el record de datos medidos en su totalidad.

Tratamiento de valores atípicos:

Los valores atípicos representan observaciones verdaderamente aberrantes que no son representativas de la población en general, (Hair., 1999). Los atípicos pueden ser causados por una variación aguda y breve de las señales o por mal funcionamiento de las mediciones. A pesar de que la presencia de varios valores atípicos puede ser un clúster importante de datos reales dentro una población determinada, proporcionando gran información del sistema. Gujarati (2004) y Ljung (2007b) plantean que de no ser eliminados de los datos se puede ver adversamente afectado el proceso de modelación. Algunas alternativas para detectar los atípicos son:

• El método estándar para la detección de atípicos multivariantes. Una estimación robusta de los parámetros de la Distancia de Mahalanobis y compararlos con un valor crítico de la de la distribución Chi- cuadrado (χ^2) (Rousseeuw. & Van Zomeren., 1990). El hecho de recurrir a estas bien conocidas medidas de distancias se debe a que ellas toman en cuenta la matriz de covarianzas, la cual cuantifica la forma y la talla de los datos multivariantes (Sultan. & Ahmed., 2010). Además para datos multivariantes normalmente distribuidos los valores de estas distancias son

aproximados a los de la distribución Chi- cuadrado con p grados de libertad (χ_p^2); siendo p el número de variables consideradas bajo estudio (Filzmoser., 2005). Los atípicos multivariantes no pueden ser simplemente definidos como aquellas observaciones que tienen una larga Distancia de Mahalanobis al cuadrado (D²). Por esta razón un determinado cuantil de la distribución Chi – cuadrado tiene que ser considerado. Dada la naturaleza de los test estadísticos Hair (1999) sugiere que se use un nivel conservador, quizás 0.001, como valor umbral para la designación de casos atípicos.

- Representar los datos en un gráfico de tiempo e identificar los valores que parezcan no-confiables.
- Después de estimado el modelo, graficar los residuos e identificar los valores inusualmente grandes. Entonces evaluar los datos que sean responsables de los residuos largos.

Una vez identificados los valores atípicos, Ljung (2007b) propone las siguientes técnicas para eliminar o minimizar sus efectos.

- Extraer las porticiones de datos informativos en segmentos y combinarlos en un set de datos multiexperimentos.
- Manualmente reemplazar los valores atípicos de modo que se muestren como datos perdidos. Este enfoque trata los atípicos como valores perdidos.
- Eliminar valores atípicos por medio del filtrado de los datos para contenidos de alta frecuencia ya que, a menudo, resultan de cambios abruptos.

Remoción de tendencias

La remoción de tendencias en lo datos significa eliminar valores medios y tendencias lineales de las señales. No resulta útil substraer valores medios cuando se tengan datos de respuesta en estado transitorio del sistema (tales como respuestas al paso o al impulso). Cuando se trabaje con respuesta transitoria y la salida no sea igual a cero para una entrada igual a cero, se puede optar por eliminar los valores constantes correspondientes al tiempo anterior a que la entrada haya sido aplicada.

Con frecuencia, los niveles de las medias cambian lentamente de manera no-controlada de un punto a otro durante el experimento. Estos movimientos lentos indeseados pueden eliminarse removiendo una tendencia lineal o varias tendencias lineales en forma de fragmentos en presencia de datos en dominio del tiempo.

Siempre que una tendencia se substraiga de los datos usados para la estimación, se debe substraer de los datos reservados para la validación (Ljung., 2007b).

Re-muestreo de los datos.

El re-muestreo de los datos cambia la razón de muestreo de la señal por reducción o interpolación. Si los datos están muestreados más rápido que lo necesitado durante el experimento, se pueden reducir estos sin pérdida de la información. Ljung (2007b) aconseja, también reducir los datos cuando estos contengan ruido de altas frecuencias más allá de los rangos de frecuencia de la dinámica del sistema. Si por el contrario, los datos son muestreados más lento que lo requerido, existe la posibilidad de que se pierda importante información sobre la dinámica del sistema a altas frecuencias. Para mitigarlo se pueden remuestrear los datos con una mayor razón de muestreo. Los valores re-muestreados toman lugar entre las muestras medidas que representan la información tomada sobre el sistema. El re-muestreo está fuertemente ligado al resultado que se obtenga en la etapa de la planificación del experimento, producto de la selección del intervalo de muestreo T. En lugar de re-muestrear, en un sentido teórico, puede ser mucho más eficaz repetir nuevamente las corridas experimentales, sin embrago, en muchas aplicaciones prácticas esto no es viable del punto de vista técnico –económico.

Combinación de set de datos:

Es una situación muy común en la práctica que se tenga que manejar un número de corridas experimentales separadas. La razón puede ser que la planta no está disponible para una corrida experimental larga de manera continua, o que solo es permitido manipular una entrada a la vez en corridas separados. Otra de las razones es cuando los datos de una corrida son analizados para probar su conveniencia en el proceso de modelación, pues con frecuencia ocurre que hay porciones de datos malos, información no relevante y pérdidas de datos, los cuales pueden ser difíciles o computacionalmente costosos de reconstruir. La presencia de estas situaciones presiona a tener que seccionar los datos recopilados en varios segmentos separados. Para trabajar con todos estos segmentos separados, en opinión de Ljung (1999) no se puede simplemente concatenarlos, pues los puntos de conexión provocan situaciones transitorias que pueden destruir la estimación.

Una solución es, para un caso general, con modelos de predictores que tengan infinitas respuestas impulso, la sugerida por el propio Ljung (1999); la misma está formada de la siguiente manera:

$$V(\theta) = \sum_{t \in T^1} (y(t) - \hat{y}(t \mid \theta))^2 + \dots + \sum_{t \in T^n} (y(t) - \hat{y}(t \mid \theta))^2$$
(2.6)

Donde T^1 es un set indexado de los *i*-ésimos segmentos, excluyendo aquellos *t* para los cuales $\varphi(t)$ no son completamente conocidos. Aquí los filtros que calculan $\hat{y}(t \mid \theta)$ para cada uno de los segmentos deber ser reinicializados con cero como condición inicial (o asociados con un set separado de condiciones iniciales que sean estimadas). La ecuación 2.6 para lograr su solución, se basada en un algoritmo de minimización con $V(\theta)$ como función criterio a ser minimizada.

2.2.2.1- Pre-procesamiento de los datos para modelos no-lineales de sistemas:

El pre-procesamiento de los datos para obtener modelos no-lineales puede realizarse de una forma análoga a cuando se trata de obtener modelos lineales para los casos de interpolación de valores perdidos, la combinación de set de datos y el re-muestreo usando diferentes intervalos de tiempo. Sin embargo, no se deben remover tendencias antes de buscar modelos no-lineales ya que se convierte en un tema problemático y puede ser que tengan que estar integradas dentro del modelo identificado Ljung (2007b); Wigren & J. Schoukens (2013).

2.2.3 Fase III: Selección de un set de modelos:

La selección de una estructura de modelo apropiada \mathcal{M} es crucial para el éxito de la modelación matemática. Esta selección debe estar basada tanto en el conocimiento de los procesos de modelación, como en una clara percepción y conocimiento del sistema que se desea modelar, además de tener bien concebido el fin para el cual se pretende el modelo. Un modelo determinado puede ser de los mejores disponibles, pero la una cuestión fundamental es si realmente sea lo suficiente bueno para la intención propuesta (Ljung., 1999).

2.2.3.1-Selección según linealidad.

Un aspecto general, pero a la vez fundamental, es la selección entre modelos lineales y modelos no-lineales. Aquí entra a jugar el criterio de Ljung (1999) sobre el objetivo de los usuarios, el cual debe ser "obtener un buen modelo a un bajo precio" La selección del

modelo ciertamente tiene un efecto considerable tanto en la calidad del resultado como en el precio que se pague por conseguirlo.

La calidad del modelo está basada en que las distorsiones del resultado estadístico por la exclusión de variables y la varianza sean pequeñas (Ljung., 1999). Según este autor, tal situación naturalmente genera un conflicto, ya que para reducir tales distorsiones básicamente se tienen que emplear modelos más grandes y más flexibles, requiriéndose mayor cantidad de parámetros; luego la varianza típicamente incrementa con el incremento del número de parámetros estimados. Por estas razones, según el criterio de Ljung (1999), para conseguir un buen modelo se debe lograr una especie de compromiso entre la flexibilidad y la parsimonia, lo cual constituye el meollo de los problemas de modelación.

El precio del modelo es asociado con el esfuerzo requerido para su cálculo. Calcular un modelo depende fuertemente de la estructura en particular y en él influyen la complejidad del algoritmo y las propiedades de la función criterio. Existe además un precio asociado con el uso del modelo, modelos complejos de orden elevado son más difíciles de usar para la simulación y el diseño de control.

Por todas estas razones, Ljung (1999) aconseja tratar siempre las cosas simples primero; debiendo irse hacia los modelos de estructuras sofisticadas solo si las más simples no pasan las pruebas de validación. Aunque ante determinadas aplicaciones no se esté en presencia de un sistema lineal, los métodos lineales son sin embargo una buena y primera selección para un problema de modelación; pues además conducen a esquemas robustos y simples (Ljung., 1999).

A pesar de la conveniencia de la modelación lineal, un problema particular forma una excepción; la prueba de efectos no-lineales. Esto es, según Ljung (1999), verificar si es probable que los datos puedan ser explicados por una relación lineal o si un modelo de estructura no-lineal es requerido. Tales pruebas están basadas en la relación entre las correlaciones de elevado orden (superior al segundo) y el espectro que sigue de una descripción lineal. En presencia del segundo caso, tomar una decisión racional es optar por los modelos no-lineales, pese a las dificultades que esto trae.

2.2.3.2- Selección según el tipo de parametrización.

Al tratarse de la parametrización de sistemas físicos, se presentan variantes en cuanto a la selección del tipo de modelo. En las situaciones que se disponga de la adecuada percepción

y conocimiento de las leyes físicas que describen al sistema, es posible construir modelos fundamentados físicamente, a partir de modelos de parametrización física. Si las condiciones son diferentes, y solo fuera posible poco más que acceder a corridas de datos de las variables de entradas y salidas del sistema, se deben usar modelos de parametrización de caja negra. Aspectos como la complejidad del algoritmo y la forma de la función criterio son favorecidos en modelos obtenidos mediante caja negra, los cuales poseen flexibilidad para adaptarse a los parámetros de los datos, sin la posibilidad de una interpretación física de sus valores, lo cual constituye una de sus desventajas según Ljung (1999). Sin duda alguna estas alternativas son un problema dependiente de la aplicación.

2.2.3.2- Selección según la intención de uso.

Con respecto a seleccionar modelos en dependencia de la intención de uso, se pueden sugerir algunas ideas. Por ejemplo, Ogata (1998) plantea que en dependencia del sistema que se trate y las circunstancias específicas, un modelo matemático puede ser más provechoso que otros. De esta forma el autor mencionado sugiere que en problemas de control óptimo, es más útil usar representaciones en el espacio de estados; mientras que en presencia de análisis de respuestas transitorias o de la respuesta en frecuencia de sistemas lineales e invariantes con el tiempo, la representación mediante funciones transferenciales puede ser muy conveniente. Retomando la idea del tipo de parametrización, los modelos físicos son útiles cuando el usuario necesita conocer determinados coeficientes de las leyes y fenómenos de la física que tienen lugar en el proceso, en cambio, los modelos de caja negra son una representación matemática del sistema, útiles a considerar cuando la inamica del proceso y del ruido como tal.

2.2.4 Fase IV: Obtención del modelo.

Una vez que un set de modelos candidatos ha sido seleccionado, la búsqueda del mejor modelo dentro de este conjunto se convierte en el problema de determinar o calcular el vector de parámetros θ ; por supuesto, teniendo presente obtener un buen modelo a un bajo precio (compromiso entre la flexibilidad y la parsimonia). Para lograrlo se debe usar conocimiento previo del sistema, intuición e ingenuidad; además de una serie de herramientas de cálculo. Este hecho presiona a que la modelación matemática no pueda ser tratada netamente con procedimientos completamente automatizados según criterio de Ljung (1999). Sin embargo, para lograr la modelación de un sistema se debe buscar apoyo en programas computacionales tales como el paquete de software estándar MATLABTM System Identification Toolbox Ljung., (2007b); desarrollado por MathWorks, Inc. y uno de los más utilizados en opinión de Wigren & J. Schoukens (2013); el paquete CONTSID Toolbox a criterio de Pitta & Odloak (2012); entre otros.

2.2.4.1- Obtención de modelos lineales.

Retomando la idea que ha sido seleccionado un cierto modelo de estructura \mathcal{M} , con modelos particulares $\mathcal{M}(\theta)$ que representan una forma de predecir las salidas futuras, y tratando las cosas simples primero, el predictor puede ser un filtro lineal de la forma:

$$\mathcal{M}(\theta): \ \hat{y}(t \mid \theta) = w_y(q, \theta)y(t) + w_u(q, \theta)u(t)$$
(2.7)

Como se ha colectado un conjunto de datos del sistema:

$$Z^{N} = [y(1), u(1), y(2), u(2), \dots, y(N), u(N)]$$
(2.8)

El problema que se encara es decidir cómo usar la información contenida en Z^N para seleccionar los valores adecuados $\hat{\theta}_N$ del vector de parámetros, y en este sentido un miembro adecuado $\mathcal{M}(\hat{\theta}_N)$ en el set \mathcal{M}^* (usualmente generado como un rango de estructura de modelo).

$$\mathcal{M}^* = \left\{ \mathcal{M}(\widehat{\theta}_N) \mid \theta \in D_M \right\}$$
(2.9)

Con D_M set de valores sobre los cuales θ se mueve en una estructura de modelo.

Se tiene que determinar una función de los datos Z^N del set D_M . Tal función es el resultado de aplicar métodos de estimación de parámetros. Una prueba para evaluar la habilidad de diferentes modelos de estructuras para describir los datos observados está basada en el error de predicción dado por un cierto modelo $\mathcal{M}(\theta_*)$.

$$\varepsilon(t,\theta_*) = y(t) - \hat{y}(t \mid \theta_*) \tag{2.10}$$

Un buen modelo es aquel que sea capaz de predecir, esto es, uno que produzca errores de predicción pequeños cuando sea aplicado a datos observados (Ljung., 1999). En cuanto a la cuestión de cómo calificar lo que puede significar pequeño, pueden ser utilizados dos enfoques. Uno es formar una norma valorada de manera escalar o una función criterio que mida el tamaño del error. El otro enfoque exige que $\varepsilon(t, \theta_*)$ sea no-correlacionado con una secuencia de datos dados. Tales enfoques conllevan a la aplicación de métodos de estimación de parámetros, métodos que son seleccionados dependiendo de la aplicación que se trate, según su aplicabilidad y la facilidad de cálculo. Debido a las dificultades en la

modelación a lazo cerrado, varios de los métodos del error de predicción, vistos en el Capítulo I de la presente investigación, no deben ser directamente aplicados, en su lugar variantes alternativas son sugeridas Ljung (1999) y Van den Hof. et al (2013), las mismas se referencian a continuación:

<u>Identificación directa:</u> para la minimización del error de predicción con énfasis en incluir la identificación de modelos exactos de ruidos/disturbios.

<u>Identificación indirecta:</u> tales como el método de dos-fases y el método de IV, con énfasis en el uso de excitación externa o señales de pruebas.

<u>Identificación conjunta de entradas-salidas (IO)</u>: donde las señales de entradas y salidas son modeladas como las salidas de un proceso estocástico estacionario.

Los diferentes algoritmos y las condicones típicas que deben existir para lograr la consistencia en la estimación a partir de estos métodos son descritas por (Van den Hof. et al., 2013).

2.2.4.2- Obtención de modelos no-lineales.

En aquellas situaciones en las que la dinámica del sistema sea no-lineal y el enfoque lineal mencionado anteriormente no sea capaz de conducir a un buen modelo aproximado del sistema, o incluso, cuando las esencias de la dinámica del fenómeno no-lineal estén involucradas y el usuario necesite capturar estas dinámicas en el modelo; interesantes modelos de estructuras no-lineales pueden ser utilizados.

En estos casos el predictor puede ser un filtro no-lineal, que escrito como una función general de datos antiguos Z^{t-1} queda de la siguiente forma:

 $\mathcal{M}(\theta): \quad \hat{y}(t \mid \theta) = g(t, Z^{t-1}; \theta) \tag{2.11}$

Al decidir cómo usar la información contenida en Z^N para seleccionar los valores adecuados $\hat{\theta}_N$ del vector de parámetros, y en este sentido aplicar métodos de estimación de parámetros, se puede optar en algunos casos por aplicar métodos de estimación del error de predicción de una manera análoga a como se hace en el caso de modelos lineales. Este hecho es posible debido a que en la teoría de la estimación las propiedades asintóticas y los algoritmos básicos son los mismos para algunos modelos lineales y no-lineales (Ljung., 1999). Sin embargo se recomienda aplicar algoritmos y métodos que son particulares para los modelos no-lineales. Los mismos pueden ser seleccionados en cuanto a criterios como las características de la no-linealidad, el tipo de modelo compatible y la facilidad de tratar con las entradas del sistema, o sea, dependientes de la aplicación en cuestión. La siguiente tabla resume estos aspectos para los múltiples estimadores no-lineales contenidos en el paquete de software estándar MATLABTM System Identification Toolbox según Ljung (2007b).

Estimadores según		Facilidad para
características de la no-	Tipo de modelo compatible	tratar con
linealidad		entradas múltiples
Red personalizada	Hammerstein-Weiner & Polinomiales	Si
Zona muerta	Hammerstein-Weiner	No
Red neuronal	Polinomiales	Si
Linealidad por secciones	Hammerstein-Weiner	No
Saturación	Hammerstein-Weiner	No
Redes en forma de "S"	Hammerstein-Weiner & Polinomiales	Si
Partición triple	Polinomiales	Si
Redes onduladas	Hammerstein-Weiner & Polinomiales	Si

Tabla 2.1. Estimadores para modelos no-lineales. Fuente: (Ljung., 2007b).

2.2.5 Fase V: Validación del modelo.

Los procedimientos de estimación de parámetros seleccionan el "mejor" modelo dentro del modelo de estructura elegido. Un punto clave entonces es verificar si este "mejor" modelo es lo suficientemente bueno, lo que constituye el problema de la validación Ljung (1999).

El enfoque más natural de buscar un modelo de estructura que sea apropiado, según las necesidades de los usuarios, es simplemente probar con varios de ellos y luego comparar los modelos resultantes Ljung (1999); Pillonetto & De Nicolao (2011). Cuando varios modelos hayan sido obtenidos mediante los métodos seleccionados, se debe decidir qué comparar entre ellos y cómo evaluar estas comparaciones. Un modelo a evaluar es genéricamente denotado por $m = \mathcal{M}(\hat{\theta}_N)$, el cual pertenece al tipo de estructura \mathcal{M} . Del mismo se supone que tiene una dimensión del vector fila θ (dim θ) igual a d_{\mathcal{M}} parámetros libres. Apuntando al objetivo de la validación, las siguientes herramientas pueden ser de útil aplicación; permitiendo desechar modelos, o bien, aumentar la confianza en ellos.

Selección del tamaño de un set de modelos:

Comprende temas como la selección del orden de modelos en el espacio de estados, el grado de modelos polinomiales y el número de neuronas en una red neuronal; contiene además el problema de cuáles variables incluir en la descripción del modelo. Como tal selecciona \mathcal{M} de una determinada cadena de estructuras en incremento de la forma:

$$\mathcal{M}_1 \subset \mathcal{M}_2 \subset \mathcal{M}_3 \dots \tag{2.12}$$

Tal selección entre \mathcal{M}_0 y \mathcal{M}_1 , sujeta a la ecuación 2.12, puede ser enfocada por la teoría de las pruebas estadísticas mediante el contraste de las hipótesis (Ljung., 1999):

Hipótesis nula: H_o : los datos han sido generados por $\mathcal{M}_0(\hat{\theta}_N^{(0)})$ (2.13)

Hipótesis alternativa: H_1 : los datos han sido generados por $\mathcal{M}_1(\hat{\theta}_N^{(1)})$ (2.14)

Suponiendo que \mathcal{M}_1 sea más largo que \mathcal{M}_0 , se debe introducir un prejuicio contra \mathcal{M}_1 . Esto significa que se prefiere \mathcal{M}_0 a no ser que haya evidencias convincentes de que la hipótesis alternativa H_1 sea cierta. Se procede tomando una decisión entre H_0 y H_1 tal que el riesgo (la probabilidad) de rechazar H_0 cuando en realidad sea cierta (error de tipo I) sea menor que un cierto número α . En la medida que α sea seleccionado más pequeño, un mayor prejuicio es infligido contra \mathcal{M}_1 . Al mismo tiempo es favorable maximizar la probabilidad de que H_0 sea rechazada cuando H_1 sea cierta, lo cual es conocido como la potencia del test.

Prefijando un valor $\hat{\theta}_*^{(k)}$ que signifique el límite estimado en el modelo de estructura \mathcal{M}_k ; si este valor da una descripción correcta del sistema ($S \in \mathcal{M}_k$), puede ser mostrado bajo condiciones generales que:

$$N \cdot \frac{V_N(\theta_*^{(k)}, z^N) - V_N(\hat{\theta}_N^{(k)}, z^N)}{V_N(\theta_*^{(k)}, z^N)} \in As\chi^2(d(k))$$
(2.15)

O sea que la variable aleatoria del miembro izquierdo converge a la distribución *Ji* cuadrado (χ^2) con d(k) grados de libertad (Ljung., 1999).

$$d(k) = \dim \theta^{(k)} \tag{2.16}$$

Sujeto a la hipótesis nula en 2.13 y deducido de la ecuación 2.15 se tiene que:

$$N \cdot \frac{v_N(\hat{\theta}_N^{(0)}, z^N) - v_N(\hat{\theta}_N^{(1)}, z^N)}{v_N(\hat{\theta}_N^{(1)}, z^N)} \in As\chi^2(d(1) - d(0))$$
(2.17)

Luego, según Ljung (1999), la hipótesis nula puede ser probada para cualquier nivel de significación (α) usando la expresión 2.17.

Criterio del error final de predicción (FPE) de Akaike.

El criterio del error final de predicción de Akaike (FPE –siglas en inglés) proporciona una medida de la calidad del modelo para la simulación en situaciones donde sea probado para set de datos diferentes. Una vez que hayan sido estimados varios modelos; incluso de diferentes tipos de estructuras, estos pueden ser comparados usando tal criterio Ljung (2007b). Acorde a la teoría de Akaike, el modelo más exacto posee el menor valor FPE. El criterio está definido por la siguiente ecuación:

$$FPE = \bar{J}_p(\mathcal{M}) \approx \frac{1 + (d_{\mathcal{M}}/N)}{1 - (d_{\mathcal{M}}/N)} V_N(\hat{\theta}_N, Z^N) = \frac{1 + (d_{\mathcal{M}}/N)}{1 - (d_{\mathcal{M}}/N)} \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t, \hat{\theta}_N)$$
(2.18)

Muestra cómo modificar la función de pérdidas llega a un estimado razonable de la validación y el criterio de comparación \overline{J} , solo a partir de datos de estimación. El ajuste observado debe ser compensado por el uso del número de parámetros estimados para dar un cuadro justo de la calidad del modelo (Ljung., 1999).

El valor de FPE puede ser negativo cuando el número de parámetros estimados excede al número de muestras en los datos, lo cual puede ocurrir para modelos con múltiples salidas. Para los casos de salidas múltiples la suposición de que d_M/N es pequeña no es válida. En presencia de estos casos se recomienda usar el criterio de información, también de Akaike (Ljung., 2007b).

Criterio de información (AIC) de Akaike.

El criterio de información de Akaike (AIC –siglas en inglés) proporciona una medida de la calidad del modelo para la simulación en situaciones donde sea probado para set de datos diferentes. Ljung (2007b) lo recomienda emplear para llevar a cabo comparaciones relativas de modelos con diferentes estructuras. Un valor más pequeños del criterio indica un mejor modelo Gujarati (2004) y Ljung (2007b).

AIC está definido por la siguiente ecuación:

$$AIC = logV_N(\hat{\theta}_N, Z^N) + \frac{2d_{\mathcal{M}}}{N}$$
(2.19)

Para valores de $d_{\mathcal{M}} \ll N$ queda de la siguiente forma:

$$AIC = \log\left(V_N(\hat{\theta}_N, Z^N) + \left(1 + \frac{2d_{\mathcal{M}}}{N}\right)\right)$$
(2.20)

Intervalos de confianza de los parámetros.

Un procedimiento para chequear si un determinado modelo contiene demasiados parámetros es comparar los valores estimados con la correspondiente desviación estándar estimada (Ljung., 1999). Si los intervalos de confianza contienen el valor cero, se puede considerar si el parámetro debe ser removido. Este es usualmente de interés solo cuando los correspondientes parámetros reflejan una estructura física; tal como orden del modelo y retardo de tiempo. Según Ljung (1999), si las desviaciones estándar estimadas son todas largas, la matriz de información está cerca de ser singular, lo cual es una indicación de órdenes del modelo demasiado elevados.

Comparación de modelos con set de datos nuevos.

Como es lógico, un modelo debe ser capaz de reproducir los datos de estimación, o sea aquellos que son utilizados para su estimación. Una prueba real es, si además, es capaz de describir datos de validación, formados por un conjunto de datos disponibles que no han sido usados para construir ninguno de los modelos que se pretenden evaluar. Es importante, según Hair (1999) y Ljung (1999), que estos datos hayan sido colectados en situaciones que estén cercanas a las condiciones de operación deseadas y asegurarse de que las condiciones medias durante el tiempo que se toman las muestras no hayan cambiado. Una forma atractiva y herramienta principal de comparación de dos modelos m_1 y m_2 es evaluar su desempeño al reproducir datos de validación, en términos de simulación y predicción. Los modelos pueden ser evaluados tanto por inspección visual de los gráficos de y(t) e $\hat{y}_k(t | m)$, o por medio del valor numérico de la suma de cuadrados medios de la predicción $(J_k(m))$ (Ljung., 1999).

$$J_k(m) = \frac{1}{N} \sum_{t=1}^{N} |y(t) - \hat{y}_k(t \mid m)|^2$$
(2.21)

La medida de calidad $J_k(m)$ depende del record de datos reales para el cual se hace la comparación. Es natural además considerar el valor esperado de esta medida, donde la expectación es tomada con respecto a los datos. Como $m = \mathcal{M}(\hat{\theta}_N)$ es en sí misma una variable aleatoria, siendo estimada del ruido de los datos, el valor esperado del ajuste del modelo con respecto a $\hat{\theta}_N$ da una medida de la calidad para la estructura del modelo \mathcal{M} , quedando expresado como:

$$\bar{J}_k(\mathcal{M}) = E\bar{J}_k\left(\mathcal{M}(\hat{\theta}_N)\right)$$
(2.22)

Al ser útil dar alguna medida normalizada de este ajuste, y asegurando que en las salidas y(t) hayan sido removidas las tendencias de los valores medios iguales a cero, Ljung (1999) define:

$$R^{2} = 1 - \frac{J_{k}(m)}{\frac{1}{N}\sum_{t=1}^{N}|y(t)|^{2}}$$
(2.23)

Siendo *R* la parte de la variación de la salida del sistema que no es explicada por el modelo y puede ser expresada en por ciento.

Con respecto a tomar la decisión, se favorece entonces al modelo que muestre mejor desempeño, o sea, mayor por ciento de varianza explicada por el modelo (R^2) Ljung (1999) y Gujarati (2004). Tal procedimiento es conocido como validación cruzada. Un rasgo atractivo del mismo es su carácter pragmático: la comparación tiene sentido sin ningún argumento probabilístico y sin ninguna suposición sobre el sistema real. También es aplicable al comparar modelos de estructuras diferentes. Una desventaja es que se tienen que guardar datos frescos para validar (datos de validación), teniendo presente el elevado costo de obtener la información en la mayoría de los casos (Ljung., 1999). Otros problemas de utilizar R^2 según Gujarati (2004) son: no se garantiza que este pueda predecir observaciones que se alejen de la muestra para la cual fue estimado; al comparar dos o más R^2 la variable de salida debe ser la misma; y lo más importante, R^2 no puede disminuir cuando más variables sean adicionadas al modelo, sin embargo, no puede ser maximizado por la simple adición de más variables ya que se puede incrementar la varianza del error de predicción.

Análisis de residuos.

Se pueden validar modelos paramétricos lineales y no-lineales chequeando la conducta de los residuos, ya que estos representan la porción de los datos de validación no explicados por el modelo (Ljung., 2007b). Según este mismo autor el análisis de residuos consiste de dos pruebas: la prueba de Whiteness y la prueba de independencia.

Prueba de Whiteness:

El hecho de que exista correlación entre los mismos residuos, por ejemplo, si los valores de $\hat{R}_{\varepsilon}^{N}(\tau)$ no son pequeños para $\tau \neq 0$, entonces parte de $\varepsilon(t)$ puede estar siendo predicha por los datos antiguos, lo cual es un signo de deficiencia del modelo (Ljung., 1999).

$$\widehat{R}_{\varepsilon}^{N}(\tau) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t) \varepsilon(t-\tau)$$
(2.24)

Los $\hat{R}_{\varepsilon}^{N}(\tau)$ también comprenden información relacionada con el hecho de que los residuos puedan ser considerados como blancos. Para tener una idea de lo grande que estos valores pueden ser si ciertamente los $\varepsilon(t)$ son blancos, se supone que son una secuencia de ruido blanco con media igual a cero y varianza λ . Entonces, según Ljung (1999), estos convergen a la distribución normal:

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \begin{bmatrix} \varepsilon(t-1) \\ \vdots \\ \varepsilon(t-M) \end{bmatrix} \varepsilon(t) \in AsN(0, \lambda^2 \cdot I)$$
(2.25)

Como se asume que los errores del modelo son blancos, consecuentemente esto significa que pueden ser asintóticamente distribuidos por la función *Ji* cuadrado (Ljung., 1999).

Entonces la prueba de Whiteness como tal es verificar si el vector de correlación $(\zeta_{N,M})$ pasa la prueba de ser *Ji* cuadrado distribuido $\chi^2(M)$, al probar si $\zeta_{N,M} < \chi^2_{\alpha}(M)$, siendo α el nivel de confianza de la distribución $\chi^2(M)$.

$$\zeta_{N,M} = \frac{N}{\left(\hat{R}_{\varepsilon}^{N}(0)\right)} \sum_{\tau=1}^{M} \left(\hat{R}_{\varepsilon}^{N}(\tau)\right)^{2}$$
(2.26)

Acorde al criterio de este test, un buen modelo tiene la función de auto-correlación de los residuos ubicada entre sus intervalos de confianza, indicando que los residuos son no-correlacionados. Los intervalos de confianza corresponden a rangos de valores con probabilidad de ser estadísticamente insignificantes para el sistema (Ljung., 2007b). Prueba de independencia.

Es importante que la covarianza entre residuos y entradas pasadas sea pequeña. Cuando esto no ocurre, es una señal de que probablemente existan trazas de las entradas en los errores, entonces hay una parte de y(t) que no está predicha por los datos pasados, o sea, y(t) no ha sido verdaderamente explicada por el modelo m (Ljung., 1999). En este sentido el modelo puede ser mejorado.

Asumiendo que las salidas de un modelo probabilístico para un sistema dinámico son:

CAPÍTULO 2: Procedimiento para modelar sistemas en procesos industriales a gran escala.

$$y(t) = g(t, Z^N; \theta_N) + \varepsilon(t)$$
(2.25)

Donde $\varepsilon(t)$ tiene la propiedad de ser independientes para cada uno de los datos antiguos y puede ser obtenido por medio de:

$$\varepsilon(t) = y(t) - g(t, Z^N; \theta_N)$$
(2.26)

Luego la cuestión de validación del modelo está dada por si es probable que el record Z^N en realidad haya sido generado por la ecuación 2.25. En opinión de Ljung (1999), esta situación es equivalente a si es probable que $\varepsilon(t)$ sea una secuencia de variables aleatorias independientes con función de densidad de probabilidad (PDF) $f_e(x, t; \hat{\theta}_N)$.

Por las razones antes vistas, es importante que los residuos no dependan de las entradas particulares usadas en Z^N . Para chequearlo, según Ljung (1999), es razonable estudiar la covarianza entre residuos y entradas pasadas $\hat{R}_{\varepsilon u}^N(\tau)$.

$$\widehat{R}_{\varepsilon u}^{N}(\tau) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t) u(t-\tau)$$
(2.27)

Si estas magnitudes son pequeñas y se puede justamente cuantificar lo que puede significar pequeñas, existen razones para creer que el modelo nunca produce residuos mayores que S_1 o que un error medio S_2 . Entonces S_1 y S_2 puede tener relevancia, además, cuando el modelo sea aplicado a otras entradas.

$$S_1 = \max_t |\varepsilon(t)|$$
 $S_2^2 = \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t)$ (2.28)

Según (Ljung., 1999), un simple uso de la ecuación 2.27 es considerar que para un valor de τ dado, se cumple que $\sqrt{N}\hat{R}_{\varepsilon u}^{N}(\tau)$ converge a la distribución normal con media igual a cero y matriz de covarianza P_1 :

$$\sqrt{N}\hat{R}_{\varepsilon u}^{N}(\tau) \in AsN(0, P_{1}) \qquad P_{1} = \sum_{k=-\infty}^{\infty} R_{\varepsilon}(k)R_{u}(k) \qquad (2.29)$$

Con el objetivo de verificarlo, las siguientes hipótesis son planteadas:

Hipótesis nula: H_o : los datos analizados siguen una distribución normal(2.30)Hipótesis alternativa: H_1 : los datos analizados no siguen una distribución normal(2.31)El criterio para la toma de decisión entre las dos hipótesis, según Ljung (1999), se obtienechequeando:

$$\left|\hat{R}_{\varepsilon u}^{N}(\tau)\right| \leq \sqrt{\frac{P_{1}}{N}} N_{\alpha} \quad \text{No existen evidencias para rechazar de } H_{o};$$

$$\left|\hat{R}_{\varepsilon u}^{N}(\tau)\right| \geq \sqrt{\frac{P_{1}}{N}} N_{\alpha} \quad \text{por lo tanto } \varepsilon(t) \text{ y } u(t-\tau) \text{ son independientes.}$$

$$\left|\hat{R}_{\varepsilon u}^{N}(\tau)\right| \geq \sqrt{\frac{P_{1}}{N}} N_{\alpha} \quad \text{por lo tanto } \varepsilon(t) \text{ y } u(t-\tau) \text{ no son independientes.}$$

$$(2.32)$$

Donde N_{α} denota el nivel de significación de la distribución normal para una media igual a cero y varianza igual a uno N(0,1).

Una forma atractiva de realizar esta prueba es graficando $\hat{R}_{\varepsilon u}^{N}(\tau)$ en función de τ (Ljung., 1999). Como P_1 en las ecuaciones 2.32 y 2.33 no dependen de τ , los límites de confianza son líneas horizontales. El gráfico obtenido da una percepción valiosa de la exactitud del modelo de estructura bajo estudio.

Según, los modelos deben pasar tanto las pruebas de Whiteness como las de independencia, excepto en los casos siguientes:

- Para modelos de errores de salida y cuando se use el método de las variables instrumentales. En estos casos la modelación se enfoca en la dinámica de $G_0(q)$ y no en las propiedades de los disturbios $H_0(q)$; por lo que se debe asegurar que los modelos muestren independencia entre las entradas y los errores, y prestar menos atención al resultado de la prueba de Whiteness.
- Correlación entre residuos y entradas para períodos negativos, esto no es necesariamente una indicación de un modelo inadecuado. Cuando un residuo actual en el instante de tiempo t afecta el valor de una entrada fututa, puede ser una señal de retroalimentación en el sistema. En este caso hay que concentrarse en los períodos positivos en los gráficos de correlación cruzada durante la validación del modelo.

La consecuencia de aplicar estas herramientas a los modelos estimados debe poner en condición al usuario de decidir cuál de los modelos manejados es el más apropiado para sus necesidades. La decisión a tomar no puede ser, bajo ninguna circunstancia, una pura selección de números de por cientos en lo relativo al desempeño del modelo para explicar datos de validación. Recomendable es tener en cuenta, de conjunto, aspectos como haber pasado las pruebas de los residuos, la inclusión de variable que aporten valor a la

investigación, complejidad y racionalidad del modelo e intención de uso del mismo; y por supuesto, buen ajuste en la explicación de nuevos valores.

En el caso de que no se arribe a un modelo adecuado por la presencia de características inesperadas y/o indeseables, se deben retomar algunas de las fases previas del procedimiento. Según Ljung (2007b), deficiencias como las citadas a continuación probablemente hayan influido en los malos resultados.

- Pobre excitación de las señales de entrada.
- El orden del modelo es demasiado elevado o bajo.
- El tipo de modelo no es capaz de representar los ruidos sustanciales del sistema.
- Estimadores no-lineales producen pobre ajuste.

2.3 Conclusiones del capítulo.

- 1- La modelación de sistemas a lazo cerrado parece ser un requisito necesario para procesos complejos, obliga a disponer de métodos específicos de estimación para lograr resultados convincentes.
- 2- Los modelos matemáticos deben ser capaces de logar la intensión de uso que con ellos se pretende y hacerlo además de la forma más sencilla posible.
- 3- El proceso de validación de un modelo debe ser una decisión multicriterio, donde elementos como: el análisis de residuos, ajuste, complejidad y racionalidad del modelo, juegan un papel importante y no solo sea una decisión en función de porcentajes de ajustes.
- 4- El procedimiento diseñado consta de cinco fases, planificación del experimento, pre-procesamiento de los datos, selección de un set de modelos, obtención de modelos y la validación de los mismos, además posee un enfoque cíclico.

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

CAPÍTULO 3. Implementación del procedimiento para modelar el sistema del tope la columna desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos. Introducción.

El capítulo presenta los resultados obtenidos en el proceso de modelación matemática del sistema en estudio, el tope de la columna desbutanizadora T-401/1, perteneciente a la Refinería de Petróleos de Cienfuegos. Trabajando sobre la base de un conjunto de variables de entradas y salidas, y siguiendo el procedimiento desarrollado en el capítulo anterior, se determinan varios modelos de acuerdo a los objetivos propuestos. Posteriormente se someten a pruebas de validación, las cuales permiten establecer la comparación entre ellos atendiendo a diferentes criterios; con vista a seleccionar un modelo apropiado para la simulación del proceso.

3.1 Fase I: Planificación del experimento.

El sistema que se pretende modelar matemáticamente, ver figura 3.1, forma parte de un proceso industrial a gran escala que opera en configuración de lazo cerrado. El mismo ha sido descrito en la sección 2.1 del capítulo anterior.



Figura 3.1: Sistema a identificar para la estrategia de control que se pretende. Fuente: (Alonso Díaz. et al., 2012)

El desarrollo de pruebas a lazo abierto de larga duración, dada las condiciones de operación, aumenta el riesgo de incurrir en graves errores que ponen en peligro la industria y sus estándares de calidad según Dankers. et al (2012) y Pitta & Odloak (2012); así como las personas y el medio ambiente. Por tal motivo, se requiere de permisos espaciales y una rigurosa preparación y supervisión por parte de los operadores y técnicos de instrumentación para desarrollar pruebas abriendo lazos de control. Producto de esta situación en las corridas experimentales realizadas con el fin de la modelación del sistema

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 71 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

se cuenta con la presencia de lazo cerrado. Siguiendo el criterio de Hair (1999); Schoukens, Pintelon & Rolain (2006) y T. Chen, Ohlsson & Ljung (2012), los datos obtenidos durante las corridas experimentales se dividen a la mitad, el primer set con objetivo de la estimación de los modelos y segundo se reserva como datos frescos para la validación.

3.1.1 Selección de las señales a medir.

El control de temperatura del tope de la columna es fundamental, a través de él se puede modificar la composición (calidad) del gas producto GLP (Alonso Díaz. et al., 2012), por lo que se decide como variable de salida la temperatura en este punto de la columna. Al abrir el lazo correspondiente al TIC, las señales que influyen de forma directa en la sección estudiada son, el punto de ajuste del FIC generado por el TIC (spFIC) y la retroalimentación del FIC (feedback); valores escalares y medibles en sendos casos. Al seleccionar estas dos señales como variables de entrada la configuración a modelar es considerada como un sistema de múltiples entradas y salida simple (MISO siglas en inglés). Teniendo en cuenta que se muestrea además otra serie de variables con influencia indirecta en el sistema tratado, si en etapas posteriores del proceso de modelación se comprueba que los modelos obtenidos no son lo suficientemente explicativos, pueden tomarse como nuevas variables de entradas.

3.1.2 Selección del intervalo de muestreo.

El sistema de control distribuido (DCS) versión CENTUM CS 3000 R3, realizado por Yokogawa e implementado en la columna desbutanizadora, como cualidad programada toma muestras cada un segundo. El tiempo de muestreo total de las corridas experimentales $0 \le t \le T_N$ es limitado, producto de las condiciones de operación, pero la adquisición de los datos dentro de este tiempo es poco costosa. Ante esta situación se decide tomar T = 1scomo intervalo de muestreo, siendo ventajoso desde un punto de vista basado en la teoría de la información, muestrear tan rápido como sea posible (Ljung., 1999). El hecho de medir las variables solo en ciertos instantes de tiempo (cada un segundo) resulta en una aplicación de tiempo discreto.

3.1.3 Selección de las señales a manipular.

Al existir la configuración de lazo cerrado es de vital importancia excitar las señales de entrada en la medida de lo posible. A pesar de la presencia de disturbios que afecten al sistema durante la operación normal, usualmente estas exitaciones no son suficientes para

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 72 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

permitir la correcta modelación del proceso (Ogata., 1998); (Ljung., 1999); (R. Pintelon. & J. Schoukens., 2001), (Pitta. & Odloak., 2012); (Dankers. et al., 2012) (Van den Hof. et al., 2013). Es permitido solo interrumpir las acciones del controlador automático TIC que genera puntos de ajuste dinámicos al FIC, pues esta es una práctica que a menudo se realiza durante la producción por las limitaciones del TIC (largos períodos de estabilización) que comprometen el resto de los parámetros operacionales en opinión de Alonso Díaz. et al (2012). Para resolver el problema de la carencia de excitación natural, se utiliza la variante recomendada por Pitta & Odloak (2012), la cual consiste en adicionar una señal de excitación a la variable manipulada. Al haberse sospechado no-linealidades significantes para sistemas como el analizado S. Chen. et al (1989); R. Pintelon. et al (2003); Tóth. et al (2009) y (Sakthivel. et al., 2011), e incluso corroboradas en este caso por medio del comando *isnlarx(data, orders)* del software MATLABTM System Identification Toolbox R2011b, aunque solo para determinados modelos de estructuras polinomiales ARX (ver Anexo II); y puesto que el interés está enfocado en identificar la mejor aproximación lineal del sistema o identificar el modelo no-lineal que lo describe, la excitación deben ser representativa de la clase de excitaciones que sean aplicadas posteriormente en los dispositivos en opinión de R. Pintelon & J. Schoukens (2001). Se recurre entonces a señales generadas por funciones pasos aleatorias, y como el proceso posee configuración operadorcontrolador, las amplitudes de los pasos se basan en las experiencias de los especialistas para lograr cubrir las bandas de frecuencia de interés, o sea, los entornos del punto de operación del sistema.

Para comprobar que las señales sean persistentemente excitantes, se recurre al comando *advice(data)* del software MATLABTM System Identification Toolbox R2011b. Se verifica que el orden de excitación de las señales spFIC y feedback es [50 50]. Luego $\Phi_u(\omega)$ es diferente de cero como mínimo en 50 puntos en el intervalo de frecuencia $-\pi < \omega < \pi$ y u(t) es persistentemente excitante de orden 50.

Por lo tanto:

$$\left|M_n(e^{i\omega})\right|^2 \Phi_{\rm u}(\omega) \equiv 0 \tag{3.1}$$

Implicando que:

$$\left|M_n(e^{i\omega})\right|^2 \equiv 0 \tag{3.2}$$
CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 73 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

3.2 Fase II: Pre-procesamiento de los datos.

Las corridas experimentales tuvieron un tiempo de duración aproximada de 12 veces el tiempo de establecimiento del sistema (t_{ss}), en aras de garantizar la integridad de los parámetros que puedan variar en función de perturbaciones lentas o que puedan aparecer no-linealidades que no estén presentes en los transitorios; aunque una mejor elección hubiese sido de 20 veces el t_{ss} según el criterio de Alonso Díaz. et al (2012). En realidad este tiempo estuvo limitado por las características propias que posee una planta de procesos como la estudiada. Para la recolección de los datos se aprovecha la bondad del software MATLABTM R2011b que permite la comunicación entre el DCS versión CENTUM CS 3000 R3 implementado en la planta y el Open Process Control (OPC) Toolbox. OPC está disponible en la ayuda de MATLABTM a partir de la versión 7.0 como un intercambiador real de datos (Manjunath. & Raman., 2011). Finalmente los datos son exportados al espacio de trabajo del software MATLABTM R2011b en un archivo *.mat.

Como se trabaja en una aplicación off-line, los datos se someten a un pre-procesamiento en busca de deficiencias, pues es probable que no estén en condiciones para ser usados de forma inmediata en el algoritmo de modelación según el criterio de Ljung (1999). Primeramente se pre-procesan los valores de las variables respondiendo a los requerimientos para los modelos lineales y posteriormente para los no-lineales, ya que esto debe responder a los tipos de modelos que se pretendan usar.

3.2.1 Pre-procesamiento de los datos para modelos lineales de sistemas.

Tratamiento de valores atípicos.

Para la detección de atípicos multivariantes se utiliza la estimación robusta de los parámetros de la Distancia de Mahalanobis cuadráticas (D^2) y sus respectivas probabilidades por medio del software IBM SPSS Statistics Versión 20.0.

En este caso, no se cumple el supuesto de que cada variable continua tenga una distribución normal. Sin embargo, según (IBM Corporation.) las comprobaciones empíricas internas indican que este procedimiento es bastante robusto frente a las violaciones del supuesto de distribuciones de probabilidad, no obstante recomienda que se debe tener en cuenta hasta qué punto se cumplen estos supuestos. Tomando un nivel de significación $\alpha = 0.001$ muy conservador recomendado por Hair (1999) como valor umbral para la designación de casos, se obtienen en total 24 atípicos; ya que las probabilidades para D² (significación de la

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 74 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

prueba) arrojaron valores menores o iguales a 0.001. Una vez identificados los valores atípicos, la decisión a tomar con respecto a su tratamiento es extraer las particiones de datos informativos en segmentos, resultando en un total de cinco, los mismos son combinados en un set de datos multiexperimentos , técnica sugerida por Ljung (2007b).

Remoción de tendencias.

Al aplicar el comando *advice(data)* del software MATLABTM System Identification Toolbox R2011b a los datos, se detecta que en las entradas y/o en la salida existe la evidencia de valores medios distintos de cero. Sin embargo al estar en presencia de una aplicación de respuesta en estado transitorio del sistema (respuestas al paso), no resulta útil substraer estos valores según Ljung (2007b).

No se detectan tendencias lineales de las señales, indicado así por el propio comando *advice(data)* del software MATLABTM System Identification Toolbox R2011b.

Re-muestreo de los datos y datos perdidos.

El re-muestreo de los datos para cambiar la razón de muestreo de las señales no es requerido. Se decide trabajar con el intervalo T = 1s ya que el tiempo de muestreo total de las corridas experimentales $0 \le t \le T_N$ es limitado producto de las condiciones de operación, pero la adquisición de los datos es poco costosa al estar implementado un DCS que normalmente toma muestras cada un segundo. Tanto para entradas como salidas no se dan casos de datos perdidos.

3.2.2 Pre-procesamiento de los datos para modelos no-lineales de sistemas.

El pre-procesamiento de los datos con vistas a aplicar estructuras de modelos no-lineales coincide en este caso con el realizado en el subepígrafe anterior. No se eliminan tendencias y no existe procesamiento de datos perdidos en ninguna de sus dos variantes. Luego el set de datos multiexperimentos obtenido como resultado de haber sido tratados los atípicos es perfectamente aplicable si se tuviera que trabajar con modelos no-lineales.

3.3 Fase III: Selección de un set de modelos.

Se busca de la selección de modelos de estructuras con el enfoque de satisfacer los objetivos que con él se persiguen, o sea, cumplan con la intención de uso proyectada. Para este caso particular no es más que lograr un modelo matemático para el tope de la columna desbutanizadora del Bloque 400 que sea capaz de describir con eficiencia y eficacia el proceso tecnológico de dicho sistema. Con el modelo que de aquí resulte, se pretenden

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 75 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

simular nuevas situaciones (diferentes set de datos aunque sí alrededor de los mismos puntos de operación) durante las pruebas de la nueva estrategia de control para la planta en estudio.

Aunque no se esté en presencia de un sistema lineal, siguiendo el criterio de Ljung (1999) de tratar siempre los modelos más simples primero y prestando atención a que se trabaja en los entornos de un punto de operación, se decide iniciar este problema de modelación con estructuras lineales. Las ventajas de sus método de obtención, conducir a esquemas robustos y simples en opinión de Ljung (1999), pueden facilitar una buena aproximación lineal; capaz de convertir con eficiencia y eficacia las entradas en salidas a partir de la dinámica del proceso y del ruido como tal. En caso de que estos modelos simples no pasen las pruebas de validación, se deben usar entonces modelos de estructuras más sofisticadas como las no-lineales.

Al tratarse de un sistema físico, no disponerse de una adecuada percepción de las leyes que lo describen y solo ha sido posible poco más que acceder a corridas de datos de las variables de entradas y salidas; usar modelos compatibles con la parametrización de caja negra es lo bastante lógico y racional. Además con las variables medidas cada cierto instante de tiempo (intervalo de muestreo un segundo) se puede estar hablando de modelos de tiempo discreto, descritos por ecuaciones en diferencias. Una ventaja de este enfoque es que no es necesario usar rutinas adicionales de integración cuando se desarrollen simulaciones, como es el caso en la mayoría de las ecuaciones diferenciales que describen los modelos de tiempo continuo. La principal desventaja es que solo son válidos para el entorno del punto de operación en que han sido estimados (A. Billings. & Coca. 2000).

Ante el hecho de tratarse de un sistema con tiempos de retraso entre las señales de entradas y salidas (Alonso Díaz. et al., 2012), deben ser aplicados modelos que manejen un término adicional de retardo de manera explícita (R. Pintelon. & J. Schoukens., 2001).

Enfocado en las ideas anteriormente expuestas en esta sección, se presentan varios modelos de estructuras tanto lineales como no-lineales para ser aplicados al conjunto de datos obtenidos durante las corridas experimentales desarrolladas.

• Modelos de funciones de transferencia polinomiales lineales de tiempo discreto:

$$A(q)y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t)$$
(3.3)

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 76 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

• Modelos lineales en el espacio de estados, de tiempo discreto en la forma de innovación:

$$x(kT+T) = Ax(kT) + Bu(kT) + Ke(kT)$$
(3.4)

$$y(kT) = Cx(kT) + Du(kT) + e(kT)$$
 (3.5)

$$x(0) = x0 \tag{3.5}$$

• Modelos de funciones de transferencia polinomiales no-lineales de tiempo discreto:

$$g(Z^{t-1};\theta) = g(\varphi(t),\theta)$$
(3.6)

Donde:

$$\varphi(t) = \varphi(Z^{t-1}) \tag{3.7}$$

$$g(\varphi,\theta) = \sum_{k=1}^{n} \alpha_k \mathcal{K}(\beta_k (x - \gamma_k))$$
(3.8)

• Modelos no-lineales Hammerstein & Weiner de tiempo discreto:

$$w(t) = f(u(t)) \tag{3.9}$$

$$x(t) = \frac{B_{j,i}(q)}{F_{j,i}(q)} w(t)$$
(3.10)

$$y(t) = h(x(t)) \tag{3.11}$$

3.4 Fase IV: Obtención del modelo.

Los modelos identificados toman como base el conjunto de datos de estimación, conformado por la primera mitad de los valores de entradas y salidas del sistema, obtenidos durante el proceso experimental.

3.4.1 Identificación del tiempo de retardo del sistema.

Para comenzar a aplicar los métodos de estimación de parámetros, se deben conocer características básicas del sistema relacionadas con el control. Con este objetivo se emplea la respuesta al paso, una vía simple, segura y robusta de construir modelos lineales no paramétricos del tipo respuesta finita al impulso (FIR), la cual proporciona información sobre el tiempo de retardo con suficiente grado de exactitud según Ljung (1999); Haugen (2010) y Chen, Ohlsson & Ljung (2012). Se obtienen los siguientes resultados:

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.



Figura 3.1: Variaciones del punto de ajuste del FIC (el eje de las ordenadas "y" muestras las variaciones del punto de ajuste en m3/h; y el de las abscisas "x", el tiempo en segundos)



Figura 3.2: Comportamiento de la temperatura del tope de la columna, respecto a las variaciones del punto de ajuste del FIC (el eje de las ordenadas muestra la temperatura en °C; y el de las abscisas, el tiempo en segundos)



Figura 3.3: Respuesta al paso (integración de la señal impulsiva obtenida por correlación). De análisis gráfico de la respuesta al impulso obtenida por correlación, figura 3.3, se puede notar un tiempo de retardo estimado de 65 segundos, el cual es requerido más adelante para llevar a cabo el proceso de modelación.

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 78 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

3.4.2 Obtención de modelos polinomiales lineales a partir del método de variables instrumentales (IV):

Un buen punto de partida en la modelación es comenzar por aplicar, quizás la forma más inmediata de parametrización usando funciones rotacionales (Ljung., 1999), obteniéndose así los modelos de funciones de transferencia también conocidos como modelos polinomiales.

Se selecciona el método de las variables instrumentales porque a diferencia de otros métodos de estimación de parámetros, posee la ventaja de que no es sensible al tipo de ruido que afecta la respuesta del sistema, o sea, no asume que el ruido es blanco según Ljung (2007b); por lo que es recomendable su uso cuando se desconocen las particularidades del tipo de ruido como en este caso. Además para el tamaño muestral (N = 6365) bastante grande, el error cuadrático medio de los estimados por variables instrumentales debe tender a cero, mientras que los LSE tienden asintóticamente al cuadrado de sus desviaciones en opinión de R. Pintelon & J. Schoukens (2001).

Debido a la presencia de lazo cerrado se emplea la variante refinada desarrollada por Gilson. et al (2009). Según este autor la misma es capaz de identificar consistentemente modelos de plantas en lazo cerrado y no depende de las características en cuanto a la linealidad del controlador aplicado al proceso.

Se asume que el sistema generador los datos está dado por las relaciones:

$$y(t) = G_o(q)u(t) + H_o(q)e_0(t)$$
(3.12)

$$u(t) = C_c(q)[r(t) - y(t)]$$
(3.13)

<u>Donde</u>: r(t): para este caso es el punto de ajuste del FIC, como supuesto debe ser persistentemente excitante, lo cual es comprobado anteriormente por el uso del comando *advice(data)* del software MATLABTM System Identification Toolbox R2011b.

 $C_c(q)$: es el controlador del lazo de retroalimentación, puede ser conocido a no y está definido por:

$$C_c(q) = \frac{Q(q^{-1})}{P(q^{-1})} = \frac{q_0 + q_1 q^{-1} + \dots + q_{n_c} q^{-n_c}}{p_0 + p_1 q^{-1} + \dots + p_{n_c} q^{-n_c}}$$
(3.14)

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 79 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

3.4.3 Obtención de modelos polinomiales lineales a partir del método de minimización del error de predicción.

Se aplica la variante alternativa sugerida por Ljung (1999) y Van den Hof. et al (2013) de identificación directa ante la presencia de lazo cerrado.

Según estos autores el enfoque es una generalización directa de los métodos clásicos de modelación a lazo cerrado, donde se asume que la salida del sistema puede ser expresada en términos de las entradas como:

$$y(t) = \sum_{x_k \in \mathcal{X}} G_0(q) u_k(t) + H_0(q) e(t) + r(t)$$
(3.15)

Además el método supone que el lazo de retroalimentación sea suficientemente excitado con la señal externa r(t) y no sea algebraico, o sea, que el modelo que lo describe debe tener como mínimo un paso de tiempo de retardo. Para este caso ambos se cumplen, lo cual ha sido anteriormente analizado.

3.4.4 Obtención de modelos lineales en el espacio de estados a partir del método de minimización del error de predicción.

Igualmente se aplica la variante alternativa sugerida por Ljung (1999) y Van den Hof. et al (2013) de identificación directa ante la presencia de lazo cerrado para la aplicación iterativa de la minimización del error de predicción. Se tiene en cuanta además la opción de refinar los modelos obtenidos simpre que se pueda mejorar el resultado inicial.

3.4.5 Resumen de los modelos obtenidos.

Varios son los modelos obtenidos por medio del software MATLABTM System Identification Toolbox R2011b. No se muestran resultados del método de las variables instrumentales, ya que el algoritmo parte con la estimación de modelo ARX el cual manifiesta un pésimo desempeño al representar la estructura del sistema a partir de los datos. Con vista a que los modelos cumplan con la intención de uso, se emplea el enfoque de simulación. La siguiente tabla muestra un resumen de los modelos obtenidos.

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

Modelo	Polinomios [*] o matrices ^{**} incluidas según el modelo de estructura.							
1000010	Α	В	С	D	F	K		
Bj111365***	-	[1 1]	1	1	[3 3]	-		
Bj112265	-	[1 1]	1	2	[2 2]	-		
Bj211365	-	[2 2]	1	1	[3 3]	-		
Bj411365	-	[4 4]	1	1	[3 3]	-		
SS165 ^{***}	[129 129]	[129 2]	[1 129]	[0 0]	-	[129 1]		
Ss265	[130 130]	[130 2]	[1 130]	[0 0]	-	[130 1]		
Ss265r	[130 130]	[130 2]	[1 130]	[0 0]	-	[130 1]		
Oe2265***	-	[2 2]	-	-	[2 2]	-		
Oe2165	-	[2 2]	-	-	[1 1]	-		
Oe2365	-	[2 2]	-	-	[3 3]	-		
Oe2565	-	[2 2]	-	-	[5 5]	-		
Oe1565	-	[1 1]	-	-	[5 5]	-		
Oe4565	-	[4 4]	-	-	[5 5]	-		
Oe4365	-	[4 4]	-	-	[3 3]	-		

Tabla 3.1: Resumen de los diferentes modelos obtenidos.

* En un determinado modelo de estructura polinomial A, B, C, D y F se refieren a los polinomios que lo integran; representándose su orden por un valor escalar o las coordenadas de un vector en caso multivariable.

^{**} En un determinado modelo de estructura de estados en el espacio A, B, C, D y K se refieren a las matrices que lo integran; que tienen un determinado orden en dependencia de la cantidad de variables que describen el sistema y la cantidad de estados estimados por el modelo.

*** Bj, Ss, u Oe seguido por una determinada secuencia numérica significa un modelo de estructura de tipo Box – Jenkins (BJ), de estados en el espacio (SS) o de error de salida (OE) según el caso; de orden y retardo entre las entradas y salidas dado por la secuencia. En el caso particular de Ss poseer una letra minúscula cerrando la secuencia de números quiere decir que el modelo de estados en el espacio ha sido mejorado mediante el método de minimización del error de predicción.

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

3.5 Fase V: Validación de los modelos.

Con el objetivo de encontrar un modelo que sea lo suficientemente bueno para simular el proceso del sistema estudiado, se someten a una comparación los obtenidos en la sección anterior. Los elementos a tener en cuenta para evaluar estas comparaciones son: el por ciento de ajuste del modelo para reproducir datos nuevos en términos de simulación; el Criterio del error final de predicción (FPE) de Akaike y el análisis de los residuos en cuanto a la independencia entre los errores del modelo (Test de Whiteness) y la independencia entre las diferentes entradas del sistema con respecto a tales residuos. En este sentido se destaca que un modelo es mejor en la medida que logre un adecuado compromiso entre estos parámetros, o sea, logre un alto por ciento de varianza explicada (R^2) según Ljung (1999) y Gujarati (2004); posea un menor valor del criterio FPE acorde a la teoría de Akaike Ljung (2007b) y presente residuos que resulten independientes entre ellos y también de las entradas del modelo. En la tabla 3.2 se muestran los valores para cada uno de estos parámetros según el modelo de que se trate.

De la misma se puede observar que las estructuras Box–Jenkins presentan altos valores de ajuste y muy bajos valores del criterio FPE, además de pasar sendas pruebas de residuos. Por su parte los modelos de estados en el espacio poseen valores bajos del criterio FPE, similares a los BJ, pero son invalidados por la prueba de independencia entre los residuos y no tienen un buen desempeño del por ciento de ajuste. En cuanto a los OE, se evidencia una situación bastante normal de su comportamiento, los residuos que generan no son capaces de pasar la prueba de Whiteness. Sin embargo, autores consultados durante la investigación, R. Pintelon & J. Schoukens (2001) y Ljung (2007b), plantean que para modelos de este tipo se debe asegurar que muestren independencia entre las entradas y los errores, lo que sí es favorable en este caso. La causa es que en estas situaciones la modelación se enfoca en la dinámica de $G_0(q)$ y no en las propiedades de los disturbios $H_0(q)$, por lo que se presta menos atención al resultado de la prueba de Whiteness. En lo referido al ajuste del modelo, se nota que los de más bajo orden poseen buen ajuste, comparado a los BJ; pero en el criterio FPE muestran desventajas.

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 82 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

Modolo	Ajuste	Criterio	Análisis de residu	ios (ver Anexo III)	
WIGHEIO	(%)	FPE	Test de Whiteness [*]	Independencia ^{**}	
Bj111365	88.23	$1.40454e^{-005}$	Si	Si	
Bj112265	84.72	1.91647 <i>e</i> ⁻⁰⁰⁵	Si	Si	
Bj211365	87.86	$1.44523e^{-005}$	Si	Si	
Bj411365	81.90	$1.86817e^{-005}$	Si	Si	
SS165	55.19	$1.60247e^{-005}$	No	Si	
Ss265	37.09	2.16480 <i>e</i> ⁻⁰⁰⁵	No	Si	
Ss265r	46.28	2.16480 <i>e</i> ⁻⁰⁰⁵	No	Si	
Oe2265	84.36	398.165	No	Si	
Oe2165	84.27	86.7732	No	Si	
Oe2365	87.86	0.33621	No	Si	
Oe2565	67.48	599.499	No	Si	
Oe1565	85.46	108.496	No	Si	
Oe4565	89.43	0.10016	No	Si	
Oe4365	81.9	30.9489	No	Si	

Tabla 3.1: Valores de los parámetros a tener en cuenta en la comparación de los modelos.

* y ** se refieren a que si los modelos pasan las pruebas de Whiteness y de independencia.

3.5.1 Selección del mejor modelo.

Tomando como base los resultados de la tabla 3.2 se decide seleccionar el modelo Bj111365 como el más adecuado para el fin propuesto en la investigación, obtener un modelo capaz de simular el proceso operacional del tope de la columna desbutanizadora T- CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 83 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

401/1 con eficiencia y eficacia. El mismo presenta un buen desempeño (por ciento de varianza explicada ($R^2 = 88.23$) y buena calidad para la simulación en situaciones donde sea probado para set de datos diferentes (el valor del criterio FPE = 1.40454*e* - 005). Además pasa las pruebas de los residuos demostrando que sus errores son independientes entre ellos, así como que son idependientes de las entradas del modelo. Luego la estructura del modelo resultante, de tiempo discreto, es de la siguente forma según Ljung (2007b):

$$y(t) = \sum_{i=1}^{na} \frac{B_i(q)}{F_i(q)} u_i(t - nk_i) + \frac{C(q)}{D(q)} e(t)$$
(3.16)

.....

Proporcionando una parametrización completamente independiente de la dinámica y el ruido con funciones polinomiales rotacionales. Donde los polinomios B, C, D y F; para el sistema compuesto por dos entradas y una salida, están dados por:

$$B_1(q) = -0.001602q^{-65} \tag{3.17}$$

$$B_2(q) = 0.00811q^{-65} \tag{3.18}$$

$$C(q) = 1 + 0.005926q^{-1} \tag{3.19}$$

$$D(q) = 1 - q^{-1} (3.20)$$

$$F_1(q) = 1 - 1.367q^{-1} - 0.113q^{-2} + 0.4802q^{-3}$$
(3.21)

$$F_2(q) = 1 - 1.362q^{-1} + 0.4521q^{-2} - 0.0896q^{-3}$$
(3.22)

El cual representa un modelo matemático del sistema de orden superior, específicamente de tercer orden, coincidiendo con lo planteado por autores como Preeti & Singh Beniwal (2012).

El hecho de obtener en modelo Box-Jenkins en el proceso de identificar una buena aproximación lineal para una aplicación como la actual, no es para nada sorprendente. Resultados similares han sido notados en la bibliografía consultada por autores tales como Ljung (1999); J. Schoukens. et al (2006) y Van den Hof & Leskens (2007). Lo que ocurre en estos casos, según los autores antes mencionados, es que se debe usar un complemento que consista de la varianza del ruido perturbador más la fuente de ruido no-lineal; y un método clásico de modelación lineal que estime además un modelo de ruido, como el BJ, hace esto automáticamente.

3.6 Eficacia y eficiencia del modelo Box-Jenkins obtenido.

Tratándose de obtener un modelo capaz de simular el sistema del tope de la columna desbutanizadora T-401/1, los resultados alcanzados en la fase validación son una prueba de

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 84 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

su eficacia. El mismo explica el 88.23 por ciento de la varianza, muestra del buen ajuste al reproducir datos nuevos. Otra evidencia de la buena calidad para la simulación en situaciones donde sea probado para set de datos diferentes, está dada por el pequeño valor del criterio FPE, igual a 1.40454e - 005. Además al pasar ambas pruebas de los residuos es una señal de que las salidas han sido predichas por los datos pasados, o sea, y(t) es adecuadamente explicada por el modelo.

A pesar de que el modelo obtenido es de tercer orden, resulta ser eficiente; debido a las siguientes razones: cumple con su intensión de uso, lo que ha sido demostrado a través de su eficacia, y lo logra a un bajo precio. En primer lugar, porque es capaz de representar un sistema no-lineal mediante una estructura lineal, lo que influye en el precio que se paga por conseguirlo, pues el algoritmo mediante el cual se obtiene es menos complejo. En segundo lugar, el costo asociado al uso del modelo es relativamente bajo, además de ser una estructura lineal, la mayoría de los polinomios contenidos en las funciones rotacionales que lo componen son de bajo orden, requiriéndose menor cantidad de parámetros; convirtiéndolo más fácil de usar para la simulación y el diseño de control.

Como se ha podido ver, el modelo que resulta de la presente investigación es adecuado para la simulación de nuevos datos, siempre y cuando se trabaje en condiciones operacionales semejantes a las del momento para el cual fue creado. El mismo, desarrollado por medio del software MATLABTM System Identification Toolbox R2011b, puede ser implementado en ambientes de simulación tales como el Simulink de MATLABTM con el objetivo de diseñar controladores; debido a las bondades de estas herramientas según criterio de Ljung (2007b). Teniendo en cuenta estos aspectos, el modelo Box-Jenkins obtenido puede ser puesto al servicio de la Refinería de Petróleos de Cienfuegos con el propósito de simular el proceso operacional del tope de la columna desbutanizadora; tanto en las condiciones actual del sistema como con la nueva estrategia de control avanzado en la cual se ha estado trabajado para mejorar el control de temperatura en el tope de la columna fraccionadora de gases.

CAPÍTULO 3: Implementación del procedimiento para modelar el sistema del tope de la columna 85 desbutanizadora T-401/1 de la Refinería de Petróleos de Cienfuegos.

3.7 Conclusiones del capítulo.

- 1- El método de minimización del error de predicción con la variante de identificación directa, aplicado en función de estructuras polinomiales lineales, proporcionó el mejor modelo a utilizar en la presente investigación.
- 2- Se evidenció que aplicado el criterio de Ljung (2007b) referido a tratar los modelos más simples primero, fue posible obtener un modelo matemático relativamente sencillo, pues a pesar de ser de orden superior, posee una estructura lineal e incluye pocos parámetros.

CONCLUSIONES

- 1- Los modelos lineales son básicos y versátiles, ya que además de representar procesos lineales, son capaces de proporcionan en muchos casos una buena aproximación lineal, cuando no-linealidades han sido detectadas o sospechadas.
- 2- En la práctica de la modelación matemática, los modelos lineales poseen varias ventajas con respecto a los no-lineales, sus método de obtención son más sencillos y robustos.
- 3- Se diseñó un procedimiento para identificar sistemas que operan en configuración de lazo cerrado, enfocado en obtener modelos paramétricos, el cual requiere de la integración de conocimientos previos del proceso, intuición y herramientas de cálculo.
- 4- Se obtuvo un modelo matemático lineal, con la menor complejidad estructural posible, que evidencia un buen desempeño para describir el proceso del sistema del tope de la columna desbutanizadora T-401/1 alrededor del punto de operación para el cual fue estimado; lo que constituye una muestra de su eficiencia y eficacia.

RECOMENDACIONES

- El modelo de estructura lineal obtenido debe ser aplicado durante el desarrollo de la nueva estrategia de control, con el fin de simular el proceso operacional del sistema y no con el objetivo de predecir parámetros operacionales.
- 2- Modelar el proceso operacional del tope de la columna desbutanizadora para otros rangos de valores de los parámetros operacionales, ya que esto permitirá implementar la nueva estrategia para un mayor horizonte de control automático.

BIBLIOGRAFÍA

- Al-Odienat., A. I., & Al-Lawama., A. A. (2008). The advantages of pid fuzzy controllers over the conventional types., 5., American Journal of Applied Sciences., 6., 653-658.
- Alonso Díaz., F., Herrera Fernández., F., & Madrazo Valladares., Y. (2012, Abril 9). Propuesta de composición del para control producto final de fraccionamiento licuados en columna de gases del petróleo de la refinería de petróleo "camilo cienfuegos". La Habana, Cuba.
- Anderson., B. D. O., & Moore., J. B. (1979). Optimal filtering. . Upper Saddle River, N.J.: Prentice Hall.
- Aslam., F., & Kaur., G. (2011). Comparative analysis of conventional, p, pi, pid and fuzzy logic controllers for the efficient control of concentration in cstr., 6., International Journal of Computer Applications., 17., 12-16.
- Billings., A., & Coca., D. (2000). Identification of narmax and related model. En , Control Systems, Robotics, and Automation., VI.
- Bjorklund., S. (2003). A survey and comparison of time-delay estimation methods in linear systems., 140. Linkopings University. Recuperado a partir de http://www.control.isy.liu.se/research/reports/LicentiateThesis/Lic1061.pdf.
- Borjas., S., & Garcia., C. (2011). Subspace identification for industrial processes., Scielo.,
 Sociedade Brasileira de Matemática Aplicada e Computacional., 12.(3.).
 Recuperado a partir de http://dx.doi.org/10.5540/tema.2011.012.03.0183.
- Cramér., H. (1946). Mathematical methods of statistics. New York.: Princeton University Press.
- Chen., S., Billings., S., & Luo., W. (1989). Orthogonal least squares methods and their application to non-linear system identification., Taylor & Francis., International Journal of Control., 50.(5.), 1873 - 1896.
- Chen., T., Ohlsson., H., Goodwin., G. C., & Ljung., L. (2011). Kernel selection in linear system identification. part ii: a classical perspective. En (págs.. 4326-4331.).
 Orlando, FL, USA.: IEEE. Recuperado a partir de

http://www.nt.ntnu.no/users/skoge/prost/proceedings/cdc-ecc-2011/data/papers/0979.pdf.

- Chen., T., Ohlsson., H., & Ljung., L. (2012). On the estimation of transfer functions, regularizations and gaussian processes-revisited., Elsevier., Automatica., 48.(8.), 1525-1535.
- Dankers., A., Van den Hof., P. M., Heuberger., P. S., & Bombois., X. (2012). Dynamic network structure identification with prediction error methods - basic examples. Belgium.
- Filzmoser., P. (2005). A multivariate outlier detection method. Department of Statistics and Probability Theory, Vienna, Austria.
- Fisher., R. (1912). On an absolute criterion for fitting frequency curves. , 155.
- Fuyin., D., & Weifeng., D. (2009, Agosto 21). Design of a three-input fuzzy logic controller and the method of its rules reduction. Huangshan, P. R. China.
- Gilson., M., Garnier., H., Young., P. C., & Van den Hof., P. M. (2009). Refined instrumental variable methods for closed-loop system identification. 15th IFAC Symposium on System Identification. Recuperado a partir de http://www.dcsc.tudelft.nl/~pvandenhof/Paperfiles/Gilson&etal_SYSID2009Prepr.p df.
- Gujarati., D. N. (2004). Basic econometrics. (Fourth Edition.), 1002. New York.: The McGraw-Hill Companies.
- Hair., J. F. (1999). Análisis multivariante (Quinta Edición.), p. 832. Madrid.: Prentice Hall Iberia.
- Haugen., F. (2010, Julio). Ziegler-nichols' closed-loopmethod. Recuperado a partir de http://techteach.no/publications/articles/zn_closed_loop_method/zn_closed_loop_m ethod.pdf.

IBM Corporation. Ibm spss statistics. tutorial.

ŁAWRYNCZUK., M., & TATJEWSKI., P. (2010). Nonlinear predictive control based on neural multi-models., Int. J. Appl. Math. Comput. Sci., 20.(1.), 7–21.

- Liu., T., & Gao., F. (2012). Industrial process identification and control design., 467. New York.: Springer-Verlag.
- Ljung., L. (1999). System identification. theory for the user. (2.), 609. U.S.: Prentice Hall.
- Ljung., L. (2007a). Some aspects on nonlinear system identification. Sweden.
- Ljung., L. (2007b). System identification toolbox 7. user's guide. (Versión 7.0.). U.S.: The MathWorks, Inc.
- Ljung., L. (2010). Perspectives on system identification. Division of Automatic Control, Linköpings Universitet, Sweden. Recuperado a partir de http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.184.8654&rep=rep1&typ e=pdf.
- Ljung., L., & Vicinio., A. (2005). Special issue on identification., IEEE Trans., Automatic Control., AC-50.
- López Guillén., M. E. (1999). Identificación de sistemas. Aplicación al modelado de un motor de continua.
- Luenberger., D. G., & Ye., Y. (2008). Linear and nonlinear programming. (Third Edition.), 546. New York, USA.: Springer Science + Business Media, LLC.
- Lyzell., C., Glad., T., Enqvist., M., & Ljung., L. (2009). Identification aspects of ritt's algorithm for discrete-time systems. Saint-Malo, France.
- Manjunath., R. M., & Raman., S. J. (2011). Fuzzy adaptive pid for flow control system based
 opc., IJCA Special Issue., Computational Science New Dimensions & Perspectives., 8.
- Medeiros., J., & Meneghetti Ugulino., F. (2007, Noviembre 5). Artificial intelligence techniques applied in a simulated oil distillation system. , 19th International Congress of Mechanical Engineering - Brasília, D. , 7.
- Ogata., K. (1998). Ingeniería de control moderna. (Tercera edición.), p. 1015. México.: PRENTICE-HALL HISPANOAMERICANA, S.A.
- Paraskevopoulos., P. N. (2002). Modern control engineering., 736. New York.: Marcel Dekker, Inc.

- Pillonetto., G., & De Nicolao, G. (2011). Kernel selection in linear system identification. part i: a gaussian process perspective. En (págs.. 4318-4325.). Orlando, FL, USA.: IEEE.
- Pintelon., R., Schoukens., J., & Rolain., Y. (2003). Uncertainty of transfer function modeling using prior estimated noise models. Lugar. Recuperado a partir de http://wwwtw.vub.ac.be/elec/Papers%20on%20web/Papers/RikPintelon/IFAC03Pin telon-2.pdf.
- Pintelon., R., & Schoukens., J. (2001). System identification. a frequency domain approach. , 605. New york.: IEEE Press.
- Pitta., R. N., & Odloak., D. (2012). Closed-loop re-identification of an industrial debutanizer column., The International Federation of Automatic Control Furama Riverfront, Singapore.
- Preeti., M., & Singh Beniwal., D. N. (2012). Comparison of conventional and fuzzy p/pi/pd/pid controller for higher order non linear plant with high dead time., 8., International Journal of Scientific and Research Publications., 2., 5.
- Rao., K., & Reddy., B. (2010). Fuzzy pi and integrating type fuzzy pid controllers of linear, nonlinear and time -delay systems. , 6., International Journal of Computer Applications., 1.
- Rousseeuw., P., & Van Zomeren., B. (1990). Unmasking multivariate outliers and leverage points., JAMER STATIST ASSN., Journal of the American Statistical Association., 85.(411.), 633-651.
- Sakthivel., G., Anandhi., T., & Natarajan., S. P. (2011). Modelling and real time implementation of digital pi controller for a non linear process., Global Research Publishing., Journal of Innovative Research in Engineering and Sciences., 2.(5.), 274 - 290.
- Schoukens., J., Pintelon., R., & Rolain., Y. (2006). Identification of nonlinear and linear systems, similarities, differences, challenges. 14th IFAC Symposium on System

Identification. Recuperado a partir de http://wwwtw.vub.ac.be/elec/Papers%20on%20web/Papers/RikPintelon/IFAC2006 Pintelon.pdf.

- Sultan., F., & Ahmed., M. (2010). Selection of best outlier detection method using regression analysis., IJCSR International Journal of Computer Science and Research., 1(1), 1-9.
- The MathWorks, Inc. (2011, Agosto 14). Matlab system identification toolbox. getting started.
- Tóth., R., Heuberger., P. S., & Van den Hof., P. M. (2009). Asymptotically optimal orthonormal basis functions for lpv system identification., Elsevier., Automatica., 45., 1359–1370.
- Van den Hof., P. M., Dankers., A., Heuberger., P. S., & Bombois., X. (2013). Identification of dynamic models in complex networks with prediction error methods—basic methods for consistent module estimates., Elsevier Ltd., Automatica 49., 2994– 3006.
- Van den Hof., P. M., & Leskens., M. (2007). Closed-loop identification of multivariable processes

with part of the inputs controlled., Taylor & Francis., International Journal of Control., 80.(10.), 1552–1561.

- VAN DONKELAAR., E. T. (2000). Improvement of efficiency in identification and model predictive control of industrial processes. a flexible linear parametrization approach. Netherlands.
- Van Overschee., P., & DeMoor., B. (1996). Subspace indentification of linear system. theory, implementation, aplication. Kluwer Academic Publishers.
- Verheagen., M. (1994). Identification of the deterministic part of mimo state space models, given in innovation form from input-output data. , Automática., 30.(1.), 61-74.
- Wald., A. (1949). Note on the consistency of the maximum likelihood estimate. En (págs.. 595-601.).
- Weiyong., Z., & Dexian., H. (2008). Adaptive state feedback predictive control and expert control for a delayed coking furnace., Chinese Journal of Chemical Engineering., 16.(4.), 590—598.

- Wigren., T., & Schoukens., J. (2013). Three free data sets for development and benchmarking in nonlinear system identification. En (págs.. 2933-2938.). Zürich, Switzerland.: IEEE. Recuperado a partir de http://wwwtw.vub.ac.be/elec/Papers%20on%20web/Papers/JohanSchoukens/ECC2 013-Three%20free%20data%20sets-Wigren.pdf.
- Zill., D. G. (1997). Ecuaciones diferenciales, con aplicaciones del modelado. (6ta.). México.: International Thomson Editores.

ANEXOS

Anexo I: Diagrama de flujo del proceso de la columna desbutanizadora T-401/1



Fuente: Sistema de Control Distribuido (DCS, de sus siglas en inglés), realizado por Yokogawa, versión CENTUM CS 3000 R3 e implementado en la planta.

Anexo II: Comparación de modelos polinomiales ARX subyacentes de los datos.

El comando isnlarx(DATA, ORDERS) del software MATLABTM System Identification Toolbox R2011b consiste en una prueba de hipótesis sobre el modelos subyacente de los datos, contrastando que sea linear ARX contra no-lineal ARX.

Hipótesis nula H_o : No-linealidades son detectadas en los datos, el modelo lineal es rechazado.

Hipótesis alternativa H_1 : No-linealidades no son detectadas en los datos, el modelo lineal no es rechazado.

Como resultado evalúa si no-linealidades han sido detectadas en los datos. El test solo se enmarca en comparar si modelos no-lineales ARX de determinados órdenes son significantemente mejor que los lineales ARX lo cual constituye una de sus desventajas (The MathWorks, Inc.). Estadísticos de contraste NLHyp o Valor NL.

NLHyp = 0: se rechaza Ho por lo que el modelo lineal no es rechazado.

NLHyp = 1: no hay evidencias para rechazar Ho por lo que el modelo lineal es rechazado.

Valor NL (NLvalue): la parte estimada del error cuadrático medio explicado por la nolinealidad. Se debe comparar con el error cuadrático medio (Sigma del ruido) de la salida no explicada. Valor NL es cero si el modelo lineal no es rechazado.

Razón de detección (DetectRatio): La razón de las pruebas estadísticas y el umbral de detección. Valores pequeños (menores que 0.5) o valores grandes (mayores que 2) significan que es test es robusto. Valores de la razón de detección cercanos a uno (1) significa que el test está en un punto o momento justo antes de un cambio notable de la detección de la no-linealidad.

Anexo II: Comparación de modelos polinomiales ARX subyacentes de los datos. Continuación.

	NLHyp	Desviación	Dazón da	No-linealidades	
Modelo		estándar del	Kazoli uc	detectadas en los	
		ruido estimado	detección.	datos.	
isnlarx(signalexcitation,	1	0.0000750	4.1985	Si	
[1 1 65])		0.0033753			
isnlarx(feedbacksignal,	1	0.0022816	1.6812	Si	
[1 1 65])		0.0032810			
isnlarx(multisignal, [1 1	1	0.0032948	0 1868	Si	
1 65 65])	1	0.0032748	0.1000	51	
isnlarx(signalexcitation,	1	0.0020214	1.186	Si	
[3 3 65])	1	0.0030214			
isnlarx(feedbacksignal,	0	_	-	No	
[3 3 65])	Ū				
isnlarx(multisignal, [3 3	0	_	-	No	
3 65 65])	Ū				
isnlarx(signalexcitation,	0	_	-	No	
[5 5 65])	Ū				
isnlarx(multisignal, [5 5	0	_	_	No	
5 65 65])	Ū				
isnlarx(signalexcitation,	0	_	-	No	
[10 10 65])	Ŭ				
isnlarx(signalexcitation,	0	_	_	No	
[50 50 65])			-	110	
isnlarx(signalexcitation,	1	0.0032122	1.8013	Si	
[2 2 65])				~1	



Anexo III: Gráficos de análisis de residuos.

Modelos de estructuras polinomiales lineales Box-Jenkins.

Anexo III: Gráficos de análisis de residuos.

Continuación.

Modelos de estructuras en espacio de estados.



Anexo III: Gráficos de análisis de residuos.

Continuación.

Modelos de estructuras polinomiales lineales de errores de salida.

